

2009 RAW DATA NOW! (*)

Alessandra Donnini

Etcware S.r.l.

(*) Tim Berners Lee http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html



... al web dei dati

Read more at Wikipedia...

WIKIPEDIA This entry is from Wikipedia, the user-contributed encyclopedia. It may not have been reviewed by professional editors and is licensed under the GNU Free Documentation License. If you find the biography content factually incorrect, defamatory or highly offensive you can edit this article at Wikipedia. Find out more about our use of this data.

Links & Information

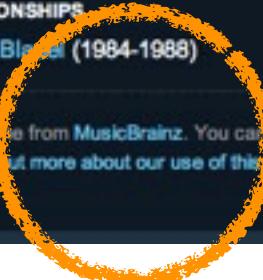
LINKS

- Official homepage at eltonjohn.com
- MySpace at myspace.com/eltonjohn
- Wikipedia article on Elton John
- IMDb at imdb.com/name/nm0005056
- Last.fm page on Elton John
- Discogs at [discogs.com/artist/Elton John](http://discogs.com/artist/Elton+John)
- MusicBrainz entry on Elton John

PERSONAL RELATIONSHIPS

Married to Renate Blaauw (1984-1988)

Links & information come from MusicBrainz. You can add or edit information about Elton John at musicbrainz.org. Find out more about our use of this data. The BBC is not responsible for the content of external sites.



Ken Bruce
BBC Radio 2

Alex Lester
BBC Radio 2

The Chris Evans Breakfast Show
BBC Radio 2

Jeremy Vine
BBC Radio 2

Richard Allinson
BBC Radio 2

Get It On With Bryan Burnett
BBC Radio Scotland

Information displayed about artists played on BBC programmes is incomplete at present. Find out more about this artist play count information.

Ci sono sempre più dati su web

- Pubblica Amministrazione (open data)
- Salute
- Conoscenza generale
- Ristoranti, alberghi
- Musica



Dire *dati* ... non è abbastanza

- I dati devono essere pubblicati su web
- I dati devono essere collegati fra loro
- I dati devono essere integrati con il web
- Le tecnologie del web semantico devono essere al servizio di questo web dei dati.





Standard

RDF Resource Description Framework

URI Uniform Resource Identifier (URL+URN)



Più standard!



- Sistemi di metadati:
 - dublin core
 - FOAF
- SKOS (Simple Knowledge Organization System)
- RDFS (RDF Schema)
- OWL (Web Ontology Language)
- SPARQL (Simple Protocol And RDF Query Language)



7

Il web dei dati!



Il web dei dati!

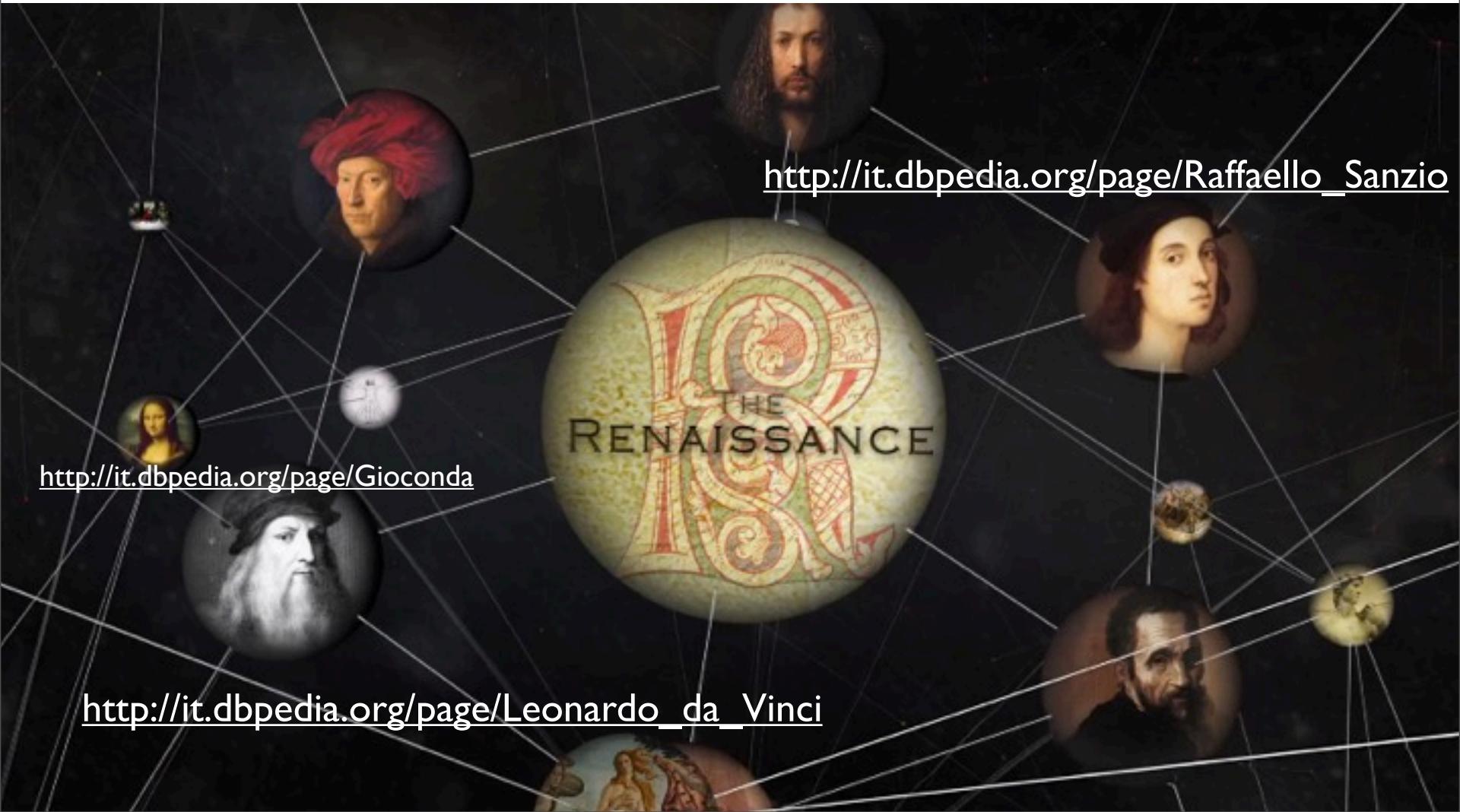


http://it.dbpedia.org/page/Raffaello_Sanzio

Il web dei dati!

