Linked Data

Semantic Universe Webcast
April 21, 2010
David Wood

2010 Semantic Technology Conference

JUNE 21 - 25 SAN FRANCISCO, CA

Linked Enterprise Data Tutorial

Tuesday, June 22, 2010

08:30 AM - 12:00 PM

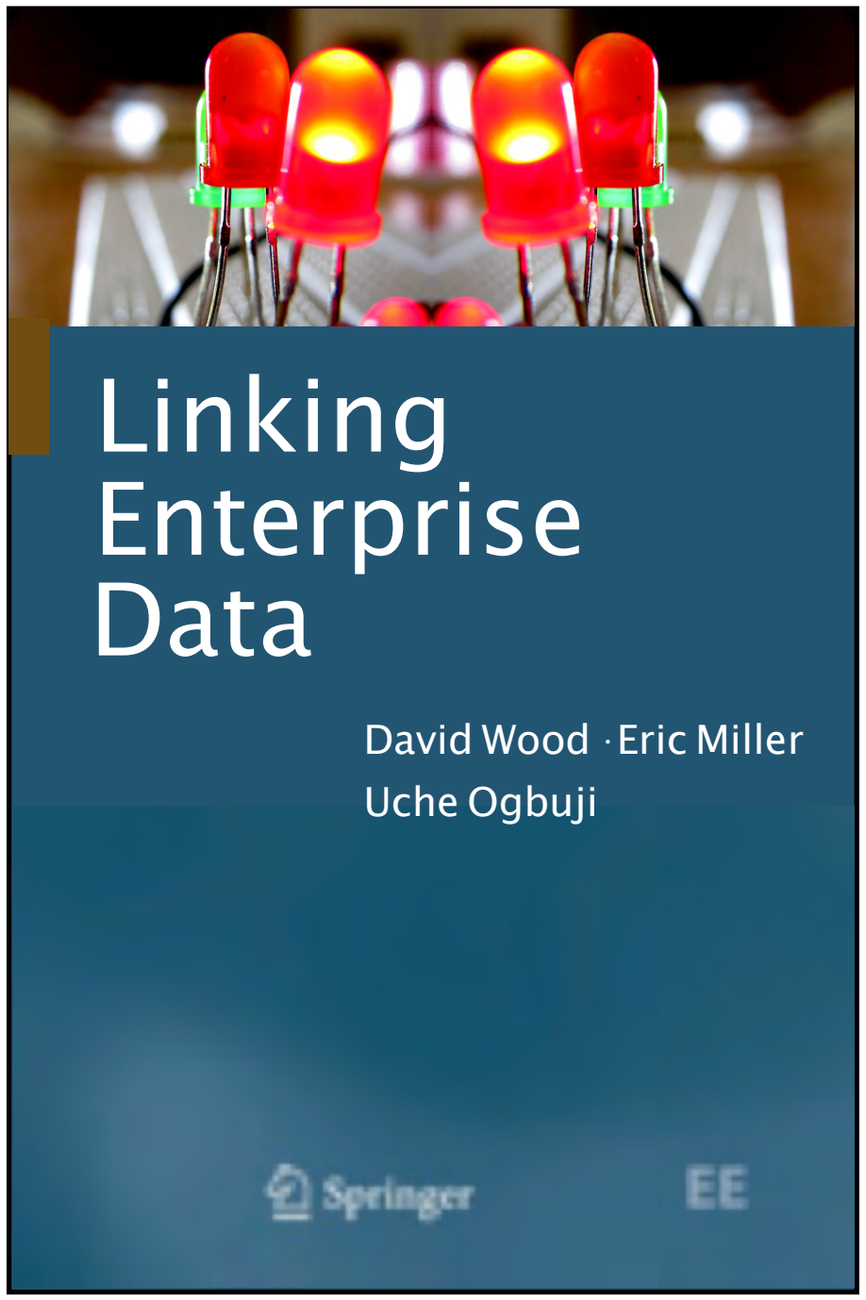
