

Subdirección General de Archivos

noviembre 2019 **JORNADA TÉCNICA**

Reutilización de la información del sector público en los archivos **ARCHIVOS ABIERTOS**

**Web semántica y el intercambio de datos archivísticos: casos de estudio**

Ana María López Cuadrado

Archivos de la Comunidad de Madrid



Subdirección General de Archivos

noviembre 2019

JORNADA TÉCNICA

Reutilización de la información  
del sector público en los archivos

ARCHIVOS ABIERTOS

# Web semántica y el intercambio de datos archivísticos: casos de estudio

Ana María López Cuadrado



# 1. Qué es la normalización de la descripción archivística

Descripción estandarizada de entidades archivísticas permitiendo el intercambio de datos entre diferentes sistemas archivísticos.



Realizar la tarea de descripción archivística mediante procesos reglados y establecidos, comunes a toda la institución, teniendo una misión y objetivo claro y definiendo los proyectos a realizar

# ¿Qué ha sucedido?

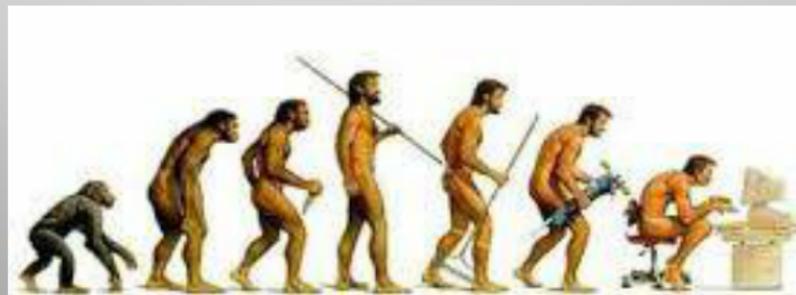
Cambios reflejados estadísticamente en el perfil de nuestros usuarios y en su forma de acceso a la información (vocabulario pensado para archiveros)

Adaptación a la normativa nacional e internacional de contenido y estructura de datos

Multilingüismo

Avances tecnológicos y digitales (web semántica, Humanidades Digitales, Big Data)

Desarrollo de las teorías de organización de las instituciones y la gestión por procesos.



# ¿En qué situación nos encontramos?



- Única o múltiples bases de datos relacionales abiertas o no al público.
- Multitud de reglas y normativas
- Profesionales muy formados con conocimientos especializados



- Información sin definir ni normalizar con ruido y falta de organización
- Falta de organización en el trabajo, repetición de tareas



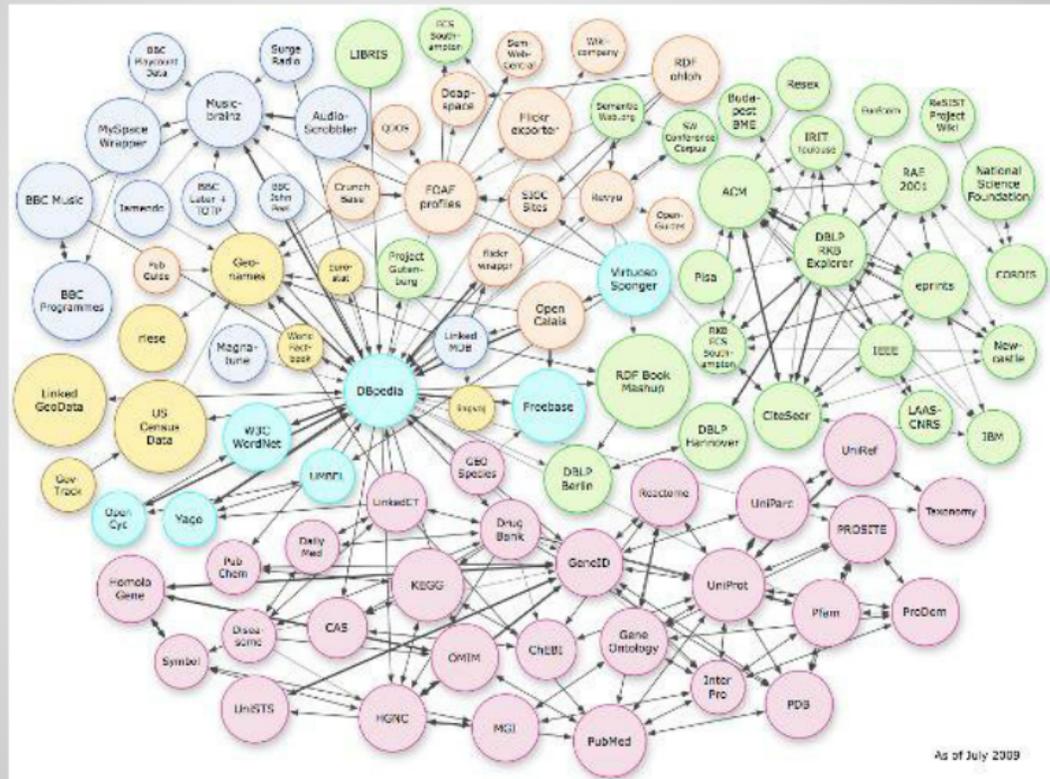
# ¿Por qué la normalización en las descripciones archivísticas?