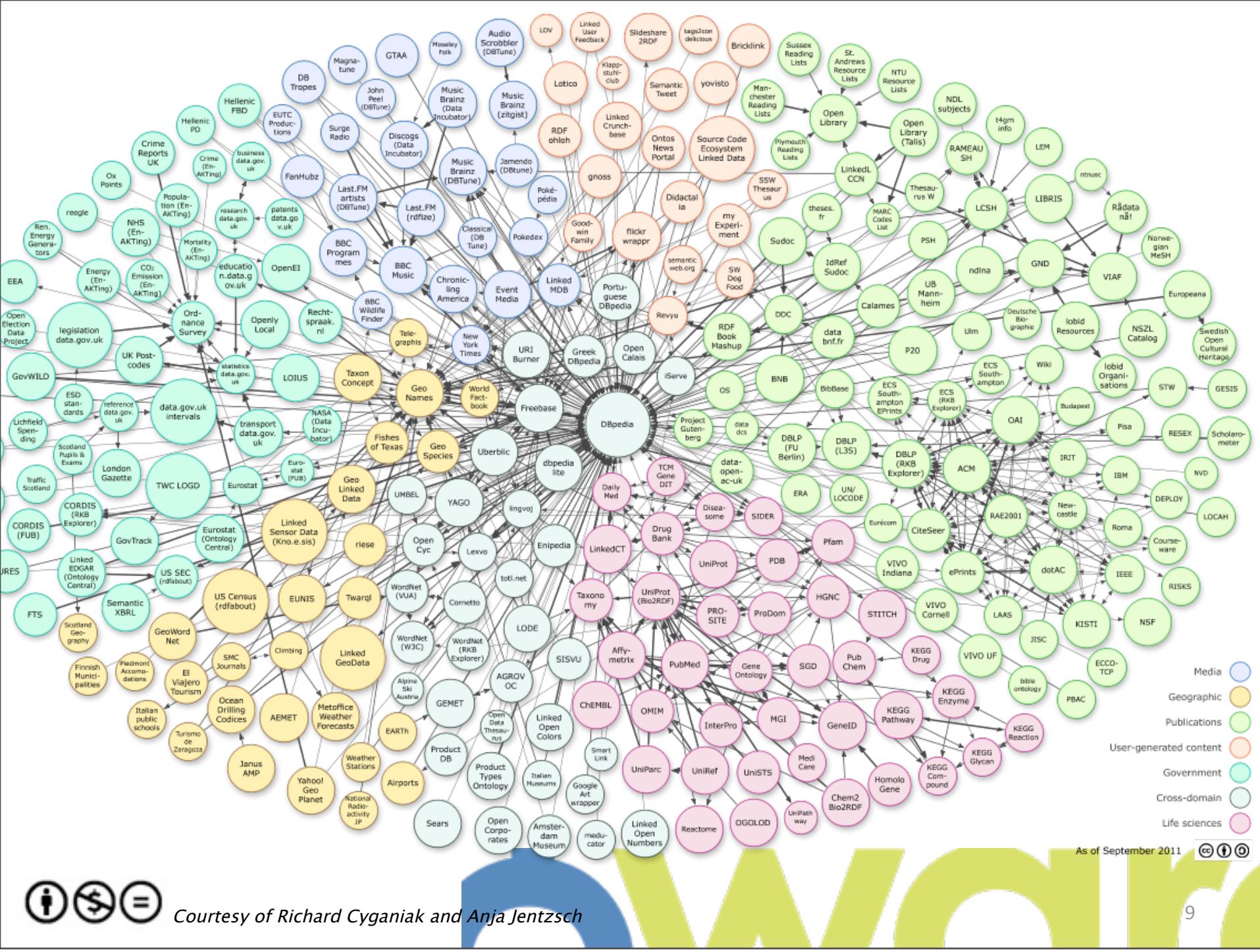


Il web dei dati!