Level: Technical - Introductory

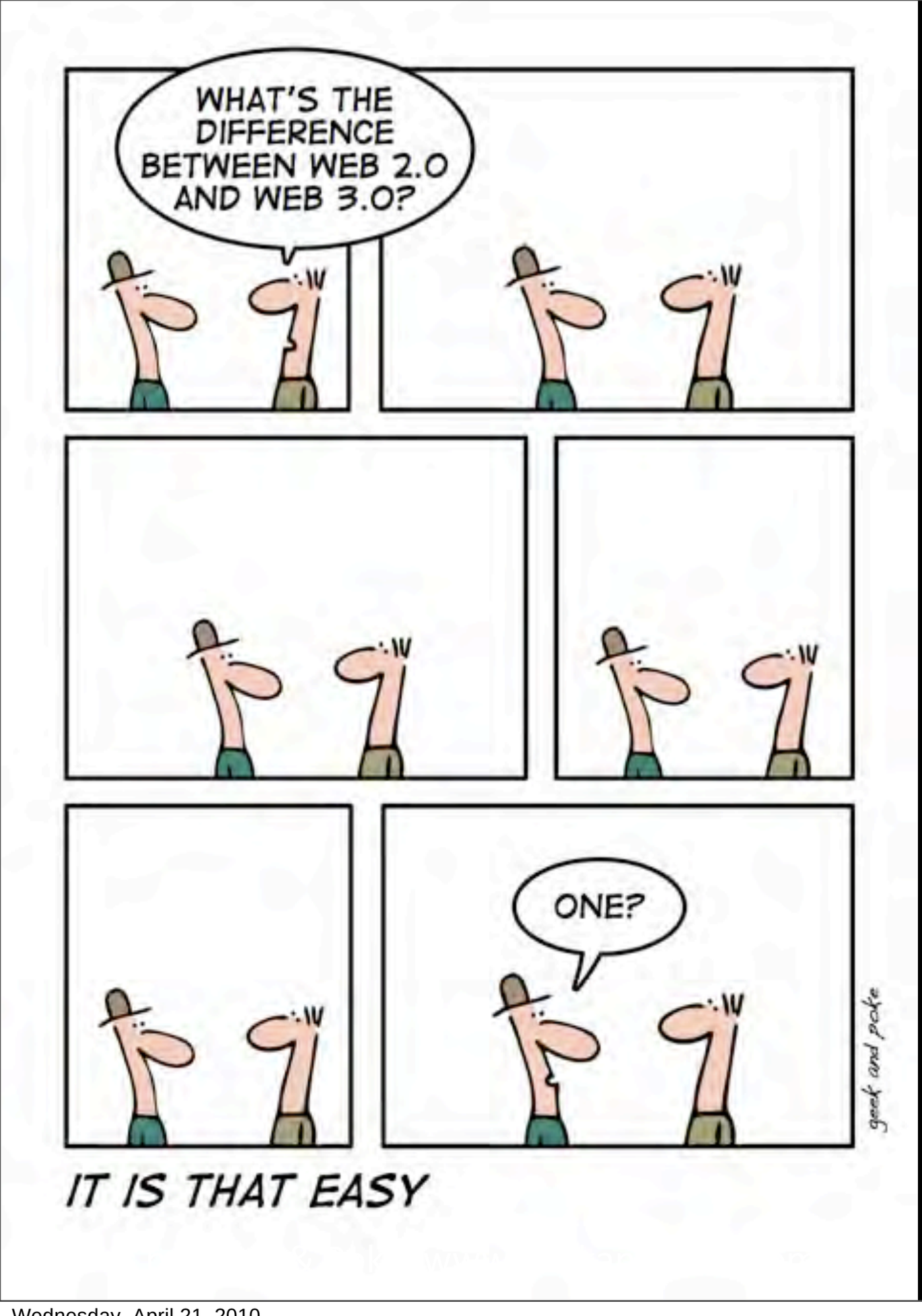
David Wood,

Bernadette Hyland

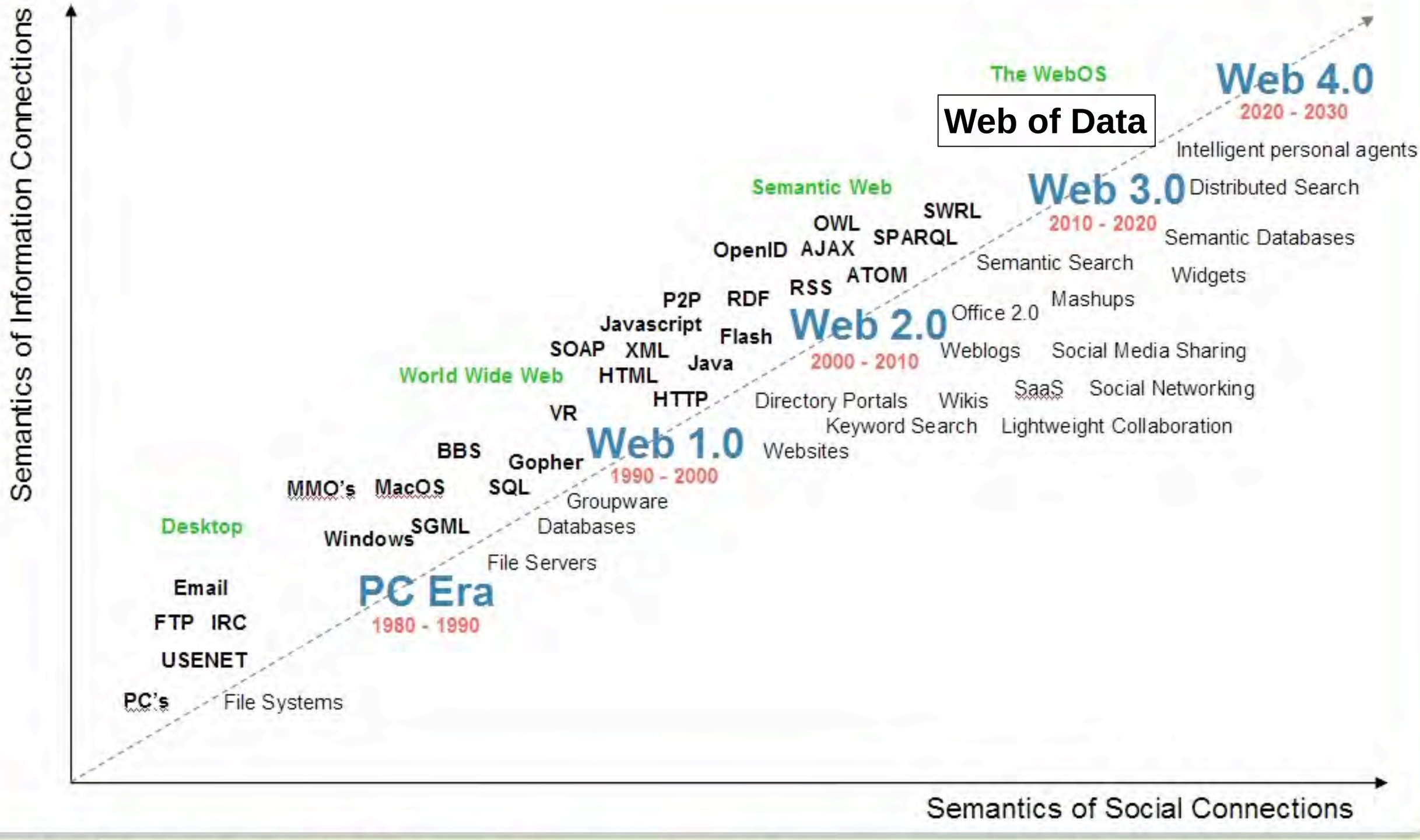
Springer will be publishing a book of our collected research in early 2011.