¿Qué beneficios aporta la normalización en la descripción archivística?

1. Mayor eficacia y mejor gestión de los recursos
2. Mayor facilidad en las comunicaciones entre unidades o personas
3. Uso de indicadores homogéneos para comparar la ejecución y el desarrollo de las mismas actividades en diversos ámbitos geográficos o temporales
4. Intercambio de información entre instituciones con contenidos similares, lo que reduce tiempo y carga de trabajo al ser más efectivo, optimizando los resultados obtenidos por los demás y compartiendo tus resultados.



# 3. La Web Semántica y los Datos Abiertos Enlazados



¿Qué es el enriquecimiento de registros?



Dos roles, con fronteras cada día más diluídas



CONSUMIDORES

WWW

- Somos consumidores cuando usamos la Web consultando **fuentes de información**:
  - Para organizar un viaje: compra de billetes, hotel, sitios de interés, etc.
  - Para comprar un libro: desde un autor que nos gusta a un tema determinado
- Ya que muchas veces resulta tedioso/difícil encontrar lo que buscamos, **los servicios más exitosos son aquellos que son capaces de agregar/derivar información** que nos interesa: google, twitter, etc.



PRODUCTORES

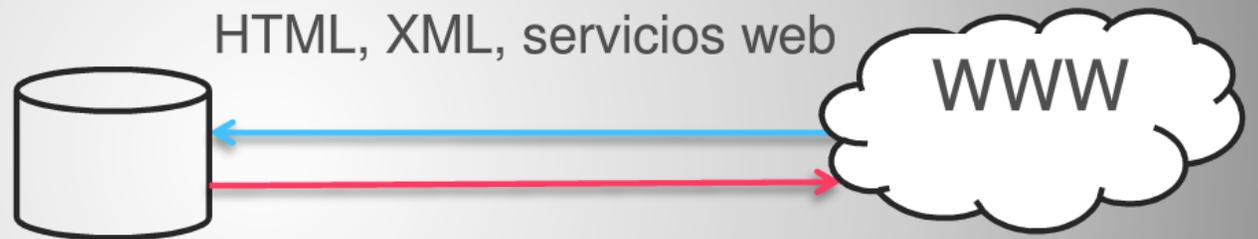
WWW

- Necesitan que su información/servicio sea **visible, útil y relevante**
- **Reutilizar** información de otros aumenta las capacidades del servicio
- **Compartir** información de una manera sencilla y estándar amplía el potencial de la misma y la visibilidad de la organización

