Interactive Information Access on the Web of Data

Lynda Hardman,
Jacco van Ossenbruggen, Alia Amin and Michiel Hildebrand

Interactive Information Access
http://www.cwi.nl/interactive_information_access

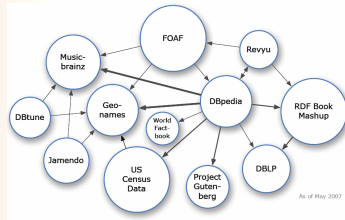
What is the Web of Data?

- linked data – data from your database(s)
- URIs, possibly identifying media fragments
- + annotations (tags)
- + links among fragments & annotations

How much linked data is there?
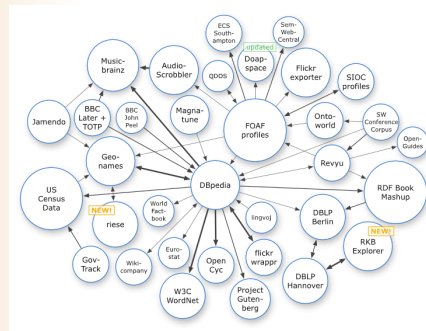
May 2007

Credit: *Chris Bizer*

3

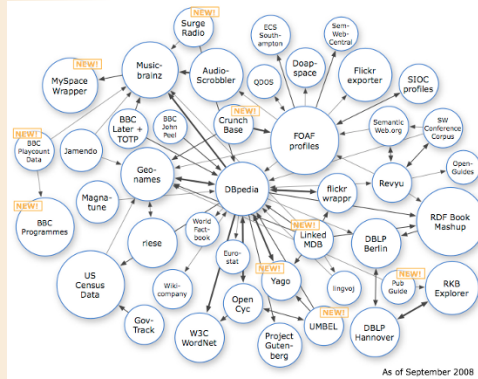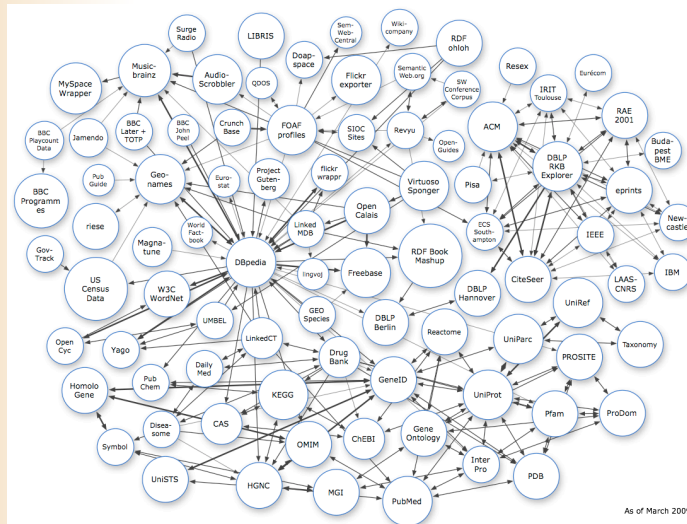# Linked data cloud March 2008

Credit: *Richard Cyganiak*

4

Linked data cloud September 2008

http://www4.wiwiss.fu-berlin.de/bizer/pub/lod-datasets_2008-09-18.png

5

Linked data cloud March 2009

> 4.5 billion RDF triples, interlinked by around 180 million RDF links
http://www4.wiwiss.fu-berlin.de/bizer/pub/lod-datasets_2009-03-05.png

6

Who are the users?
Why would they use the cloud?
What tasks can be supported?
How will the semantics help?

As of March 2009
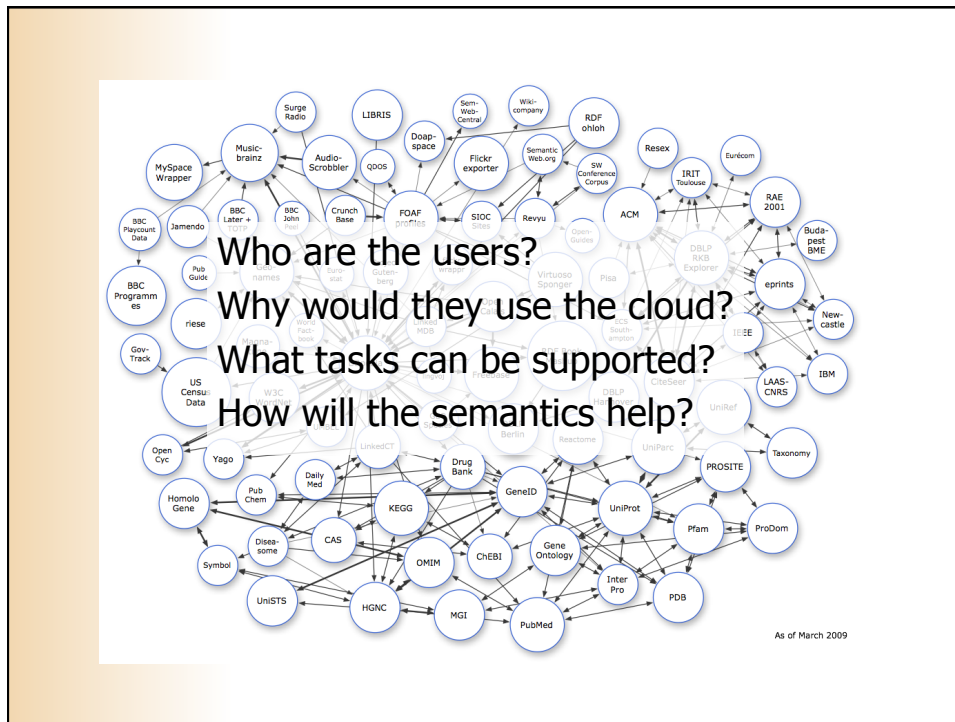
How can semantics help
with interactive information access?