Contributions include a dozen detailed case studies regarding real-world uses of Linked Data techniques in enterprise settings.





Linked Data in Context: Web 3.0 and the Semantic Web

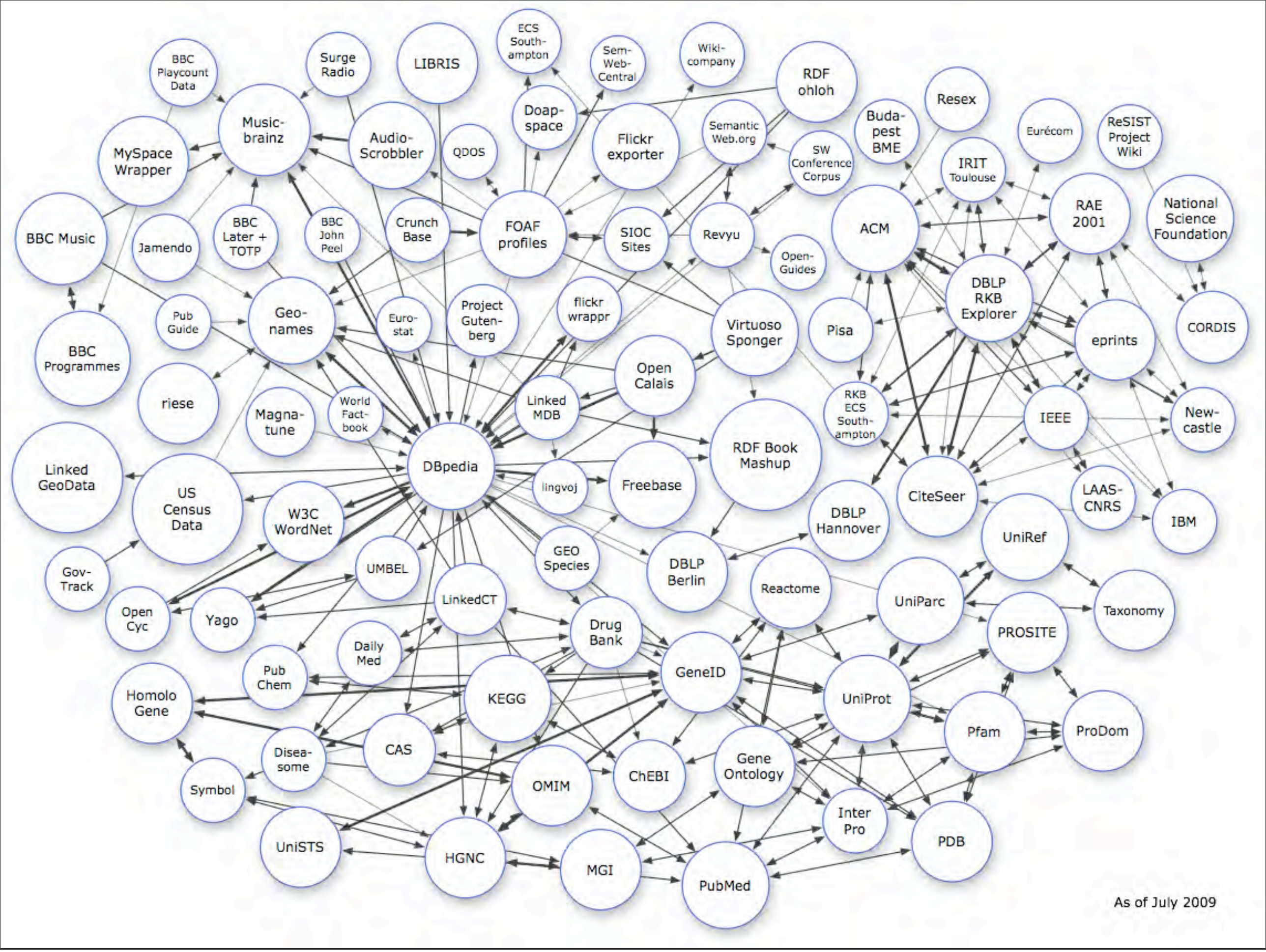


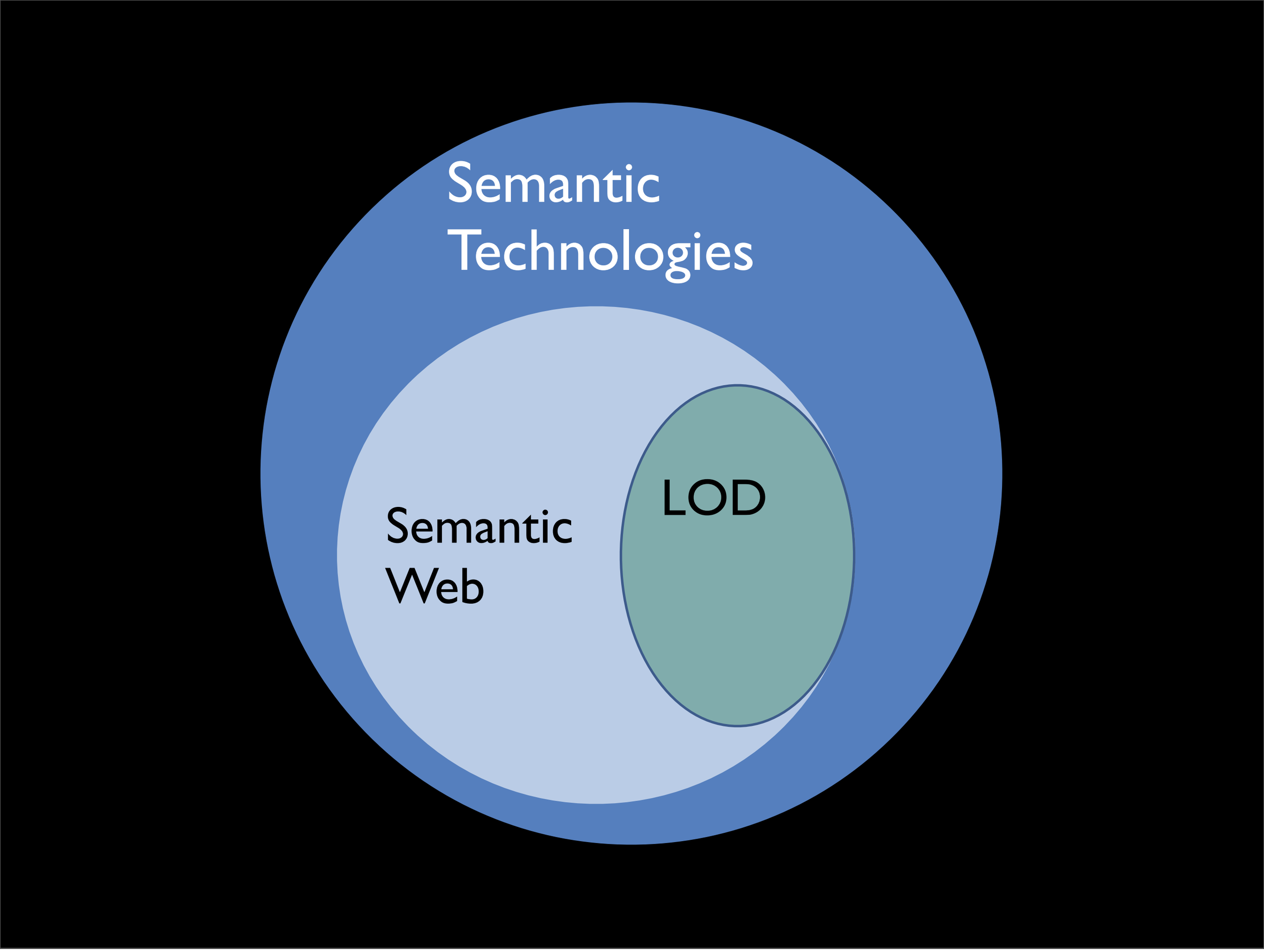
Source: Radar Networks & Nova Spivack, 2007 – www.radarnetworks.com

The Web is about links ...