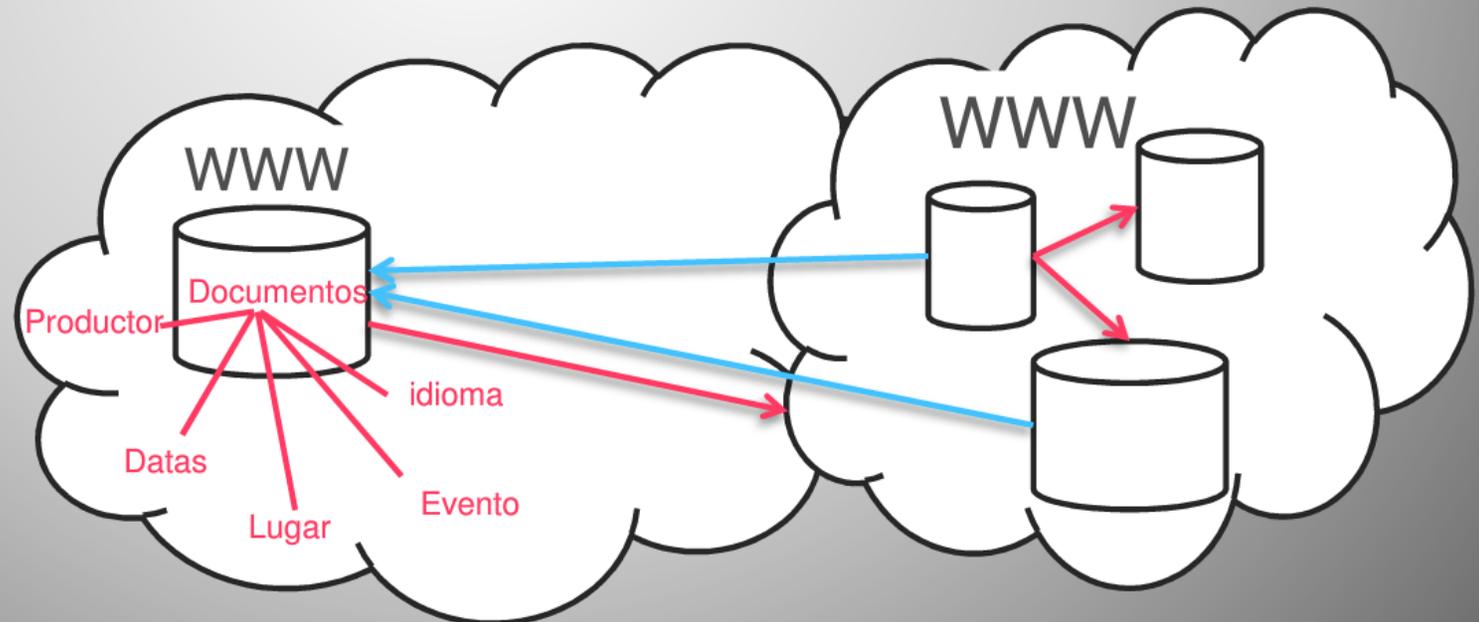
Web 1.0



Web 2.0

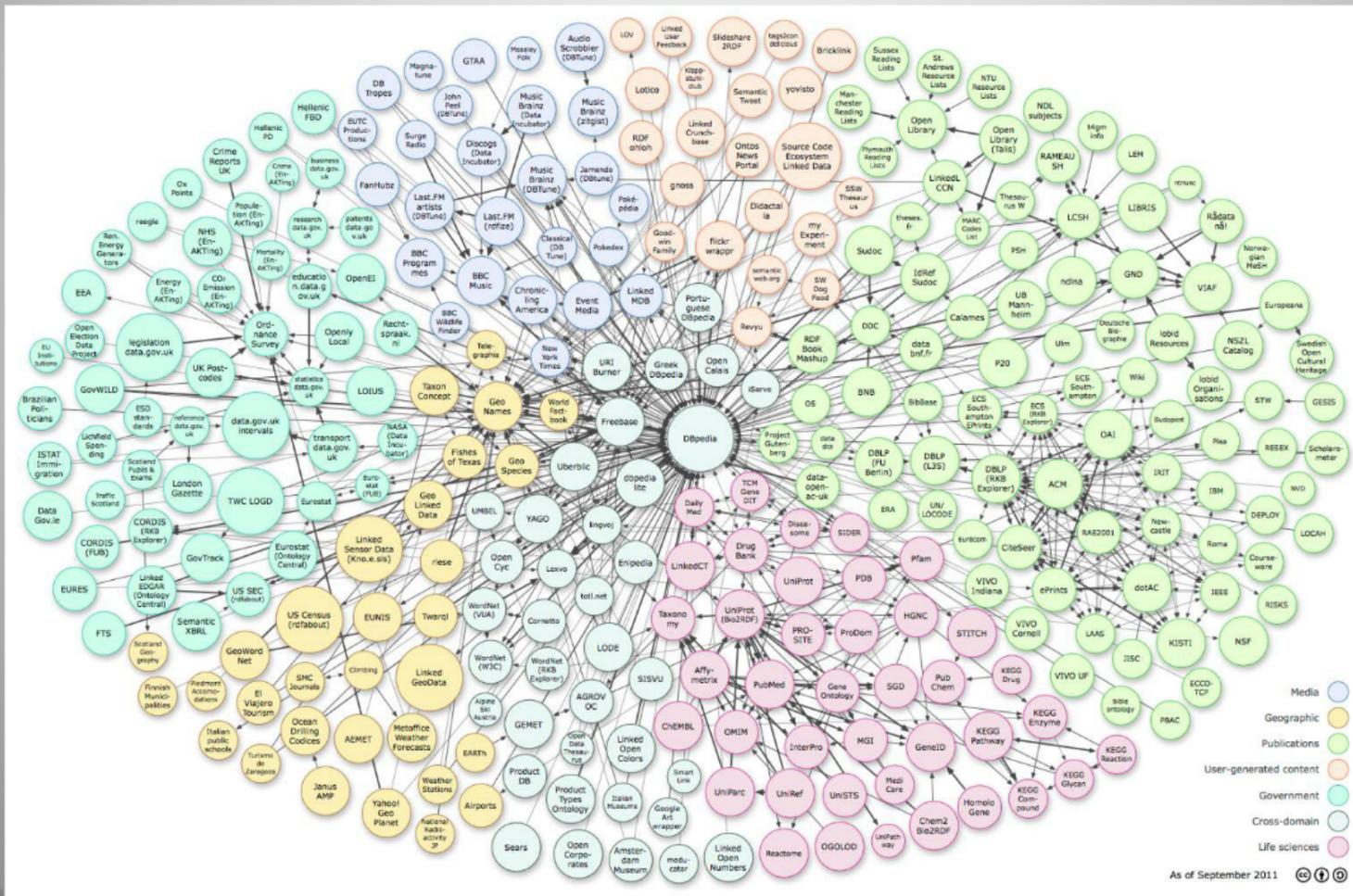


Linked Data



# Linked Data

Comparte tus datos y reutiliza los de otros



# Linked Data

- Publicación de datos en la Web:
  - Usando las **tecnologías de la Web**: HTTP, URL/URI, DNS
  - Con un **modelo de datos explícito**: RDF + vocabularios (o esquemas, ontologías..) en RDF Schema o OWL
- Permite **compartir** (aumentar) tus datos
- Permite **reutilizar** datos de otros para enriquecer tu información, reducir costes, y ofrecer mejores servicios

# Los 4 Principios\*

1. Utilizar **URIs** para nombrar cosas (recursos)
2. Usar el protocolo **HTTP** para publicar/recuperar recursos
3. Describir datos en un **formato estándar** (RDF)
4. **Enlazar** con otros recursos a través de URIs

# Ontologías

- Son los vocabularios con los que se describen las propiedades (predicados) de las tripletas de los grafos.
- Descripción de un dominio del conocimiento, de las entidades relevantes y sus relaciones dentro del dominio.
- Con ella se crea un esquema conceptual exhaustivo que facilita la comunicación e intercambio entre distintos sistemas.
- Utilizado en inteligencia artificial y en representación del conocimiento.
- Las máquinas utilizan las ontologías con diversos fines como el razonamiento, o la clasificación.
- Se escriben utilizando determinados lenguajes como RDFs o, más propiamente, OWL.

# ¿Qué necesitamos? (1)

- Bloque básico: **tripletas RDF**

**Sujeto** → **Predicado** → **Objeto**

**Universidad Central** → **es productor de** → **Expedientes para la expedición de títulos de bachiller**

**Madrid** → **forma parte de** → **España**

- **Sujeto** y **predicado** son URIs

 **Objeto** puede ser una URI o un literal

# Bloque básico para describir el recurso: Tripletas

**Sujeto** → **Predicado** → **Objeto**

Sujeto	Predicado	Objeto
Expediente de Expedición de Título de Bachiller de José Giral	Data Tópica	Madrid
Expediente de Expedición de Título de Bachiller de José Giral	Data Crónica	1896-05-09 - 1896-05-16
Expediente de Expedición de Título de Bachiller de José Giral	Es producido por	Universidad Central
Expediente de Expedición de Título de Bachiller de José Giral	Tiene como “materia”	José Giral