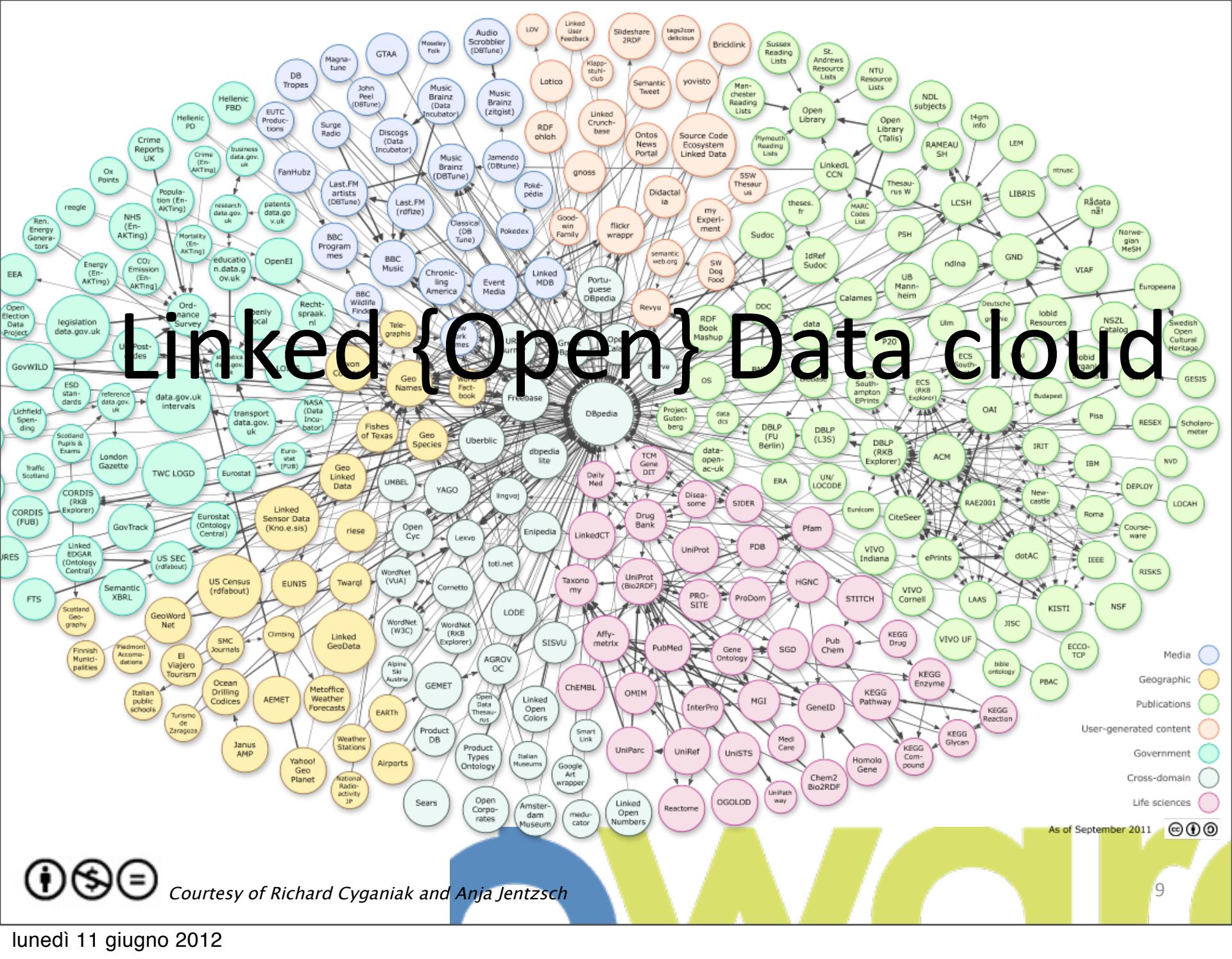
Il web dei dati!





Courtesy of Richard Cyganiak and Anja Jentzsch

linked{open}Data cloud



As of September 2011



Courtesy of Richard Cyganiak and Anja Jentzsch

Linked Open Data



- Metti i dati su web con URI
- Usa web URI
- In corrispondenza dell'URI metti altre informazioni
- Lega i dati fra loro

Usa RDF!



URI

- URIs (Unique Resource Identifiers) sono usate per identificare le cose (chiamate anche entities) nel mondo reale
 - Per esempio: persone, luoghi, eventi, aziende, enti, prodotti, film, etc (ware?)...
- Google - The knowledge network: things and not strings!



11

Come preparano i dati



oaware

Soluzione 1: “Lift”

- Tecniche di HTML scraping e natural language processing (NLP) per estrarre informazioni semantiche da siti esistenti.
- Soluzioni correnti: OpenCalais, Zemanta, Apache Stanbol
- Pro: Nessuna necessità di modificare i contenuti esistenti
- Con: Necessità di verifica



Esempio: dbpedia



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Toolbox

Print/export

Languages

Cesky
Deutsch
Espanol
Euskara
Francais
Magyar
Portugues
Pycckik

Log in Create account
Article Talk Read Edit View history Search

Liferay

From Wikipedia, the free encyclopedia

Liferay Portal is a free and open source enterprise portal written in Java and distributed under the GNU Lesser General Public License^[2] and proprietary licenses.^[1] It is primarily used to power corporate intranets and extranets.^[citation needed]

Liferay Portal allows users to set up features common to websites. It is fundamentally constructed of functional units called portlets.^[3] Liferay is sometimes described as a content management framework or a web application framework. Liferay's support for plugins extends into multiple programming languages, including support for PHP and Ruby portlets.^[4]

Although Liferay offers a sophisticated programming interface for developers, no programming skills are required for basic website installation and administration.

Liferay Portal is Java based and runs on any computing platform capable of running the Java Runtime Environment and an application server. Liferay is available bundled with a servlet container such as Apache Tomcat.^[5]

Liferay	
Screenshot	(show)
Developer(s)	Liferay, Inc
Stable release	6.1 (6.1 GAI) / January 6, 2012; 4 months ago
Development status	Active
Written in	Java
Operating system	Cross-platform
Type	enterprise portal, content management framework, Content management system, Community
License	LGPL License [1] ^[2] and proprietary ^[3]
Website	http://www.liferay.com

About:Liferay

An Entity of type : [software](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia](#)

Liferay Portal is a free and open source enterprise portal written in Java and distributed under the GNU Lesser General Public License. It is primarily used to power corporate intranets and extranets. Liferay Portal allows users to set up features common to websites. It is fundamentally constructed of functional units called portlets. Liferay is sometimes described as a content management framework or a web application framework.

