Open Data, Linked Data

Janet McKnight
Oxford University Computing Services

July 2010

Our panellists

1. Graham Klyne (Zoology)
2. Sally Rumsey (Bodleian Library)
3. Neil Jefferies (Bodleian Library)
4. Howard Noble (OUCS)

What is open data?

'A piece of knowledge is open if you are free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and share-alike.'

[From `http://www.opendefinition.org/`]

What is linked data?

'The semantic web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related data.'

[Tim Berners-Lee, from
`http://www.w3.org/DesignIssues/LinkedData.html`]

Four rules of linked data

- Use URIs for things
- Use HTTP URIs so that people can look up those names
- Return data using standards (RDF)
- Link to other things!

RDF is (stock description)

"The RDF data model is similar to classic conceptual modeling approaches such as Entity-Relationship or Class diagrams, as it is based upon the idea of making statements about resources (in particular Web resources) in the form of subject-predicate-object expressions. These expressions are known as triples in RDF terminology. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object."

http://en.wikipedia.org/wiki/Resource_Description_Framework

All true, but probably unhelpful, so let's try another approach...

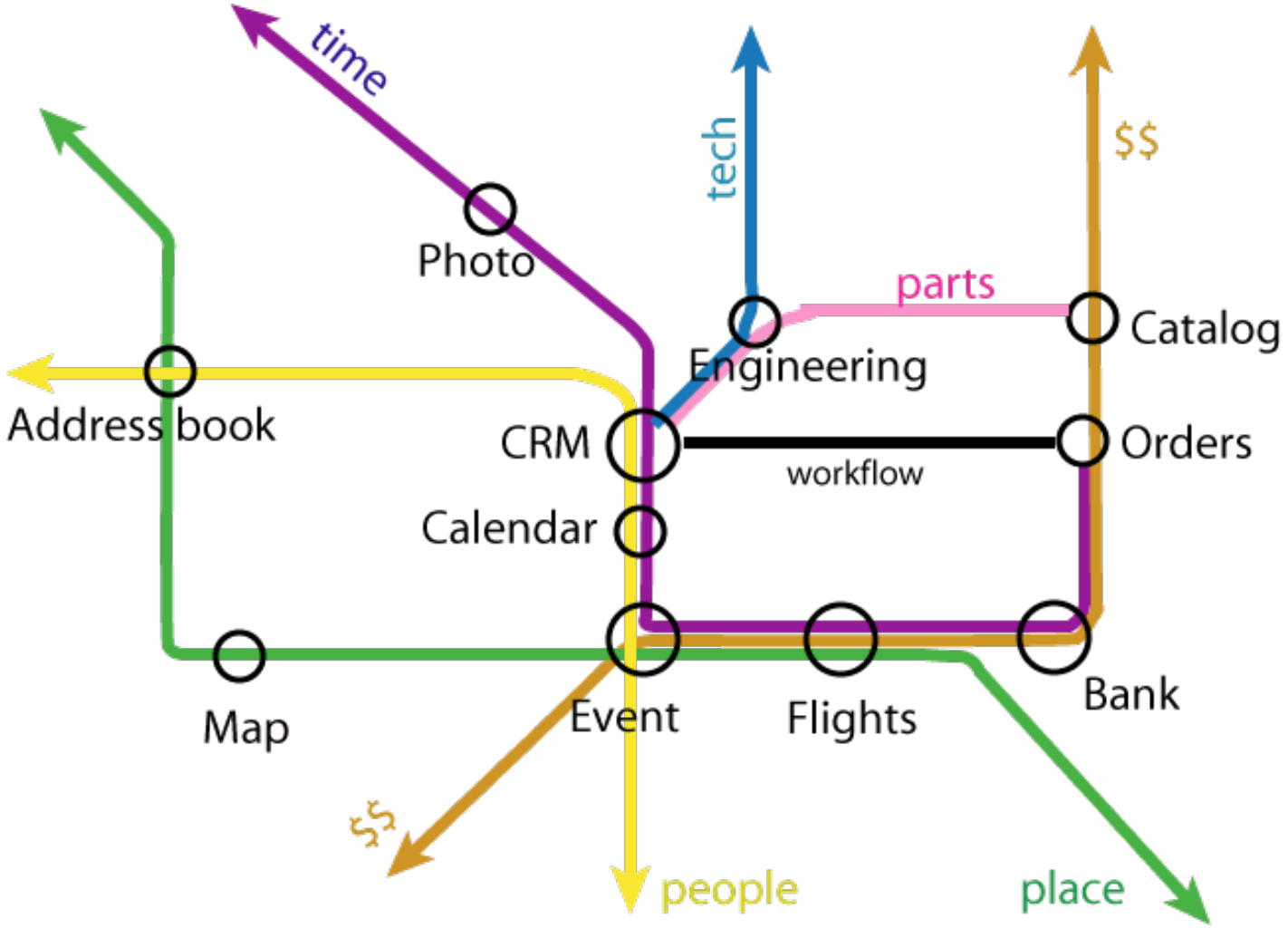
“Connolly's Bane”

