



TECNOLÓGICO  
DE MONTERREY®

## **Linked Open Data: Datos abiertos para la Web Semántica**

Día virtual de Datos Abiertos - CUDI

Dr. Hector G. Ceballos ([ceballos@itesm.mx](mailto:ceballos@itesm.mx))

Marzo 2014

# + Agenda

- Interoperabilidad en Datos Abiertos
- Web Semántica - Web de Datos
- Linked Data
- The Linking Open Data Cloud
- Vocabularios (Ontologías)
- Casos de éxito
- Conclusiones

# + Interoperabilidad en Datos Abiertos

- ¿Estaremos listos para analizar un mar de datos abiertos?
- ¿La información de los diferentes datasets puede “cruzarse”?
- Y si los podemos cruzar, ¿qué conocimiento podemos generar a partir de ellos?

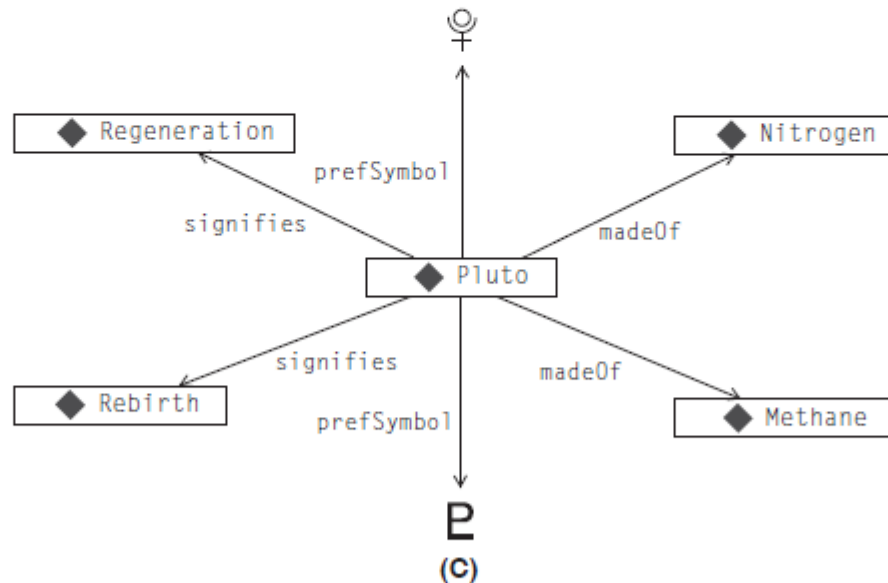


# Web Semántica

- En el año 2000, Tim Berners Lee propone una nueva versión del Web: una **Web Semántica**
  - Una web **entendible por personas y por máquinas**.
  - Una **Web de datos**, no sólo de documentos
  - Información **anotada** con etiquetas, ligas o estructuras complejas.
  - Que permitan encontrar **inequívocamente** lo que se busca.
  - Donde **agentes** realicen operaciones por nosotros.
  
- El World Wide Web Consortium (W3C).

# + Interoperabilidad

- Entendimiento común entre disciplinas y fuentes de datos



# + Aplicaciones de Tecnología Semántica en grandes organizaciones

- **Extraer conocimiento** de su propiedad intelectual.
  - Representación de conocimiento y datos no estructurados.
- Análisis de investigaciones previas para **aprender a partir de la experiencia**.
  - Administración de conocimiento.
- Proveer acceso a **memorias corporativas** para la **toma de decisiones**.
  - Organizaciones inteligentes.



# Web de Datos

- En 2013 se vuelve más realista la visión:
  - Combinando eGovernment y Semantic Web Activities.
  - Vencer la diversidad de orígenes de datos y formatos, facilitando la potencial integración y procesamiento de datos a escala Web.
  - Proveyendo estándares para formatos de intercambio de datos , modelos, herramientas y asesoría.
- Data on the Web Best Practices Working Group
  - to develop the **open data ecosystem**, to provide **guidance to publishers**, to **foster trust in the data** among developers
- CSV on the Web Working Group
  - to define mechanisms for interpreting a set of CSVs as relational data.