Property	Value
dbpedia-owl:abstract	<ul style="list-style-type: none">Liferay Portal ist eine lizenzkostenfreie Open-Source-Software, die in Unternehmen als Mitarbeiter- und Prozess-orientiertes Enterprise Portal eingesetzt wird. Mit Liferay können Informationen, Daten und Anwendungen unter einer einheitlichen Bedienoberfläche im Browser des Mitarbeiters vereinigt und personalisiert werden, um dadurch die Geschäftsprozesse im Unternehmen elektronisch zu unterstützen. Liferay Portal ist in Java entwickelt.Liferay es un portal de gestión de contenidos de código abierto escrito en Java. Se creó en 2000 en principio como solución para las organizaciones sin ánimo de lucro.Liferay Portal is a free and open source enterprise portal written in Java and distributed under the GNU Lesser General Public License. It is primarily used to power corporate intranets and extranets. Liferay Portal allows users to set up features common to websites. It is fundamentally constructed of functional units called portlets. Liferay is sometimes described as a content management framework or a web application framework. Liferay's support for plugins extends into multiple programming languages, including support for PHP and Ruby portlets. Although Liferay offers a sophisticated programming interface for developers, no programming skills are required for basic website installation and administration. Liferay Portal is Java based and runs on any computing platform capable of running the Java Runtime Environment and an application server. Liferay is available bundled with an application server such as Apache Tomcat.Liferay, Inc. é uma companhia estadunidense que desenvolve o Liferay Portal, um gerenciador de conteúdo profissional, de código-fonte aberto e que fornece documentação livre para os usuários. Além da distribuição livre do Liferay Portal, desenvolve soluções pagas, personalizadas, baseadas na versão Enterprise Edition. A empresa está localizada em Los Angeles, Califórnia, Estados Unidos.Liferay est un portal open source de gestion de contenu écrit en Java et créé en 2000 à l'origine pour une église américaine.
dbpedia-owl:genre	<ul style="list-style-type: none">dbpedia:Content_management_systemdbpedia:Enterprise_portaldbpedia:Communitydbpedia:List_of_content_management_frameworks
dbpedia-owl:latestReleaseVersion	<ul style="list-style-type: none">6.0.6 (6.0 GA4)
dbpedia-owl:license	<ul style="list-style-type: none">dbpedia:GNU_Lesser_General_Public_License
dbpedia-owl:operatingSystem	<ul style="list-style-type: none">dbpedia:Cross-platform
dbpedia-owl:programmingLanguage	<ul style="list-style-type: none">dbpedia:Java_(programming_language)
dbpedia-owl:status	<ul style="list-style-type: none">Active
dbpedia-owl:wikiPageExternalLink	<ul style="list-style-type: none">http://issues.liferay.com/secure/ReleaseNote.jspa?version=1065&styleName=Html&projectId=1095



Soluzione 2: export

- DB to RDF: <http://www.w3.org/wiki/Rdb2RdfXG/StateOfTheArt>
 1. [D2RQ](#)
 2. [Triplify](#)
 3. [SquirrelRDF](#)
 4. [Relational.OWL](#)
 5. Protege plugins
 - [DataMaster](#)
 - [OntoBase](#)
 - [DataGenie](#)
 6. [Virtuoso's RDBMS to RDF Meta Schema Language](#)
 - [Semantic Discovery Systems \(Main Site\)](#)
 - [Semantic Discovery Systems \("Trailer"/Summary Web site\)](#)
 7. [Igraps](#) mentioned in the mailinglist (out of scope)
 8. [Protege Excel Import](#)(out of scope)



Soluzione 3: riuso

- Linked Open Data: (grandi) repository di dati espressi usando il modello RDF sono disponibili su web (gratis e a pagamento)
- Deve essere definita l'interoperabilità fra questi repository, ovvero il mapping fra i dati (ad esempio mediante ontologie).



16

Dove si mettono i dati?



owcar

Dove si mettono i dati?



The Telegraph

HOME NEWS WORLD SPORT FINANCE COMMENT BLOGS CULTURE TRAVEL LIFE FAB
Film Music Art Books TV and Radio Theatre Hay Festival Dance Opera Photograph

Oscars Film Reviews Cinema Trailers Coming Soon Talking Movies Interviews DVDs

HOME > CULTURE > FILM > FILM REVIEWS

Oscars 2012: The Artist, review

The Artist, an utterly beguiling silent, black-and-white celebration of early Hollywood won Best Picture at the Oscars 2012.

★★★★★



Sulle pagine HTML



oware

17

Oscars 2012: The Artist, review – Telegraph

Source View

<http://www.telegraph.co.uk/culture/film/filmreviews/8982558/Oscars-2012-The-Artist-review.html>

Search - enhanced by Google

Monday 09 April 2012

Dating Offers Jobs

The Telegraph

CULTURE TRAVEL LIFESTYLE FASHION TECH

Books Music TV and Radio Theatre Hay Festival Dance Opera Photography Comedy Video In the Know

Coming Soon Talking Movies Interviews DVDs Film Life Film Video

2012: The Artist, review

Oscars 2012

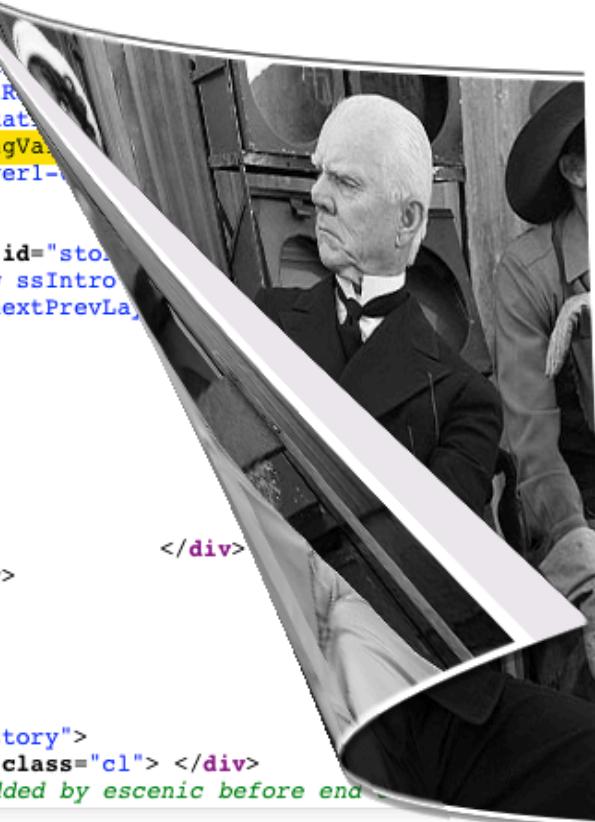
The Artist, an utterly beguiling silent, black-and-white celebration of early Hollywood won Best Picture at the Oscars