"The bane of my existence is doing things that I know the computer could do for me"

– Dan Connolly, http://www.nature.com/nature/webmatters/xml/xml.html

Many of these things involve combining information from diverse, independently-developed applications

“Semantic subway”
Combining diverse information



Tim Berners-Lee: http://www.w3.org/2006/Talks/0314-ox-tbl/

RDF is

"RDF is a general method to decompose any type of knowledge into small pieces, with some rules about the semantics, or meaning, of those pieces. The point is to have a method so simple that it can express any fact, and yet so structured that computer applications can do useful things with it."

http://rdfabout.com/quickintro.xpd

- a universal data format
- schema-free
- open-world
- uniformly structured
- uses global identifier namespace
- semantically formalized

CLAROS
Classical art data fusion

- Maps data from diverse independent data sources
- Search/browse over merged data

  Universal – target for relational, XML, and other data

  Schema-free – enhancing data structures incrementally

  Open world – new data sources

  Uniform – merging via RDF

  Global names – URIs for objects

  Semantics – inference engine

http://www.clarosnet.org

ADMIRAL
Small-scale research data curation

- Research data curation
- “Curation by addition”

  Start with what is available, figure out what we've got, enhance incrementally until ready for publication

  universal – can represent arbitrary (meta)data

  scheme-free – no foreknowledge of structures needed

  open-world – can add new data, missing isn't broken

  uniform structure – merge (meta)data from diverse sources

  global names – link to public datasets; open publication

- http://imageweb.zoo.ox.ac.uk/wiki/index.php/ADMIRAL

Use of linked data within ORA and the research registry

Three things you should know

- What?
- Why?
- Plans

Sally Rumsey, ORA Service & Development Manager

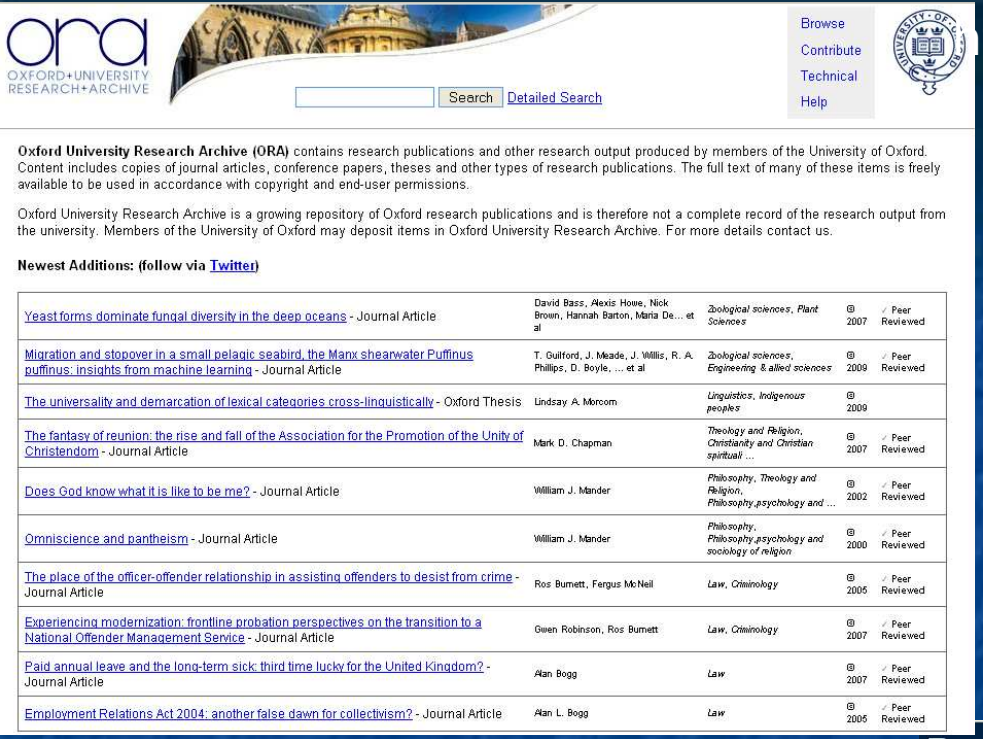
sally.rumsey@bodleian.ox.ac.uk

UNIVERSITY OF OXFORD

Bodleian Library
UNIVERSITY OF OXFORD

ora
OXFORD+UNIVERSITY RESEARCH+ARCHIVE

ttp://ora.ouls.ox.ac.uk/



Vocabulary @ OULS - a space for vocabularies.

