Untangling the Semantic Web

Louise Tutton
Senior Vice President, Scholarly Division

Overview

Introduction: the Web as we know it

Semantic Web

- What is it? Why is it important?
- Semantic Web for Publishing
- Real life examples

Conclusions

PUBLISHING TECHNOLOGY

The Web As We Know It

- Publications continue to follow traditional print model
- Volume of information on the web
- Computers understand syntax NOT meaning
- Search engines blindly retrieve all content relating to the search term
- Individual silos of data 'v' integrated resources

The Semantic Web: what is it? How can it help?

World Wide Web = Web of Documents

Semantic Web = Web of Data

"Data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications"

- *Sir Tim Berners-Lee*

The Semantic Web solution

- Enhances current WWW
- Adds meaning to data so that machines as well as humans can understand it
- Improves navigation and search
- Maximises ability to correlate disparate data sources
- Makes relevant data more visible
- Drives increased traffic to content
- Adds more value to content
- Results in richer applications

Semantic Web

The Semantic Web Solution

More possibilities

- Ability for untrained researchers to find data quickly
- Easier Research Collaboration
- Utilise cross-domain expertise to build richer applications

Possibilities are endless!

"**The most exciting thing about the Semantic Web is not what we can imagine doing with it, but what we can't yet imagine it will do**"
  - Sir Tim Berners-Lee
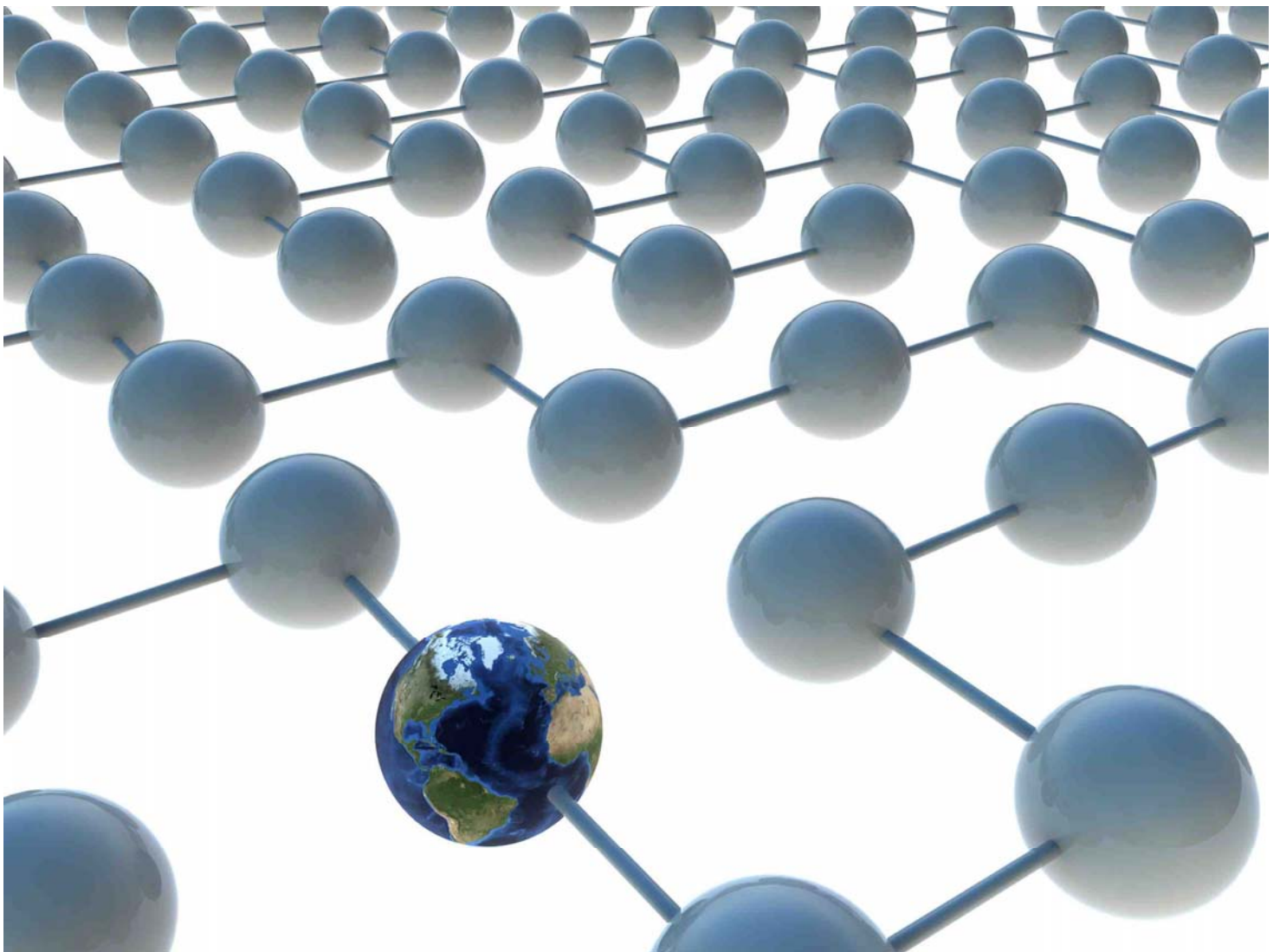
Real life examples - DBpedia