Hollywood won Best Picture at the Oscar

Rating

Art Intro

Story



To attend an Info Session or Masterclass Or download a brochure

Click here

INSEAD

The Business School for the World®

Global Executive MBA

TELEGRAPH TICKETS



```
<div class="oneHalf gutter">
    <div class="story">
        <div class="c1"> </div>
    <!-- remove the whitespace added by escenic before end -->
</div>
```

Ancora standard ...



- I dati sono nascosti nella pagina HTML e ...
- Google, Bing, Yahoo!, Yandex ed altri usano questi dati!
- Ci sono due modalità che stanno emergendo a livello di standard:
 - RDFa (HTML5, XHTML e XML in generale)
 - microdata (HTML5)



19

... c'è un grande incremento nell'uso di RDFa [...] oltre 7% delle pagine web contengono microdata.

[Mail from Peter Mika, Yahoo!](#)

*Based on a crawl evaluation by P. Mika and T. Potter
LDOW2012 Workshop, April 2012, Lyon, France*

... le pagine web che contengono dati strutturati sono aumentate del 6% nel 2010 e del 12% nel 2012.

[Hannes Mühleisen and Christian Bizer](#)

*Web Data Commons—Extracting Structured Data from Two Large Web Corpora,
LDOW2012 Workshop, April 2012, Lyon, France*

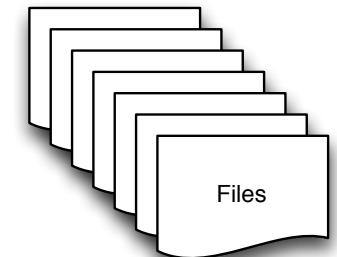
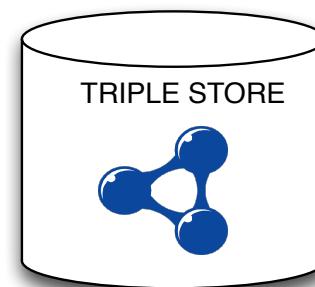


Dove si mettono i dati?



owcar

Dove si mettono i dati?



BACKEND



21

Come si accede ai dati?

- RDF
- Applicazioni che erogano dati con interfacce RESTful:

[http://www.skosware.it/rest/scheme/labc/
concept/searchFull/zucchero*.xml?
searchType=SUFFIX&language=it](http://www.skosware.it/rest/scheme/labc/concept/searchFull/zucchero*.xml?searchType=SUFFIX&language=it)

(SKOSware è una Linked data platform)





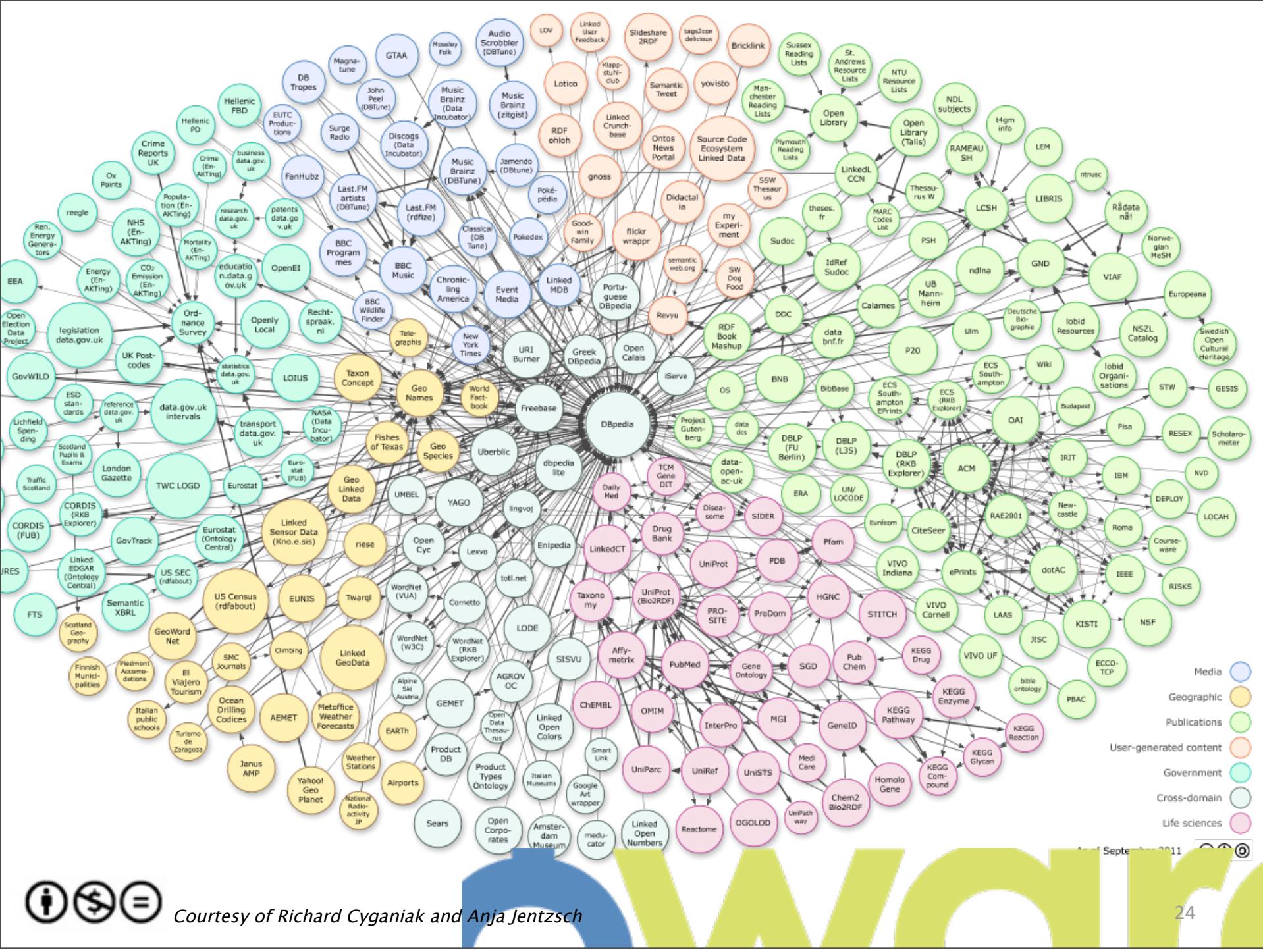
Linked data platform working group(*)