The Semantic Web is about the relationships implicit in those links

- Dan Brickley

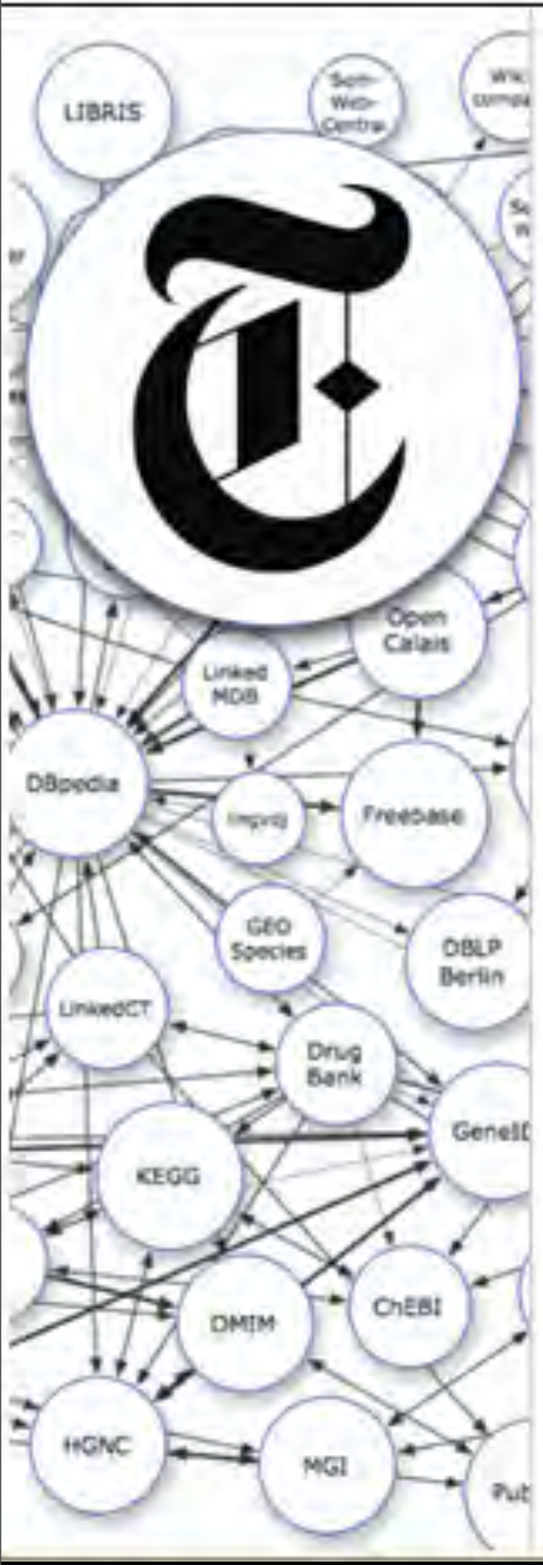

BBC Playcount Data
Surge Radio
LIBRIS
ECS South-ampton
Sem-Web-Central
Wiki-company
RDF ohloh
Resex
Buda-pest BME
Eurécom
ReSIST Project Wiki
Music-brainz
MySpace Wrapper
Audio-Scrobbler
QDOS
Doap-space
Flickr exporter
Semantic Web.org
SW Conference Corpus
IRIT Toulouse
RAE 2001
National Science Foundation
BBC Music
Jamendo
BBC Later + TOTP
BBC John Peel
Crunch Base
FOAF profiles
SIOC Sites
Revyu
Open-Guides
ACM
DBLP RKB Explorer
CORDIS
BBC Programmes
Pub Guide
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Virtuoso Sponger
Pisa
eprints
Open Calais
RKB ECS South-ampton
New-castle
riese
Magna-tune
World Fact-book
Linked MDB
IEEE
Linked GeoData
DBpedia
RDF Book Mashup
US Census Data
W3C WordNet
lingvoj
Freebase
DBLP Hannover
CiteSeer
LAAS-CNRS
IBM
UniRef
Gov-Track
UMBEL
GEO Species
DBLP Berlin
Reactome
UniParc
Taxonomy
Open Cyc
Yago
LinkedCT
Drug Bank
PROSITE
Daily Med
GeneID
Pub Chem
KEGG
UniProt
Homolo Gene
Pfam
ProDom
Disea-some
CAS
Gene Ontology
Symbol
OMIM
ChEBI
Inter Pro
UniSTS
HGNC
MGI
PubMed
PDB
As of July 2009


Semantic Technologies
Semantic Web
LOD

The New York Times

Linked Open Data BETA

Search data.nytimes.com



data.nytimes.com

For the last 150 years, The New York Times has maintained one of the most authoritative news vocabularies ever developed. In 2009, we began to publish this vocabulary as linked open data.

The Data

As of 13 January 2010, The New York Times has published approximately ,10,000 subject headings as linked open data under a CC BY license. We provide both RDF documents and a human-friendly HTML versions. The table below gives a breakdown of the various tag types and mapping strategies on data.nytimes.com.

Type	Manually Mapped Tags	Automatically Mapped Tags	Total
People	4,978	0	4,978
Organizations	1,489	1,592	3,081
Locations	1,910	0	1,910
			9,969

Browse individual data records:

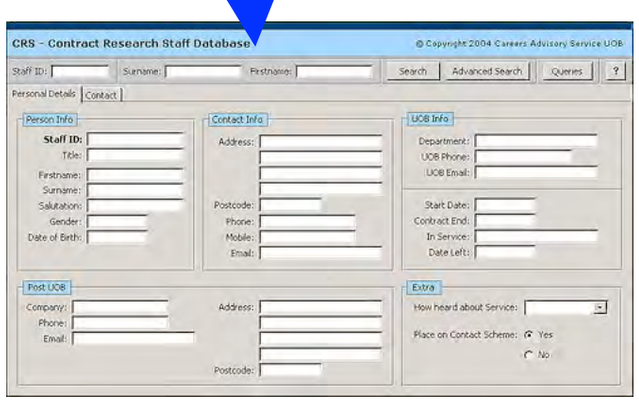
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