# Triplificando la descripción<sup>(2)</sup>:

Los objetos pueden ser URIs o literales

Sujeto	Predicado	Objeto
<a href="http://pares.mcu.es/ParesBusquedas20/catalogo/description/7264106">http://pares.mcu.es/ParesBusquedas20/catalogo/description/7264106</a>	Data Tópica	<a href="http://pares.mcu.es/ParesBusquedas20/catalogo/autoridad/82253">http://pares.mcu.es/ParesBusquedas20/catalogo/autoridad/82253</a>
<a href="http://pares.mcu.es/ParesBusquedas20/catalogo/description/7264106">http://pares.mcu.es/ParesBusquedas20/catalogo/description/7264106</a>	Data Crónica	"1896-05-09 - 1896-05-16 "
<a href="http://pares.mcu.es/ParesBusquedas20/catalogo/description/7264106">http://pares.mcu.es/ParesBusquedas20/catalogo/description/7264106</a>	Es producido por	<a href="http://pares.mcu.es/ParesBusquedas20/catalogo/autoridad/40311">http://pares.mcu.es/ParesBusquedas20/catalogo/autoridad/40311</a>
<a href="http://pares.mcu.es/ParesBusquedas20/catalogo/description/7264106">http://pares.mcu.es/ParesBusquedas20/catalogo/description/7264106</a>	Tiene como "materia"	<a href="http://pares.mcu.es/ParesBusquedas20/catalogo/autoridad/38532">http://pares.mcu.es/ParesBusquedas20/catalogo/autoridad/38532</a>

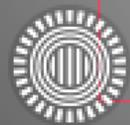
# Triplificando la descripción: Usando vocabularios

Los predicados SON URIs (de uno o varios vocabularios)

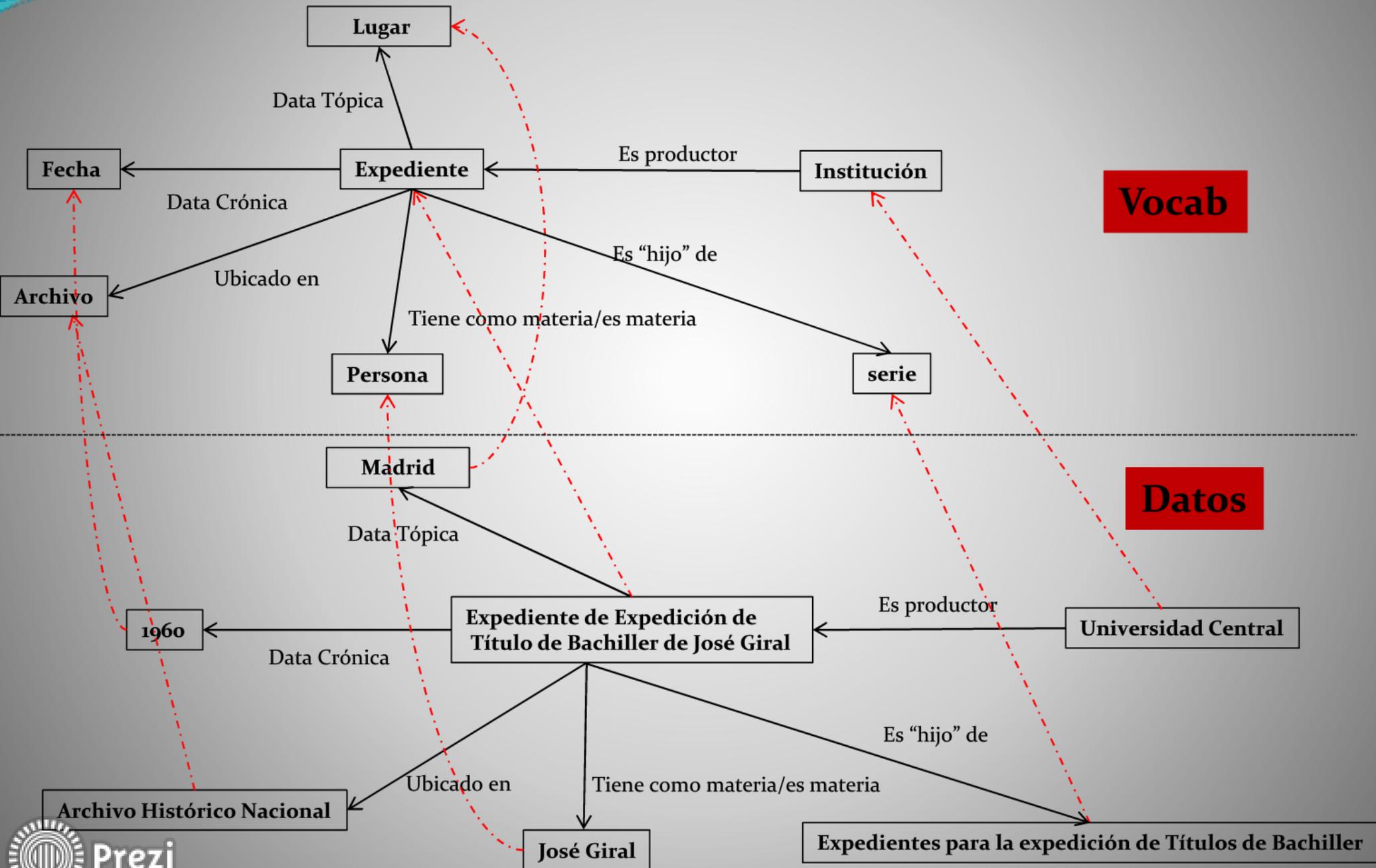
Sujeto	Predicado	Objeto
<a href="http://pares.mcu.es/ParesBusquedas20/catalago/description/7264106">http://pares.mcu.es/ParesBusquedas20/catalago/description/7264106</a>	<a href="#"><u>dc:location</u></a>	<a href="http://pares.mcu.es/ParesBusquedas20/catalago/autoridad/82253">http://pares.mcu.es/ParesBusquedas20/catalago/autoridad/82253</a>
<a href="http://pares.mcu.es/ParesBusquedas20/catalago/description/7264106">http://pares.mcu.es/ParesBusquedas20/catalago/description/7264106</a>	<a href="#"><u>dc:date</u></a>	"1896-05-09 - 1896-05-16 "
<a href="http://pares.mcu.es/ParesBusquedas20/catalago/description/7264106">http://pares.mcu.es/ParesBusquedas20/catalago/description/7264106</a>	<a href="#"><u>frbr:isProducedBy</u></a>	<a href="http://pares.mcu.es/ParesBusquedas20/catalago/autoridad/40311">http://pares.mcu.es/ParesBusquedas20/catalago/autoridad/40311</a>
<a href="http://pares.mcu.es/ParesBusquedas20/catalago/description/7264106">http://pares.mcu.es/ParesBusquedas20/catalago/description/7264106</a>	<a href="#"><u>frbr:hasAsSubject</u></a>	<a href="http://pares.mcu.es/ParesBusquedas20/catalago/autoridad/38532">http://pares.mcu.es/ParesBusquedas20/catalago/autoridad/38532</a>