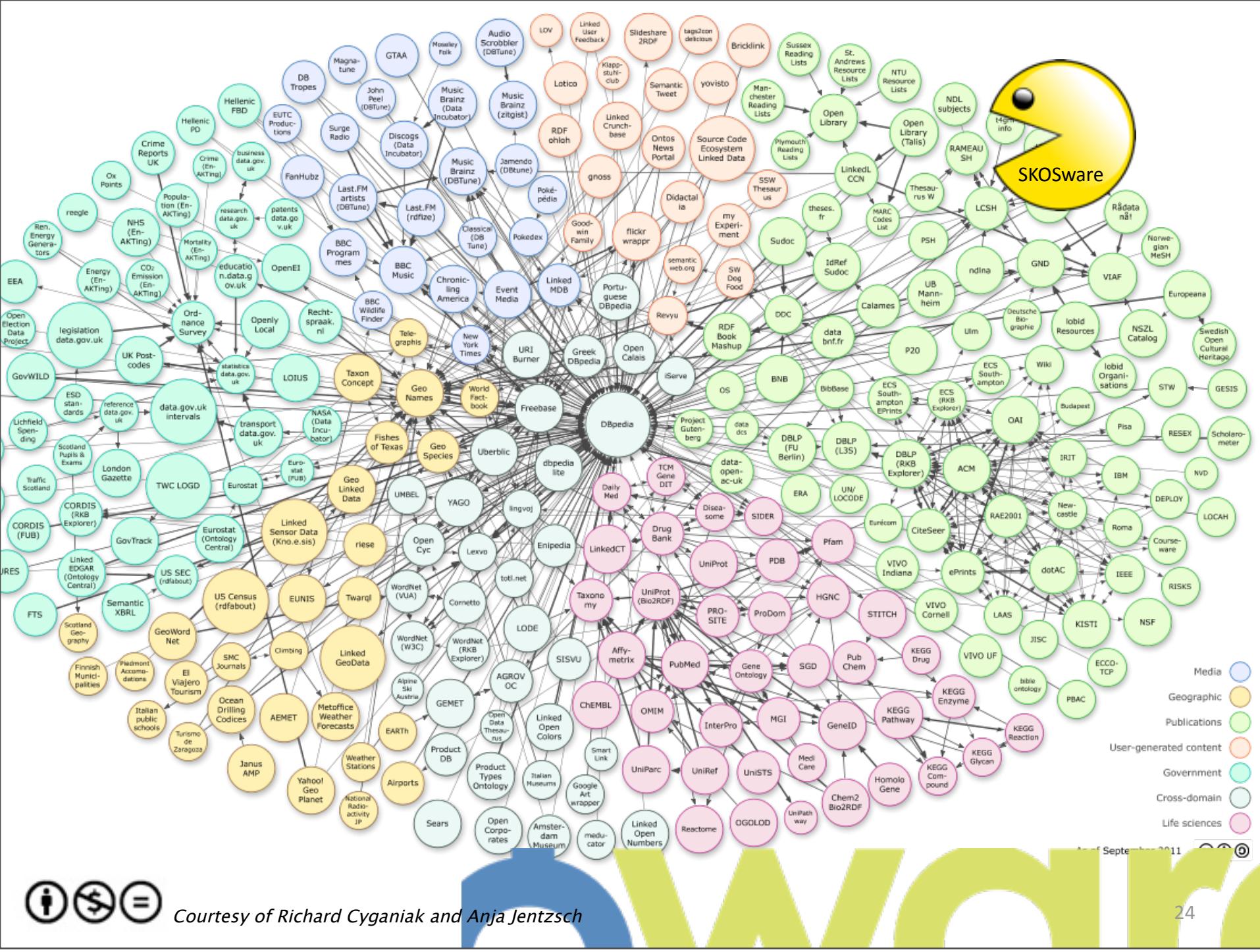
- Il LDP WG ha cominciato a lavorare questo mese e ha come obiettivo di finire nel 2014.
- Obiettivi: definire gli standard per le piattaforme di erogazione dati:
 - W3C Recommendation(s) defining a platform based on RESTful read/write Linked Data