# Linked Data

- Principios de Linked Data
  - Nombrar todas las cosas con URIs
  - Usar URIS HTTP para que la gente pueda saber a qué se refieren.
  - Cuando alguien busque un URI, proveer información útil, usando los estándares RDF y SPARQL.
  - Incluir enlaces a otros URIs, permitiendo descubrir sus relaciones on otras cosas.
- Tom Heath, Christian Bizer. Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool. 2011.
  - <http://linkeddatabook.com>

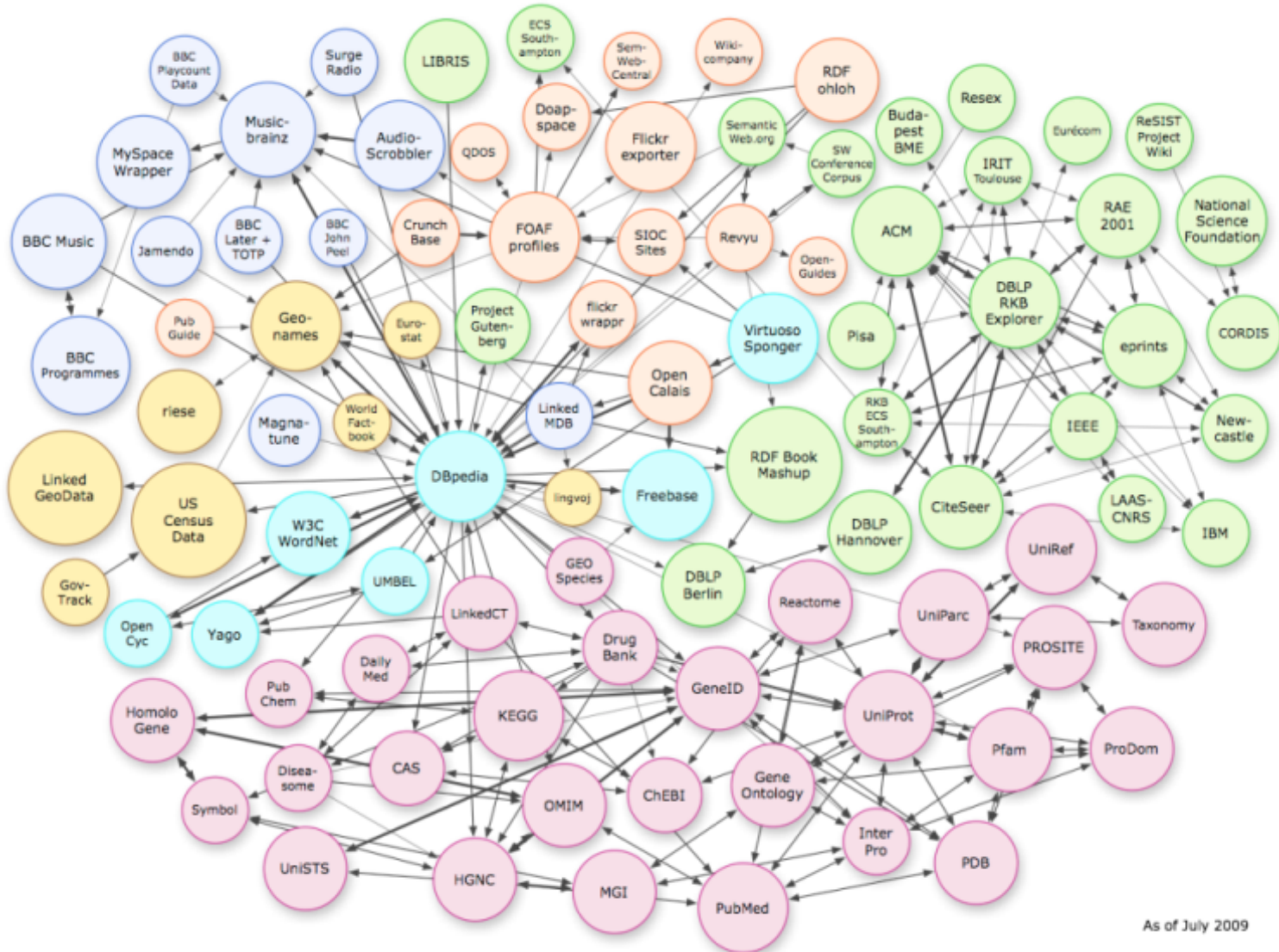




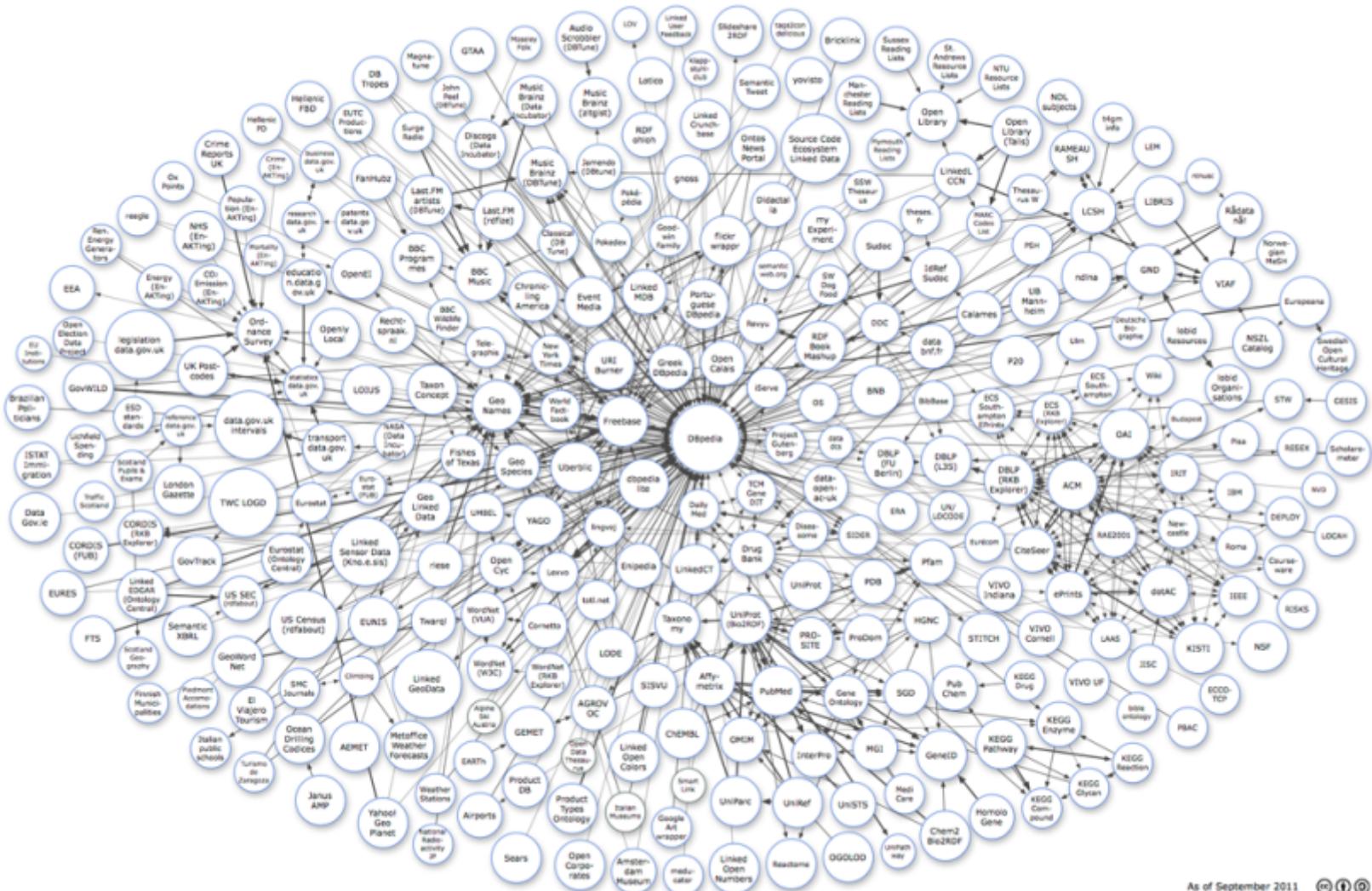
# The Linking Open Data Cloud

- Disponible en Datahub con la etiqueta LOD:
  - <http://datahub.io/dataset?tags=lod>
  - 894 de 9,088 datasets (10%)
- En formatos RDF/XML, N-Triple o Turtle.
- Algunos comparten también el SPARQL end-point.
- Publicados usando las licencias de Open Definition
  - <http://opendefinition.org/>

# + The Linking Open Data Cloud



# + En Septiembre 2011





# Principales LODs

- Mayor volumen de datos y mayor número de enlaces entrantes.