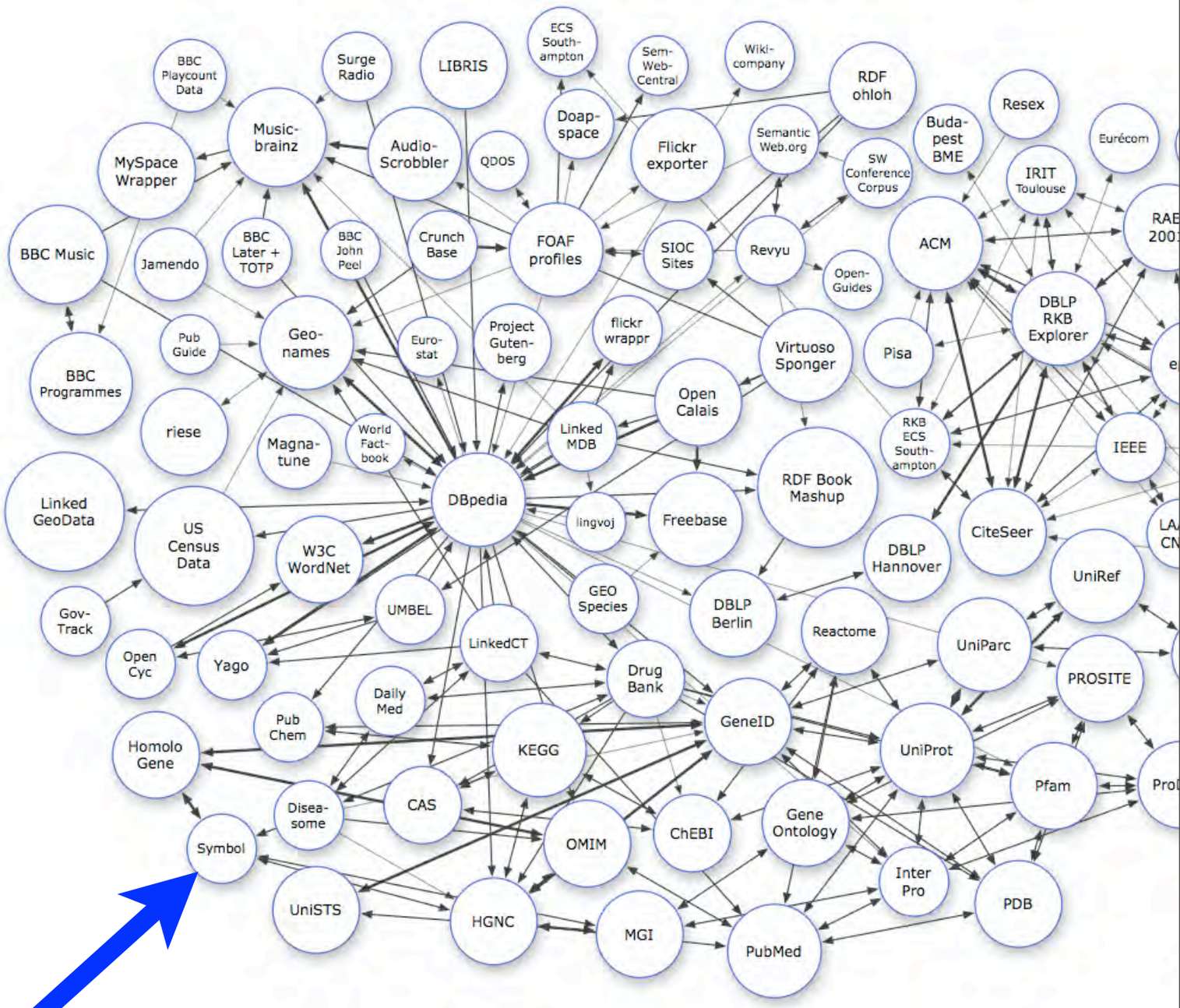
SKOS Files

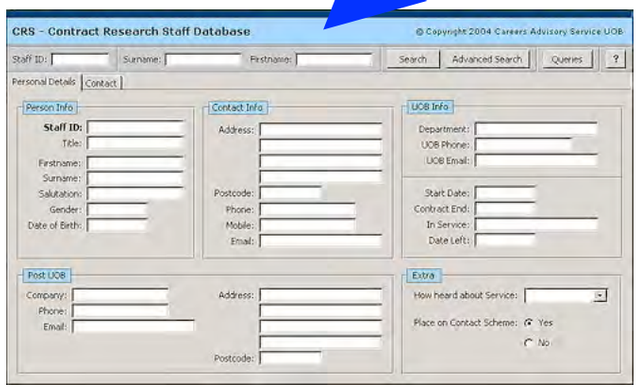
Download all of the data records as SKOS Files.