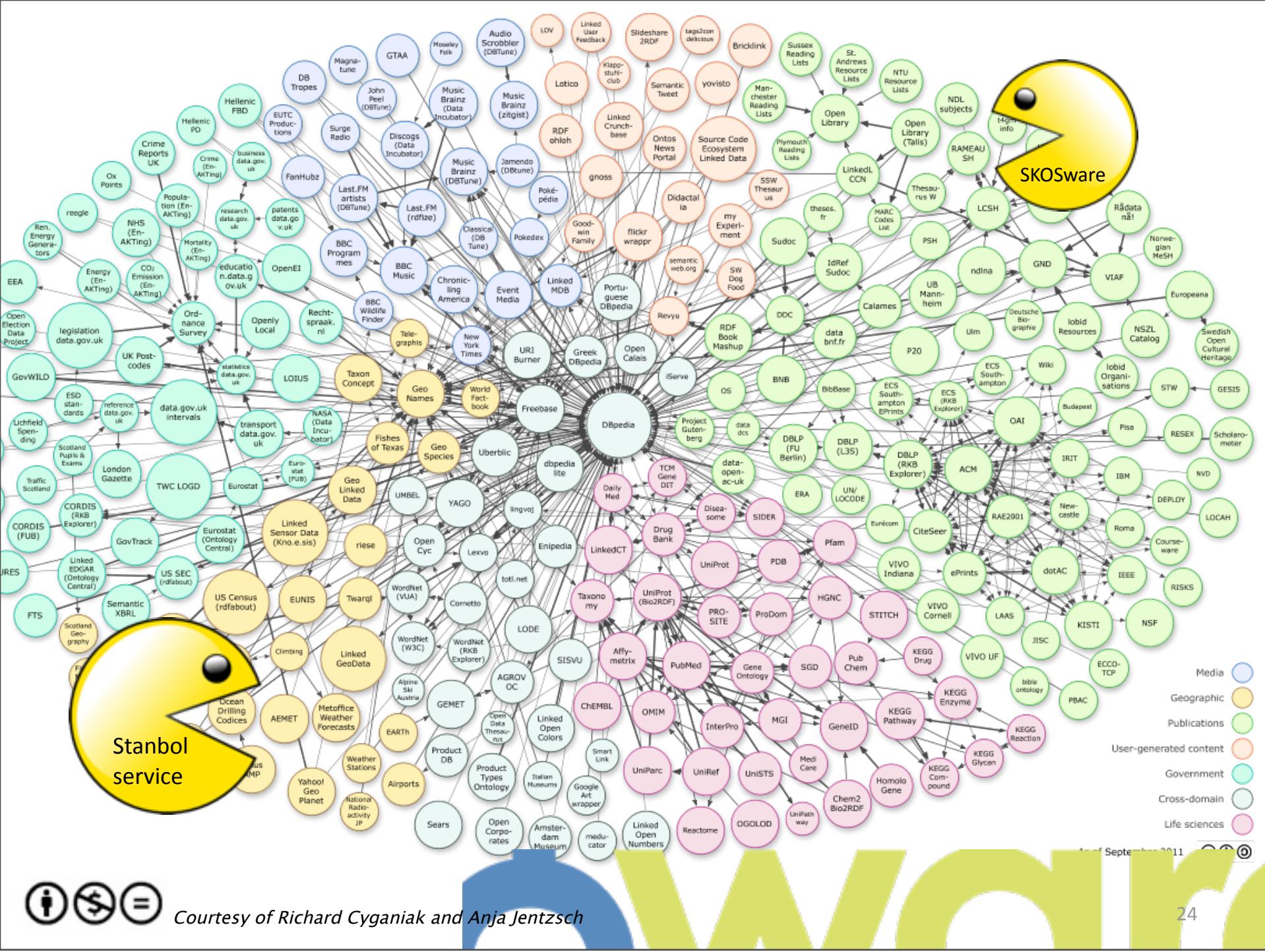
(*) <http://www.w3.org/2012/ldp/charter>