Introduction

The following vocabularies are currently hosted here:

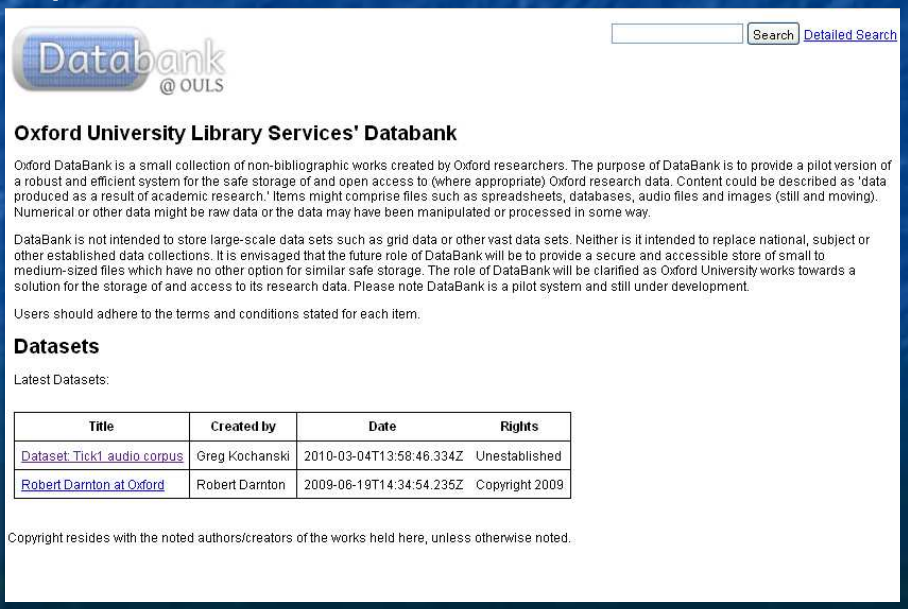
Academic Research Project Funding Ontology (ARPFO)
A vocabulary for describing the structure and relationships between organisations and individuals involved in providing funding or receiving funds for academic research or projects.

HTML produced from the toolchain at vocab.org - a copy of which can be downloaded (with example rdf from the funding vocab) here.

http://vocab.ox.ac.uk

http://163.1.127.171/index

ttp://databank.ouls.ox.ac.uk/



Why?

Semantic web used

- De facto standard
- Reflect complex structure
- Easy data sharing & re-use

Problems

- Aggregation
- Sharing
- Not a database
- Creating linked data version

Plans

- Future developments
- Timing

UNIVERSITY OF OXFORD

Google Libraries Project Bodleian Update

Neil Jefferies
R&D Project Manager
Systems & eResearch Service (SERS)
Bodleian Libraries, Oxford University

UNIVERSITY OF OXFORD

Source Material

- Over 400K digitized books, approx 300 pp
  - After dedup, pruning of collection
- JPEG2000 images + raw OCR text + “metadata”
- Volume packages generated on-demand
- 50-100TB of raw data
- Currently held on Google servers

UNIVERSITY OF OXFORD

Ingest Process

- Download volume packages
  - Bandwidth limited, checksum to validate overall package
- Unpack and validate
  - Check against manifest, zero length files, missing pages etc.
  - Manual sampling/QA of content
- Cross-reference with OLIS for bib metadata
- Generate dissemination formats
  - J2K → JPEG → PDF (Raw book vs image+OCR)
- Stored as objects with RDF descriptions