- Effort to smartly link Wikipedia's 7 million articles

- Links datasets from elsewhere on the web to Wikipedia

- Store info about people, places, music albums and films

- Describes 2.6 million "things" with 274 million "facts" using RDF

Supports more sophisticated user queries, such as:

"Give me all cities in NJ with more than 10,000 inhabitants"

"Give me all Italian musicians from the 18th century"

Friend of a Friend (FOAF) Project

- Describes people, their activities and their relations to other people

- Each RDF record and person has a unique ID (i.e. email address, IM ID, URL of blog/homepage)

- Tags include: "organization", "knows", "member of"

- Connects and integrates social networking sites (Facebook, MySpace, Blogs, LinkedIn, Flickr)

WOM Scouter

- Aims to improve shopping experience
- Offers shoppers advice on what to buy based on product barcode + buzz on blogs about a given gadget
- Uses NLP techniques to analyze what blogs are saying
- Provides straightforward positive or negative opinion
- Tested in electronics and book stores
- Could be adapted for movie, restaurant choices

Semantic Web for Publishing: How?

Making data Semantic Web ready

- Tag newly created data with meaning
- Pattern Matching + NLP for legacy data
- Bridging between relational databases & RDF
- RDF: semantic data model based on making statements about web resource

Semantic Web for Publishing: How?
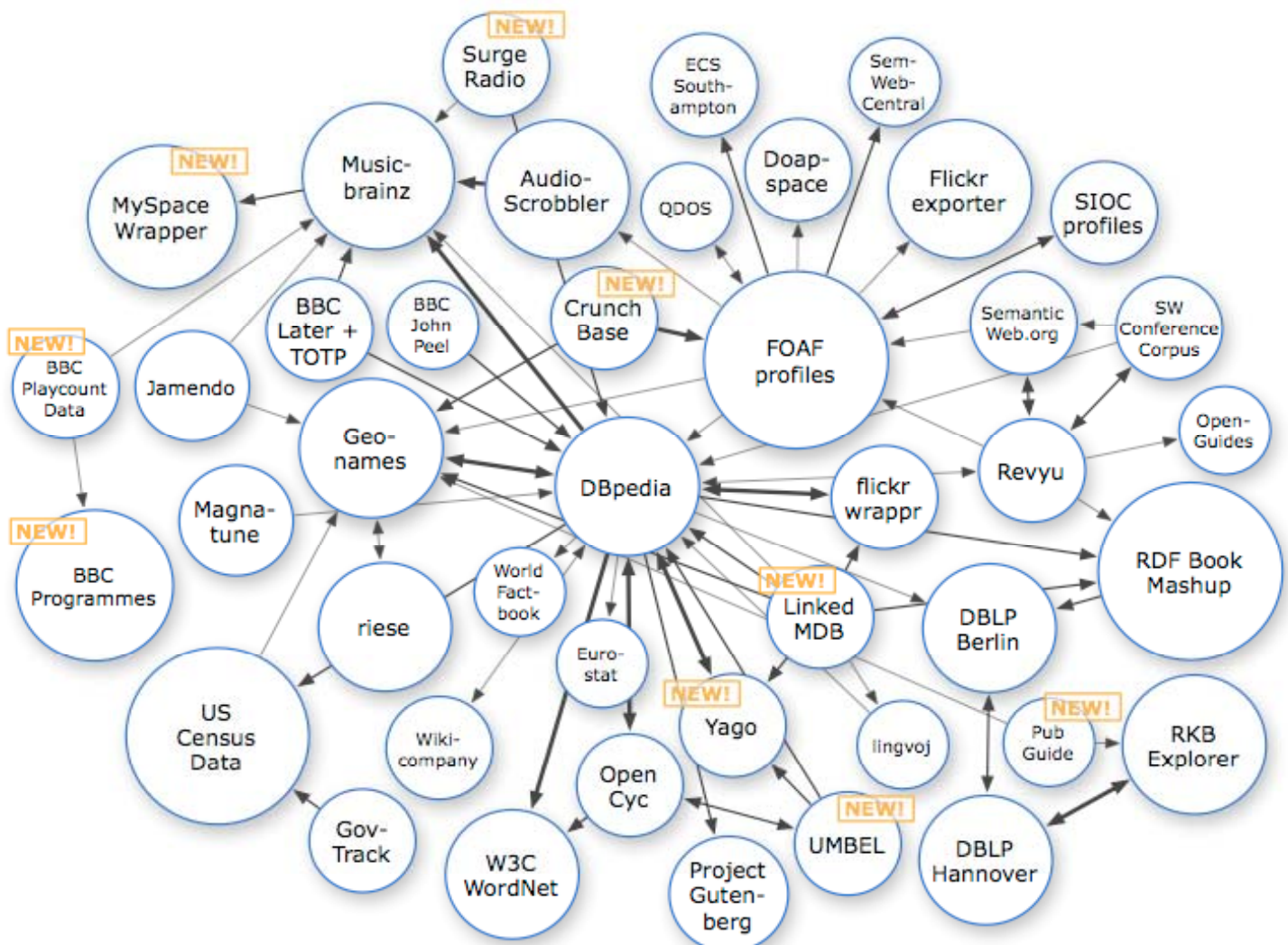
Openly available Datasets
How do you combine your data with external data