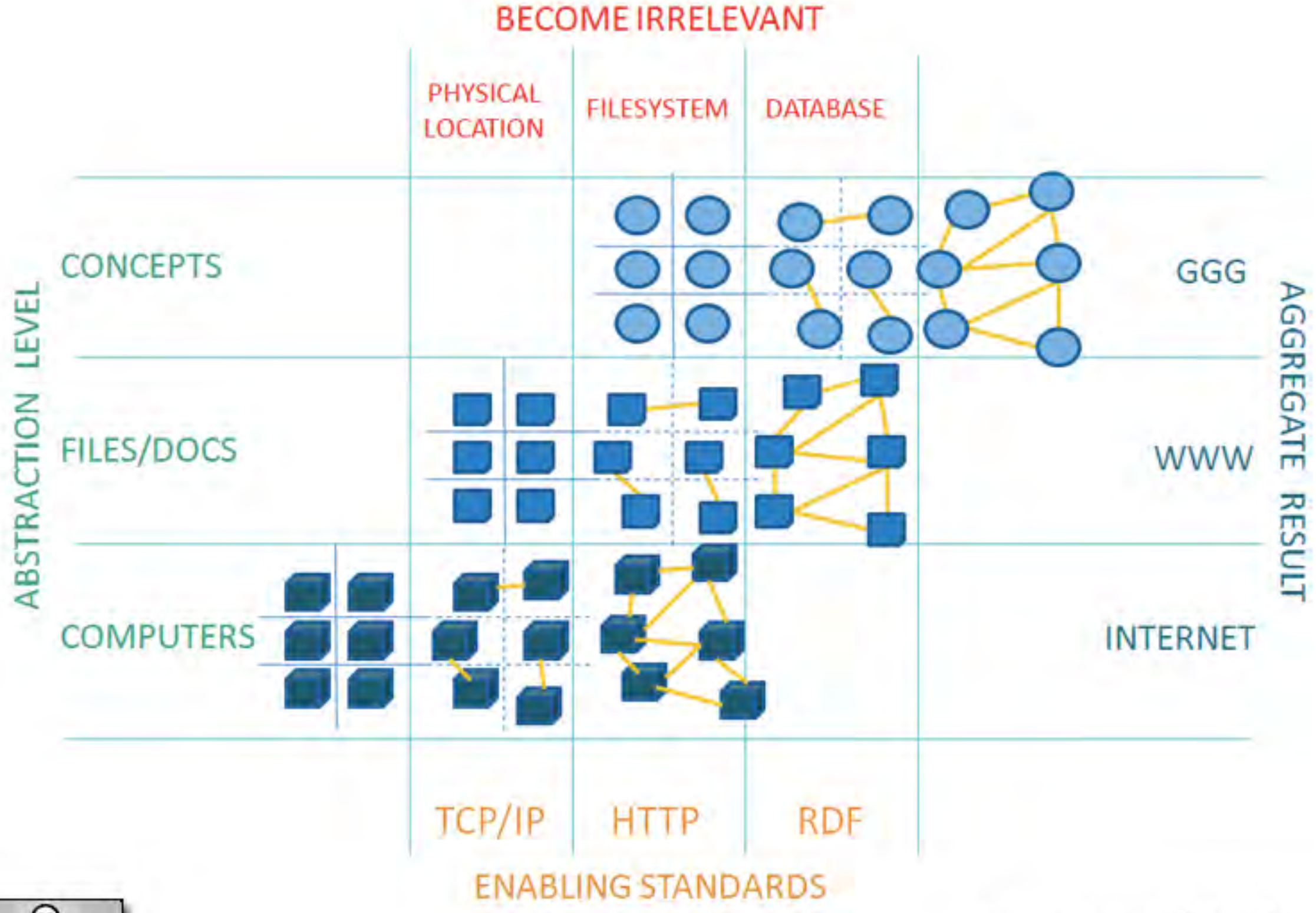
- People
- Organizations
- Locations


ORACLE
EXADATA
EXTREME PERFORMANCE
ORACLE


CRS – Contract Research Staff Database
© Copyright 2004 Careers Advisory Service UOB
Staff ID:
Surname:
Firstname:
Search
Advanced Search
Queries
?
Personal Details
Contact
Person Info
Staff ID:
Title:
Firstname:
Surname:
Salutation:
Gender:
Date of Birth:
Contact Info
Address:
Postcode:
Phone:
Mobile:
Email:
UOB Info
Department:
UOB Phone:
UOB Email:
Start Date:
Contract End:
In Service:
Date Left:
Post UOB
Company:
Phone:
Email:
Address:
Postcode:
Extra
How heard about Service:
Place on Contact Scheme:
Yes
No


BBC Playcount Data
Surge Radio
LIBRIS
ECS South-ampton
Sem-Web-Central
Wiki-company
RDF ohloh
Resex
Buda-pest BME
Eurécom
Music-brainz
Audio-Scrobbler
QDOS
Doap-space
Flickr exporter
Semantic Web.org
SW Conference Corpus
IRIT Toulouse
MySpace Wrapper
BBC Music
Jamendo
BBC Later + TOTP
BBC John Peel
Crunch Base
FOAF profiles
SIOC Sites
Revyu
Open-Guides
ACM
RAE 200
Pub Guide
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Virtuoso Sponger
Pisa
DBLP RKB Explorer
BBC Programmes
riese
Magna-tune
World Fact-book
Linked MDB
Open Calais
RKB ECS South-ampton
IEEE
Linked GeoData
US Census Data
W3C WordNet
DBpedia
lingvoj
Freebase
RDF Book Mashup
CiteSeer
Gov-Track
UMBEL
GEO Species
DBLP Berlin
DBLP Hannover
UniRef
Open Cyc
Yago
LinkedCT
Drug Bank
Reactome
UniParc
PROSITE
Daily Med
Pub Chem
GeneID
Homolo Gene
KEGG
UniProt
Pfam
CAS
Disea-some
Gene Ontology
Symbol
OMIM
ChEBI
Inter Pro
UniSTS
HGNC
MGI
PDB
PubMed


CRS - Contract Research Staff Database
© Copyright 2004 Careers Advisory Service UOB
Staff ID:
Surname:
Firstname:
Search
Advanced Search
Queries
?
Personal Details
Contact
Person Info
Staff ID:
Title:
Firstname:
Surname:
Salutation:
Gender:
Date of Birth:
Contact Info
Address:
Postcode:
Phone:
Mobile:
Email:
UOB Info
Department:
UOB Phone:
UOB Email:
Start Date:
Contract End:
In Service:
Date Left:
Post UOB
Company:
Phone:
Email:
Address:
Postcode:
Extra
How heard about Service:
Place on Contact Scheme:
Yes
No



source: www.aldobucchi.com

Tim Berners-Lee's Rules for Linked Data

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs so that they can discover more things.