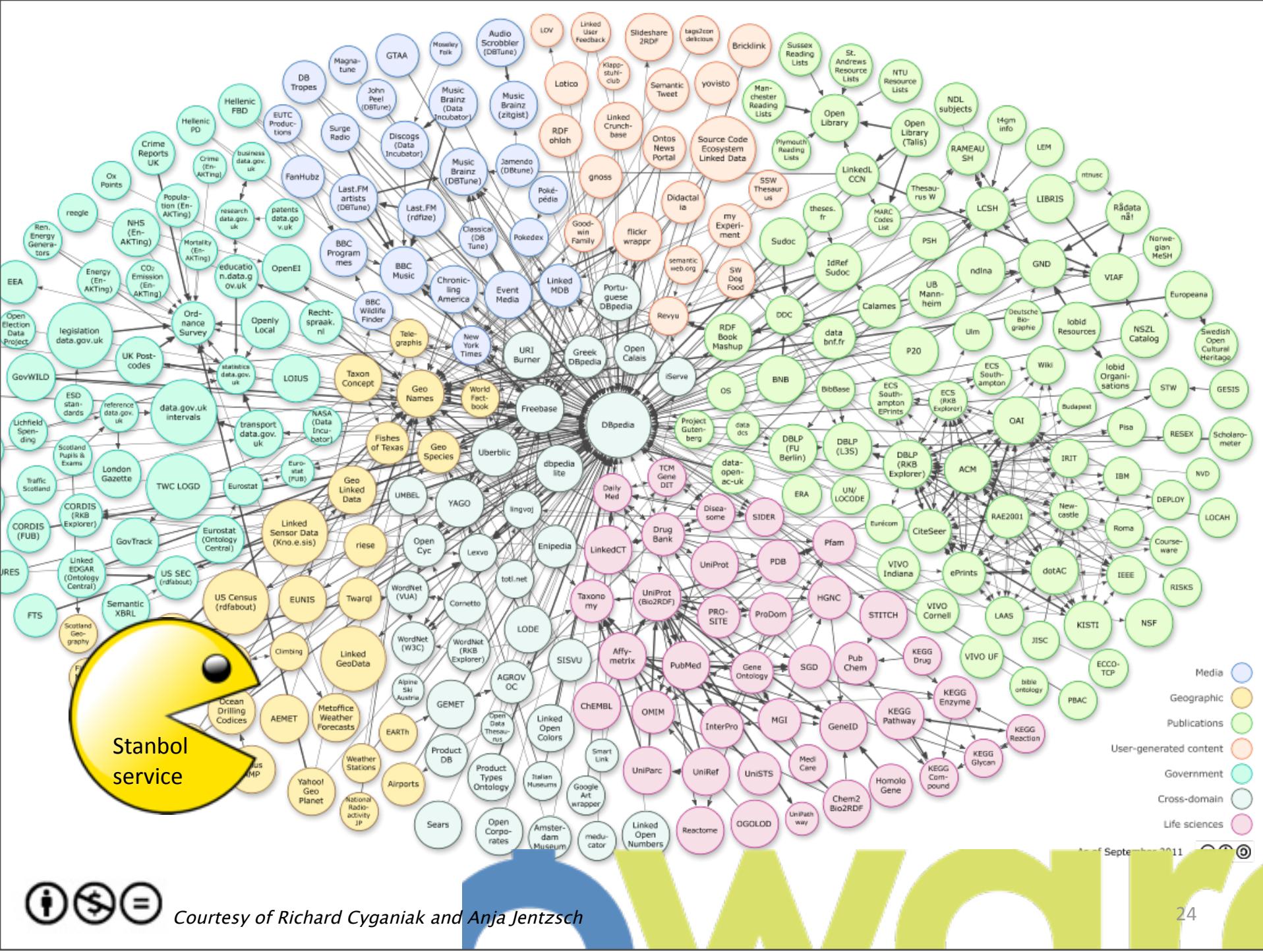


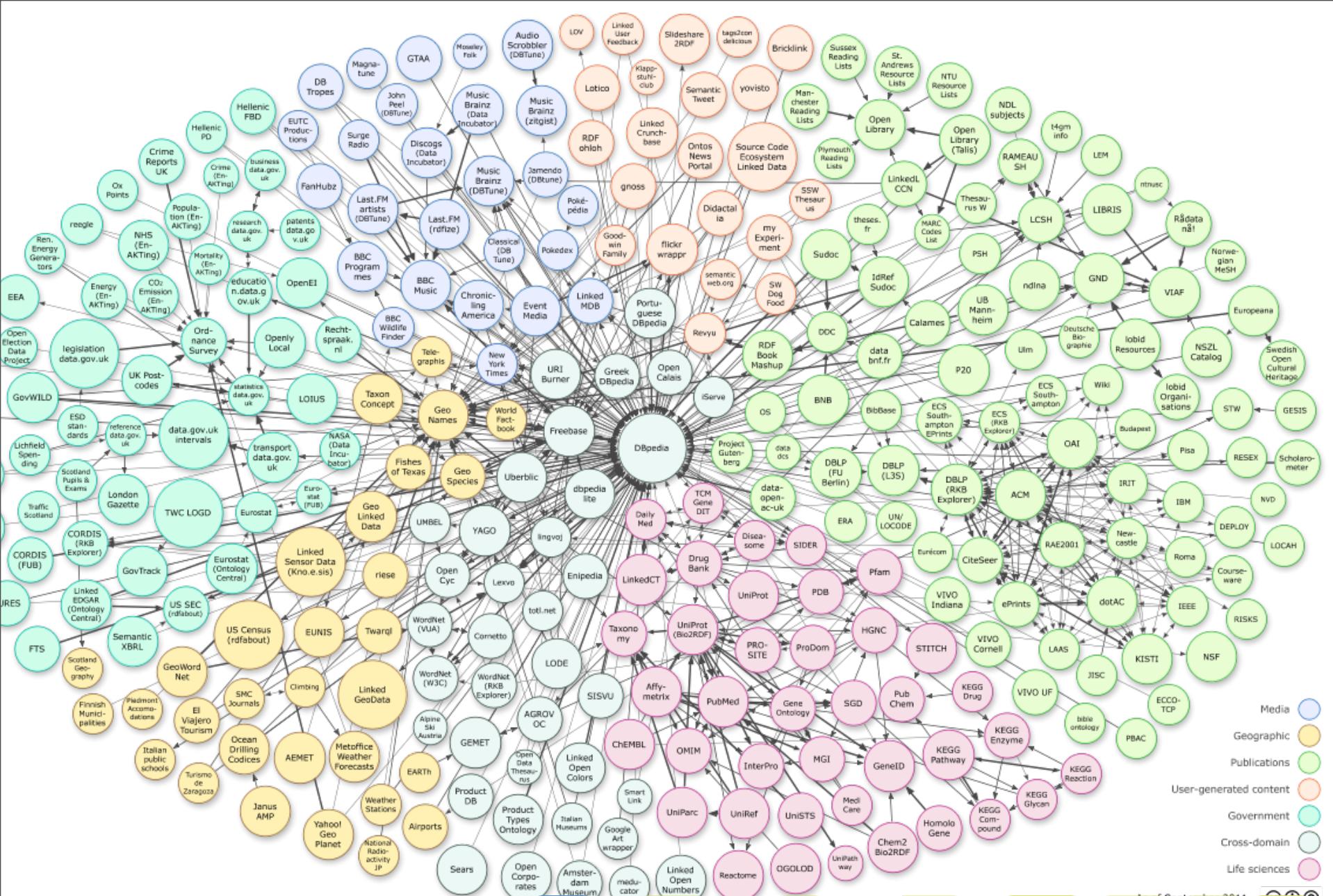
23











Come si fa a partecipare?

Trovare le parole

Trovare le metafore

Reperire fonti esistenti
(sistemi di metadati,
ontologie, tesauri)

Definire sistemi di metadati
(SKOS)

Definire ontologie

Raccontare
la storia

Produrre i dati in formato
standard

Arricchire il corpus
documentale con metadati
in formato standard



divar

DOMANDE?





dott. Alessandra Donnini

etcware S.r.l.
via Etna 13, 00141 Roma
tel./fax +39 06 64495131
<http://www.etcware.it>
a.donnini@etcware.it
cell. +39 333 8914865