En este momento los archivos no tenemos una ontología propia diseñada como tal.

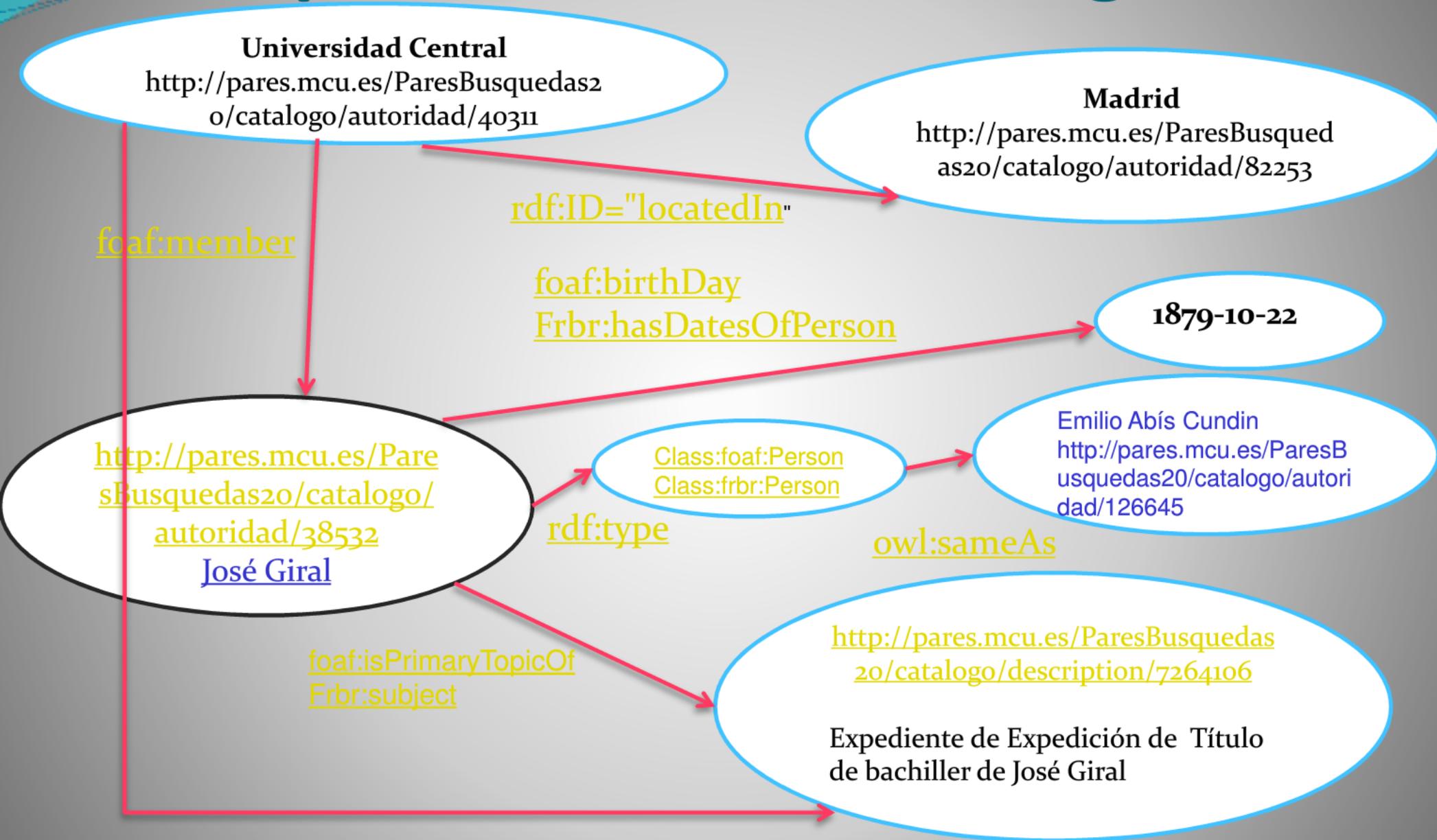
EGAD está trabajando en una ontología basada en el RIC que están poniendo en práctica en [PIAAF](#), pero no es estable todavía. PARES está a la espera de poder implementar la ontología del RIC