- Query construction
  - disambiguate input
  - selection of available terms
- (Semantic) search algorithm
  - graph traversal
  - query expansion
  - RDFS/OWL reasoning
- Presentation of search results
  - grouping by property
  - visualization on timeline, map

8

Data sets in E-Culture demo

http://e-culture.multimedian.nl/resources/datacloud/

9

Browsing annotated collections of cultural heritage artefacts

- Who: Those interested in cultural heritage
- Why: Exploring artefacts available in repository
- What: Search combined collections
- How: autocompletion to suggest topics, organise results

10

http://e-culture.multimedian.nl/demo/session/search

Use of linked data in E-Culture demo

- Query construction
  - auto-completion uses strings found in "data" and "concepts"
  - suggestions are grouped and ordered using links among items
- Result set
  - uses empirical balance between "closeness" to search string and non-intuitive path
- Result presentation
  - uses grouping of result set to show breadth of results
  - uses no particular ordering within each group

12

Subject Annotation

The execution of Johan van Oldenbarnevelt

RP-P-OB-77.320

A print that depicts the beheading of

Who: Professional annotators
Why: Subject matter annotation of 700.000 prints
What: Search in multiple thesauri for annotation terms
How: Autocompletion on who/what/where/when

Who Historical persons

person or organization

Oldenbarnevelt, Johan van [x]

What Iconclass/WordNet (en), events (nl)

(mythological) concept, object or event

beheading [x]

Where Name of place or region

geographical place

Den Haag [x]

When Date, year or period

enter date

E−Culture MultimediaN *Rijksmuseum PrentenKabinet Online*   Login | help | English

search   browse   local view   **annotate**   search

annotate: Veroordeling van Johan van Oldenbarnevelt

Veroordeling van Johan van Oldenbarnevelt

RP-P-OB-77.320

Blad met een voorstelling van de onthoofding van Johan van Oldenbarnevelt op het Binnenhof te 's-Gravenhage op 13 mei 1619. Gezicht op het plein met alle omringende gebouwen en het verzamelde publiek. In de toren linksboven het hof van prins Maurits. Om de voorstelling van de onthoofding staan de portretten van de zes andere veroordeelden, een scène met de kist van Van Ledenberg aan de galg en een gezicht op het kasteel Loevestein.

Who Historical persons

person

What Iconclass (en), WordNet (en), events (nl)

(mythological) concept, object or event

Where Name of place or region

geographical place

When Date, year or period

enter date

done | cancel

http://e-culture.multimedian.nl/pk/annotate

14

Use of linked data in annotation task

- Query construction
  - auto-completion compares string in query with terms in thesauri
- Result set (the set of terms used to construct the menu)
  - terms that contain the string
- Result presentation (in the selection menu)
  - uses grouping of results depending on entry field
  - ordering also dependent on entry field
  - presentation of additional information differs per thesaurus and annotation field

Michiel Hildebrand, Jacco van Ossenbruggen, Lynda Hardman and Geertje Jacobs. Supporting subject matter annotation using heterogeneous thesauri, a user study in web data reuse. Technical Report INS-E0902, CWI, February 2009.
http://ftp.cwi.nl/CWIreports/INS/INS-E0902.pdf

15

Study of information use by cultural heritage experts

Understand the cultural heritage experts' information seeking needs.

- Why do cultural heritage expert search?
- What are the typical experts' search task?
- What sources do they use?

16

Why do CH experts search?

- Object handling: restoration, acquisition, or loan
- Exhibition: finding themes, comparison studies
- Publication: for peers or for general public
- Managing collections' documentation: updating records
- Building thesauri: used for annotation and search

17

Key Findings

1. Information gathering as primary task
2. Searching in multiple sources
3. Communication with other experts
4. Provenance and trust

18

Prototype comparison search – bar chart

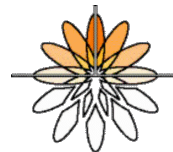

Prototype comparison search - map

Conclusions

- Build specific application
- Determine information need
- Select data sources for task
- Ensure access to provenance information without being intrusive
  - remember hyperlink markers 20 years ago?
- Investigating re-usable interface components
  - autocompletion

21

Acknowledgements

Rijksmuseum Amsterdam

http://e-culture.multimedian.nl/

22