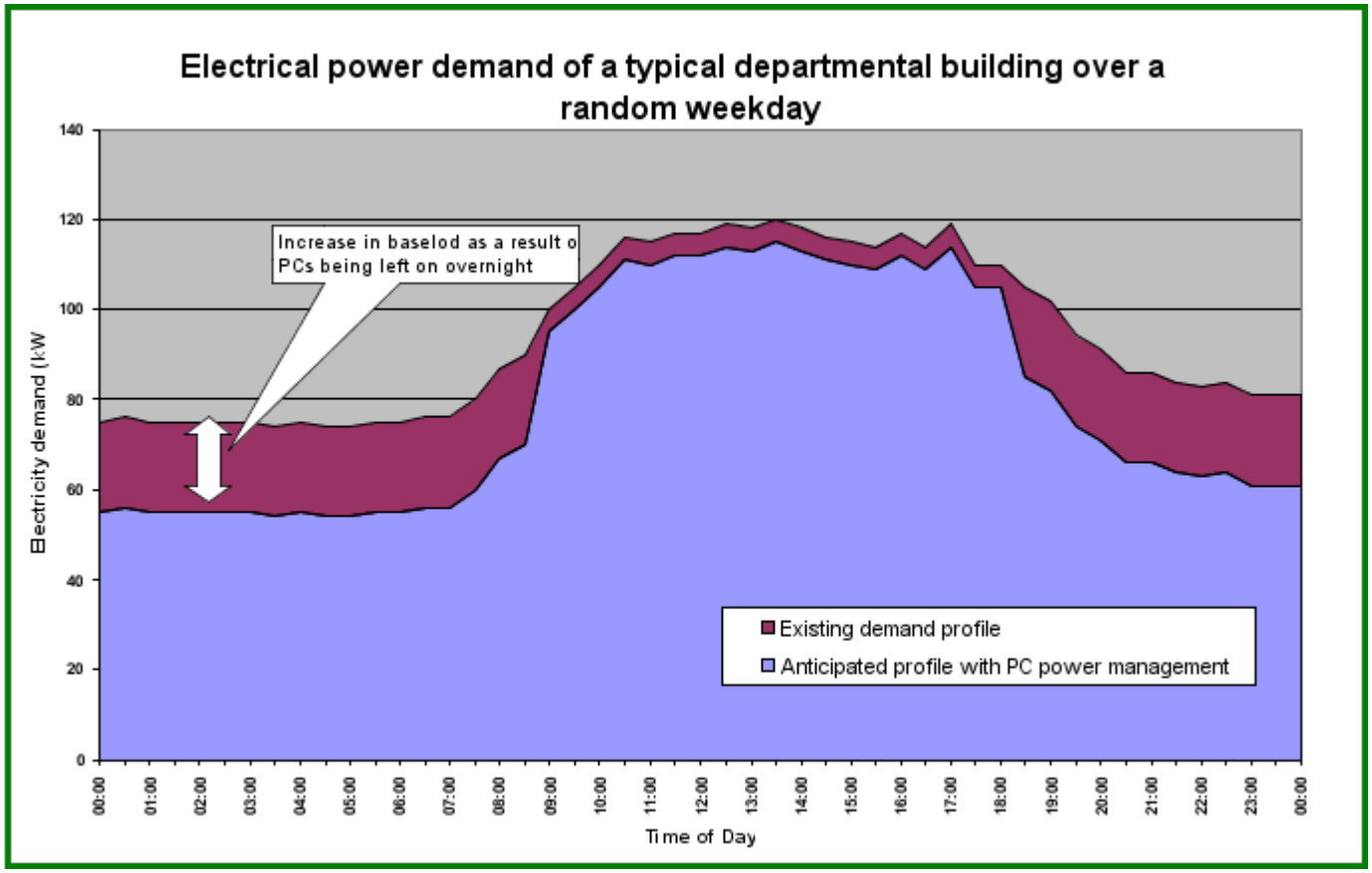
UNIVERSITY OF OXFORD

Dissemination

- Human-readable formats
  - Browse, faceted search (SOLR), page-turner, e-Citation
- Machine-readable formats
  - RDF/Linked data, OAI-PMH/ORE, RSS, Atom
  - Links to author/place thesauri such as CERL
- Added value services
  - Open Annotation (www.openannotation.org)
    - Annotation, translation, transcription – as RDF objects
    - PDF's including annotations?
  - Print-on-demand?

ICTF 2010

Open data



Three *policies:*

1. Always on: 110/1000 x (24 x 365) x 16,000 x £0.12 = £1,850,112
2. Power down after work: 110/1000 x (10 x 223) x 16,000 x £0.12 = £470,976
3. Hope for the best = £?

The social side of green IT

Consider a world where:

- 20% of people have changed their lives to reduce their energy consumption significantly (eco-warriors)
- 30% of people consume as much electricity as they can afford (sceptics and deniers)
- 50% of people consume as they see the people around them consuming (migrating birds)

Question: What effect do the eco-warriors have on total energy consumption?