# El modelo (vocabulario) y los datos



# Las tripletas conforman un grafo



# Ejemplo básico de un grafo

El grafo podría interpretarse del siguiente modo:

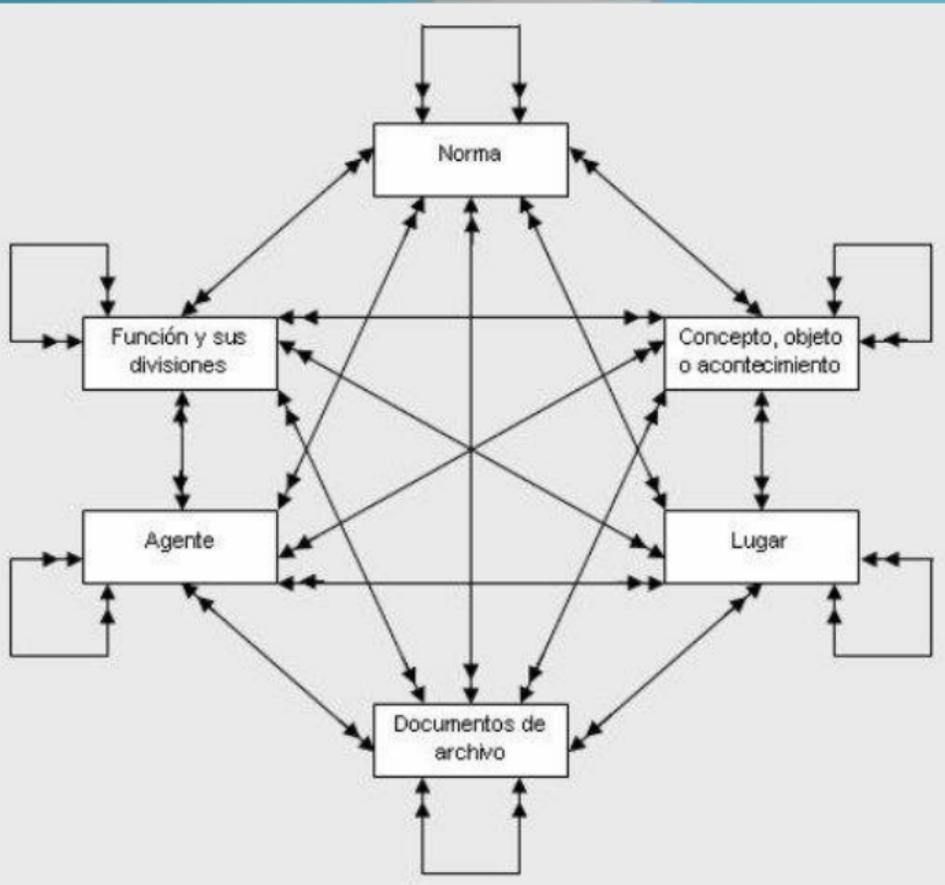
La Universidad Central, localizada en Madrid, tuvo como miembro (estudiante) a José Giral, persona, al igual que Emilio Abís Cundin.