  - Enciclopedias: DBPedia / FreeBase
  - Publicaciones científicas: ACM, Citeseer, PubMed, DBLP, Ordnance Survey
  - Geo-referencias: Geonames
  - Música: BBC Music, Last.FM
  - Dominio específico: Gene Ontology

# + Vocabularios (ontologías)

- Linked Open Vocabularies
  - <http://lov.okfn.org/dataset/lov/>
- Dublin Core
  - <http://www.dublincore.org/documents/dcmi-terms/>
- SKOS Simple Knowledge Organization System
  - <http://www.w3.org/2004/02/skos/>
- FOAF Friend of a Friend
  - <http://xmlns.com/foaf/0.1/>



# + Casos de éxito (1/2)

## ■ DBPedia

- <http://dbpedia.org/>
- Información de Wikipedia
- Infoboxes, Categorías, Geo-referencias, etc.

## ■ BioPortal

- Stanford University School of Medicine
- National Center for Biomedical Ontology (2005)
- Investigación biomédica y práctica clínica
- 18.1 Millones 2010-2015



**DBpedia**  
DBpedia Blog | Get Involved | Get Help

**About / News**  
Applications  
Use Cases  
Datasets  
Online Access  
DBpedia Live  
Downloads  
Interlinking  
Development  
Support  
Publications

**News**

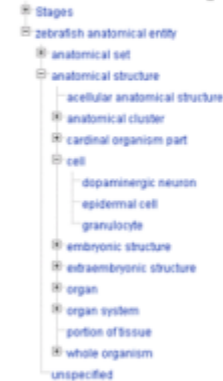
**Call for ideas and Mentors for GSoC 2014 DBpedia + Spotlight joint proposal (please contribute within the next days)**  
We started to draft a document for submission at Google Summer of Code 2014 <http://dbpedia.org/gsoc2014> We are still in need of ideas and mentors. If you have any improvements on DBpedia or DBpedia Spotlight that you would like to have done, please submit it in the ideas section now. Note that accepted GSoC students will receive about [...]

**Making sense out of the Wikipedia categories (GSoC2013)**  
(Part of our DBpedia+spotlight @ GSoC mini blog series) Mentor: Marco Fossati (@mfoos <fossati[at]ppziodati.eu> Student: Kausan Perera <kkausperera[at]gmail.com> The latest version of the DBpedia ontology has 529 classes. It is not well balanced and shows a lack of coverage in terms of encyclopedic knowledge representation. Furthermore, the current typing approach involves a costly manual mapping effort and heavily depends [...]