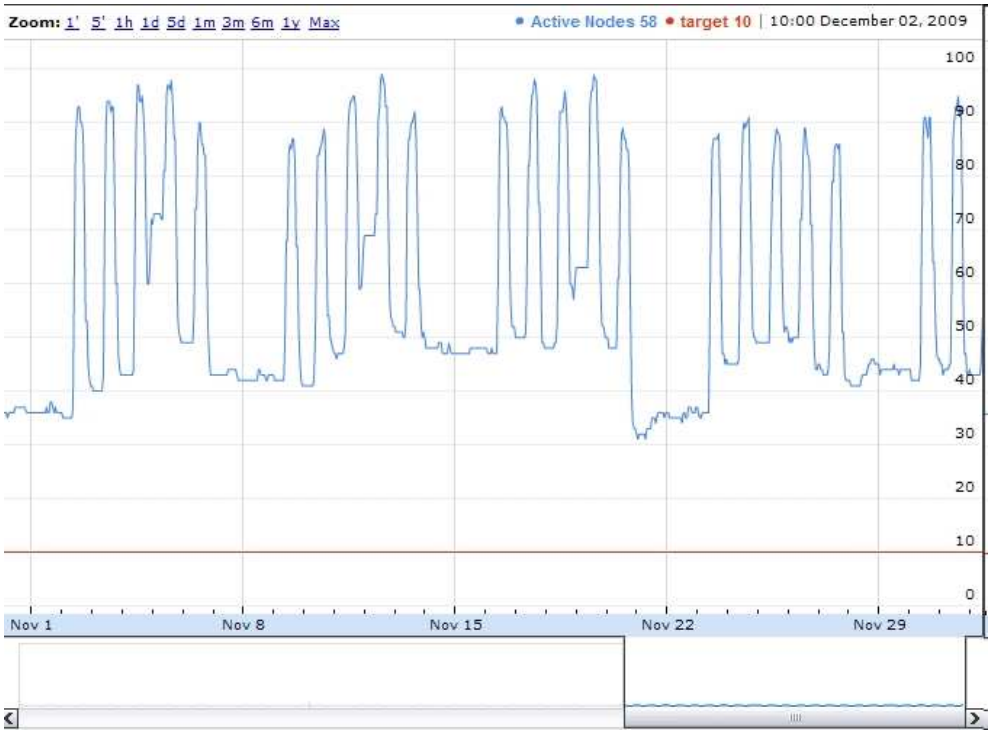
A. Total consumption will decrease

B. Total consumption will stay the same

C. Total consumption will increase

The problem with graphs and league tables

1. The internet is evolving into a space where people can easily make data available for others to use
2. People will want to visualise data in new ways, compare and make links between information that helps them understand the world
3. People will increasingly expect data to be available openly to allow them to do this



Rank	University	Policy: Env. policies	Env. Staffing	Env. Auditing	Ethical Investment	Carbon Mgmt.	Sustainable Food and Fairtrade	Student and Staff Engagement	Performance: Energy	Waste	Carbon Emissions	Water	Total Score
82=↓	Keele University		★	●		●		★		●	●	●	25.5
82=↑	The Open University			●	●		●		●		●	●	25.5
82=↑	University of Chichester		★	●		●		●	●	●	●	●	25.5
85↑	University of Essex			●	●	●	★	★	●		●	●	25.0
86↓	University of Sheffield	★		●	●	●			●	●	●	●	24.5
87=↓	Aberystwyth University			●	●	●	★	●	●	★	●	●	24.0
87=↓	Queen Mary, University of London		●	●	●			★	●		●	●	24.0
89=↓	Birkbeck College, University of London		●	●	●	●	●		●				23.5
89=↓	University of Cambridge			●	●	●	●	★	●	●	●	●	23.5
89=	University College Falmouth	●		●	●	●	●		●		●	●	23.5
89=↑	Goldsmiths, University of London			●	●		●	★	●	●	●	●	23.5
89=↓	Heriot-Watt University			●	●					●	●	●	23.5
89=↓	University of Kent	●		●	●	●	★	★	●		●	●	23.5
89=	London Metropolitan University			●	●	●	●		●		●	●	23.5
89=↓	University of Oxford			●		●	●	★	●	●	●	●	23.5

Social Capital + Open Data

Social Capital + Open Data

Questions?
Over to you...

Thank you!

- opendata@maillist.ox.ac.uk

- janet.mcknight@oucs.ox.ac.uk
- graham.klyne@zoo.ox.ac.uk
- sally.rumsey@bodleian.ox.ac.uk
- neil.jefferies@bodleian.ox.ac.uk
- howard.noble@oucs.ox.ac.uk