José Giral es el tema principal del expediente producido por la Universidad Central, para la expedición de su título de Bachiller.

Las relaciones entre URIs se realizan mediante vocabularios/ontologías controladas. En este caso. Dublin Core: <http://purl.org/dc/terms> y la ontología de Dbpedia, para la foto

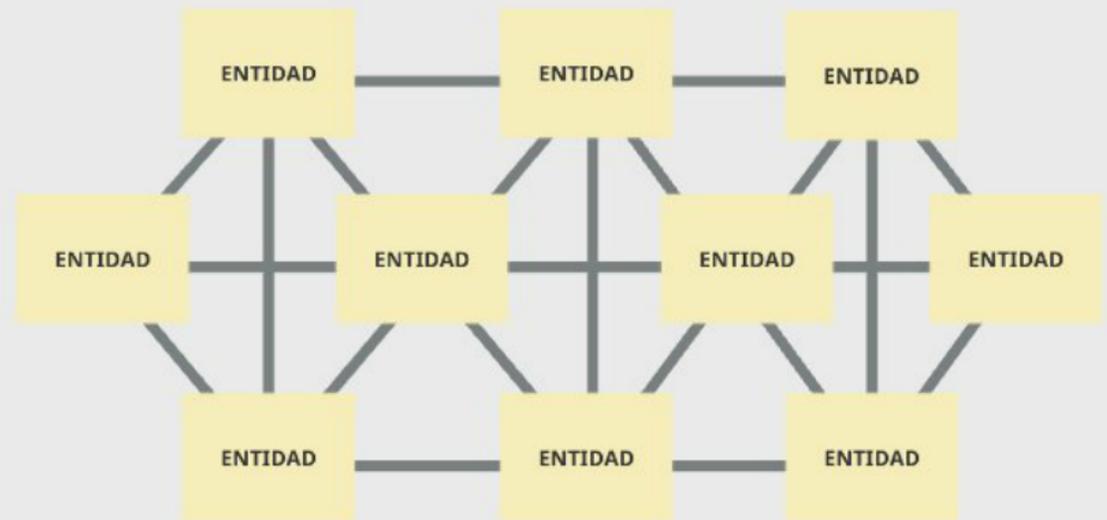
**¿Resulta aplicable a los  
archivos los modelos de web  
semántica?**

# CNEDA RIC



¿Qué se puede describir archivísticamente?

## ENTIDADES Y SUS RELACIONES



**Y....¿ para qué sirve la normalización?**

**Ejemplo práctico: la  
interoperabilidad con otros sistemas  
archivísticos: Portal Europeo de  
Archivos**

# La interoperabilidad de los archivos españoles con otros sistemas archivísticos europeos





# APEF

Assembly of Associates

Country Managers

Governing board

Content Providers

Staff: communication officer;  
developers



# APEnet – APEX – APEF

30.01.2009  
APEnet

2012  
APEX

Septiembre 2015.  
APE Foundation

# Standards



Metadata	Standard	
Finding aids holding guides	<a href="#"><u>EAD</u></a> <a href="#"><u>2002</u></a>	Encoded Archival Description. apeEAD, das APE-eigene EAD-Profil
Archival Landscape institutions	<a href="#"><u>EAG</u></a>	Encoded Archival Guide
Corporate bodies, Persons, and Families	<a href="#"><u>EAC-</u></a> <a href="#"><u>CPF</u></a>	Encoded Archival Context- Corporations, Persons, Families. apeEAC
information on digital objects	<a href="#"><u>METS</u></a>	Metadata Encoding and Transmission Standard, e.g., apeMETS
Europeana	<a href="#"><u>EDM</u></a>	Europeana Data Modell

# Europeana



[Colecciones](#) [Explorar](#) [Exposiciones](#) [Blog](#)

NUESTROS SITIOS



Subdirección General de Archivos

noviembre 2019 **JORNADA TÉCNICA**  
 Reutilización de la información del sector público en los archivos **ARCHIVOS ABIERTOS**

## Web semántica y el intercambio de datos archivísticos: casos de estudio

Ana María López Cuadrado