### Zebrafish anatomy and development

#### Tree View

Tree view constructed based on *is\_a* hierarchy



#### Class Details

General	Attributes
Class Name	cell
Database_References	ZFI
Id	CL:0000000

#### Graph View

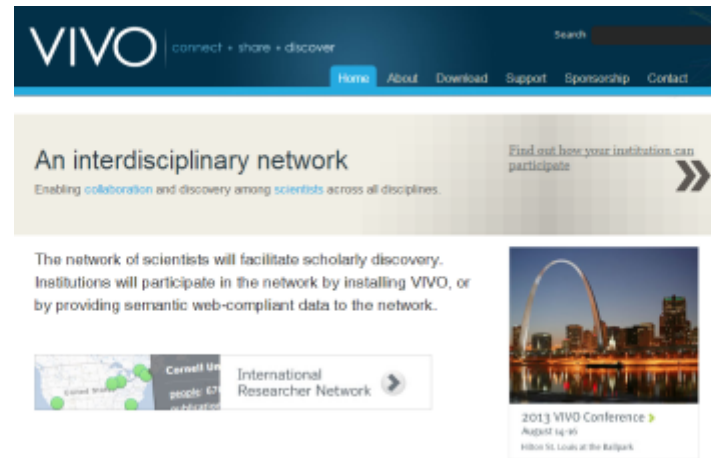
Graph Type



# + Casos de éxito (2/2)

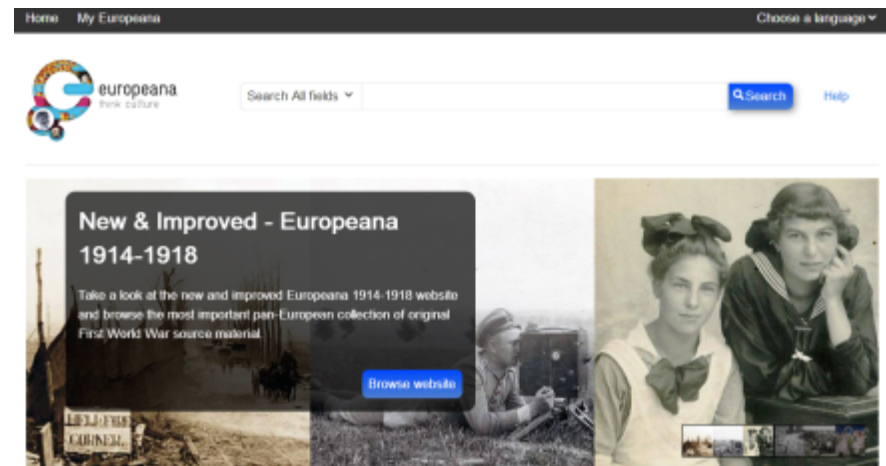
## ■ VIVO Web

- <http://www.vivoweb.org/>
- Cornell University
- Perfil científico de Universidades
- Profesores, publicaciones, proyectos, etc.



## ■ Europeana

- <http://www.europeana.eu/>
- Unión Europea
- Museos, Colecciones, Galerías
- Cuadros, Fotos, objetos, etc.





# Conclusiones

- La Web de Datos provee formatos de representación de datos que permitirán la interoperabilidad semántica entre conjuntos de datos.
- Para ello la información actual deberá adecuarse a estos nuevos formatos, usando vocabularios comunes y agregando ligas entre cosas.
- Las herramientas y plataformas actuales explotan dicha interoperabilidad ...
  - capturando el conocimiento de expertos en un dominio,
  - ofreciendo motores de búsqueda más precisos, e
  - infiriendo relaciones implícitas (razonamiento automático)





# Referencias

- Tim Berners-Lee, James Hendler and Ora Lassila. [The Semantic Web. Scientific American](#), May 2001.
- Shadbolt, N., Berners-Lee, T. and Hall, W. (2006) [The Semantic Web Revisited](#). IEEE Intelligent Systems, 21 (3). pp. 96-101. ISSN 1541-1672. W3C. Semantic Web Activity. <http://www.w3.org/standards/semanticweb/>
- OWL Premier. [www.w3.org/TR/owl2-primer/](http://www.w3.org/TR/owl2-primer/)
- Gene Phifer. Hype Cycle for Web and User Interaction Technologies, 2011. Gartner. <http://www.gartner.com/id=1752215>
- Tom Heath, Christian Bizer. Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool. 2011.