Wednesday, April 21, 2010

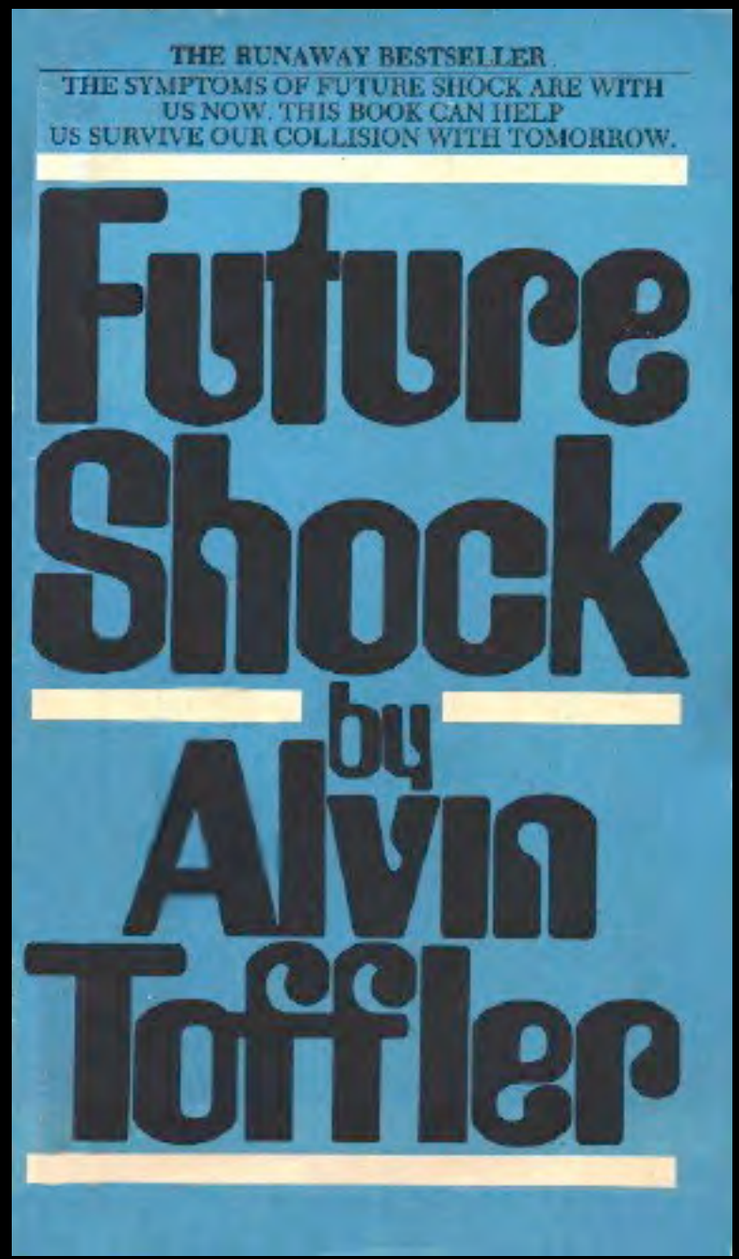
Data Inflation

- Megabyte (MB) = 2^{20}
- Gigabyte (GB) = 2^{30}
- Terabyte (TB) = 2^{40} 1,000GB
- Petabyte (PB) = 2^{50} 1,000TB
- Exabyte (EB) = 2^{60} or 1,000PB
- Zettabyte (ZB) = 2^{70} or 1,000EB
- Yottabyte (YB) = 2^{80} or1,000ZB

Jeff Pollock, Oracle

Businesses are in future shock

- Needs changing at faster pace
- M&A, new product categories, new regulations, changes in global economy accelerate needed changes.
- Information increasingly more central to the operation of any business.



"If information systems are to keep up with business,
we need to change more than technology -
we need to change how people deal with technology."
-- Jeff Pollock

Wednesday, April 21, 2010

We can, however, bridge the gap.

Semantic Technologies

Semantic Web

LOD

Linked Enterprise Data

RDBMS

CRM

BI

MIXING private and public data?

MIXING private and public data?

Absolutely!
And it is really useful to do so!


SingTel
8:48 PM
Bus Service
Find
Nearest bus stops
Keywords
Address / Bus Stop Cod
Go
You can tap the "Locate Me" button for near by bus stops
"Safari" Would Like to Use Your Current Location
Don't Allow
OK
Check iRIS for SBST bus arrival time service
Brought to you by Land Transport Authority

Problems with TimBL's Rules

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs so that they can discover more things.

New Rules

- Require labels
- Require descriptions
- Reuse vocabularies whenever possible
- Always preserve provenance (e.g. seeAlso)
- Relate when possible (skos:sameAs, etc.)
- Plan for identifier persistence

New Rules

useful info

- Require labels
- Require descriptions

links/ schema

- Reuse vocabularies whenever possible
- Always preserve provenance (e.g. seeAlso)

other URIs

- Relate when possible (skos:sameAs, etc.)
- Plan for identifier persistence

The Real Deal

We have case studies from a variety of organizations. Notice the differences between them:

- Broadcasting company
- Consumer service
- International organization
- Healthcare provider
- Publishing company
- Utility company
- Government department
- Open Source project

Attributions

http://www.flickr.com/photos/paraflyer/3179937406/

http://richard.cyganiak.de/2007/10/lod/

http://www.flickr.com/photos/joeseggiola/2722930520/

http://data.nytimes.com

http://aldobucchi.com

http://www.flickr.com/photos/oracleopenworld09/4012920029/

http://www.flickr.com/photos/ccgd/276642318/

Semantic Technology Conference banner used by permission of Semantic Universe

Springer reference used by permission of Springer; book cover is representative only

Future Shock cover (c) Random House

Other photos by David Wood