Linked Open Data – W3C SWEO project
Collectively > 17 billion RDF triples interlinked by 3 million links

Examples
- Bio2RDF
- DBPedia
- RDFBookMashup
- Project Gutenberg
- Geonames
- World Factbook

As of September 2008

Semantic Web Technologies at Publishing Technology: Metastore

Next generation metadata management

- Two years R&D investment
- RDF metadata repository behind IngentaConnect and pub2web
- Flexible, extensible data representation and tools
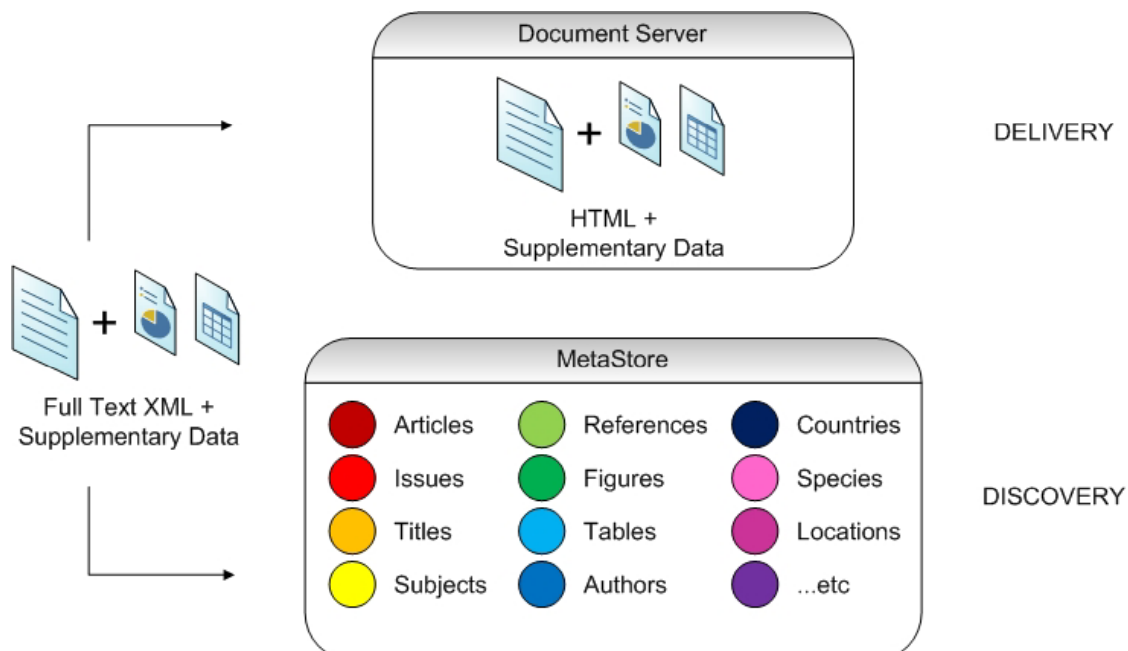
Making data more useful

- Additional discovery routes for users
- Increased visibility and cross promotion of products
- Ease of integration of external datasets with publisher data

---

Metastore – driving discovery
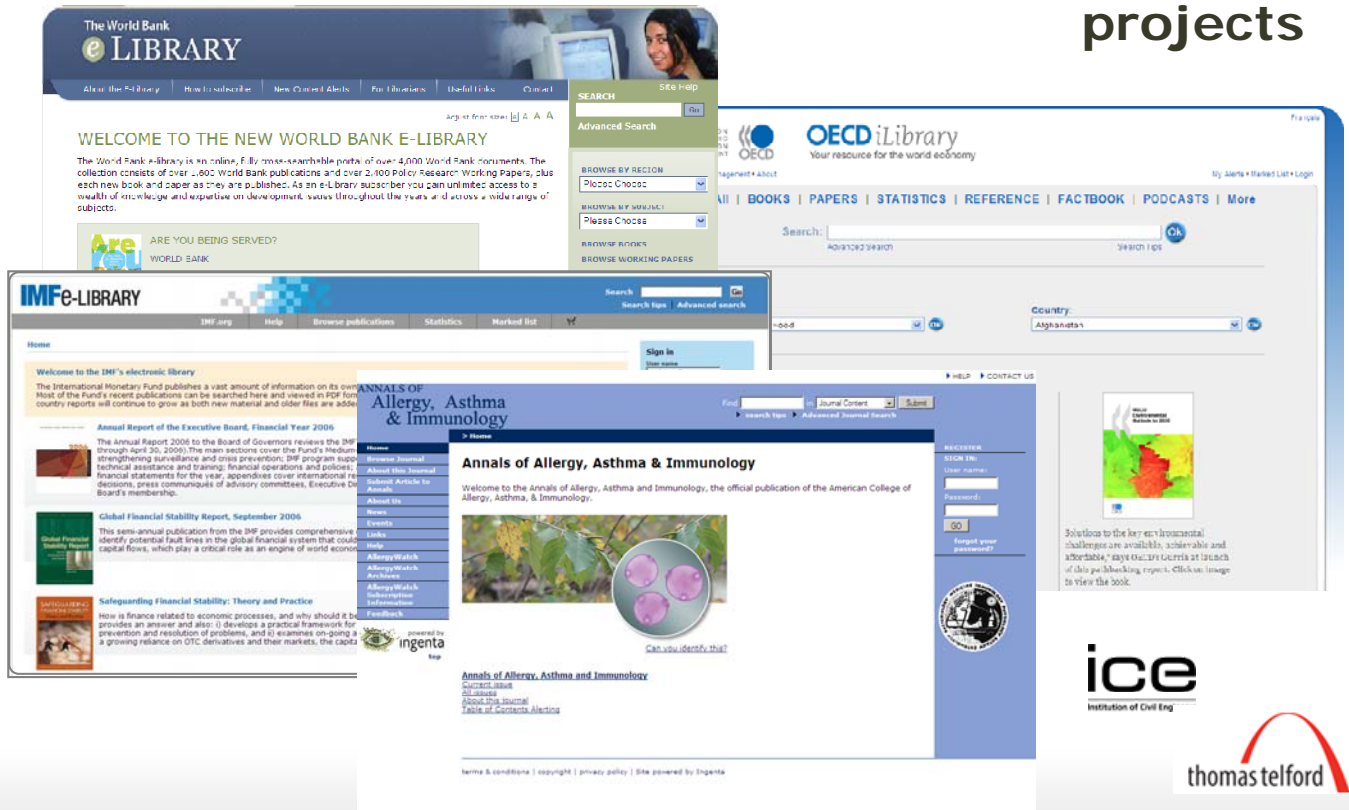
How could this look within pub2web?

Demo

IngentaConnect Mobile trial

- Targeted at undergraduate market
- Move away from traditional print approach
- Complements migration to XML
- Exposure to new markets.
  - 2.6 billion people have at least 1 mobile device.
  - India has 10 million new mobile device subscribers per month.
  - PC sales in decline in certain countries (in favour or mobile devices)

Current pub2web projects

Conclusions

- **Web of data** that can be processed by machines
- Extension of the current World Wide Web
- Many ways in which data can be made ready for the Semantic Web
- Benefits for publishers:
  - New routes for discovery of content
  - Enhanced visibility of content
  - Data re-use
  - Flexibility for future research
  - Richer applications
- Experimentation is key!