Citation Services (Overview)

Les Carr, Tim Brody
University of Southampton

International Workshop on Co-operation on Citation Services and Repositories
Utrecht, June 14th 2010

JISC

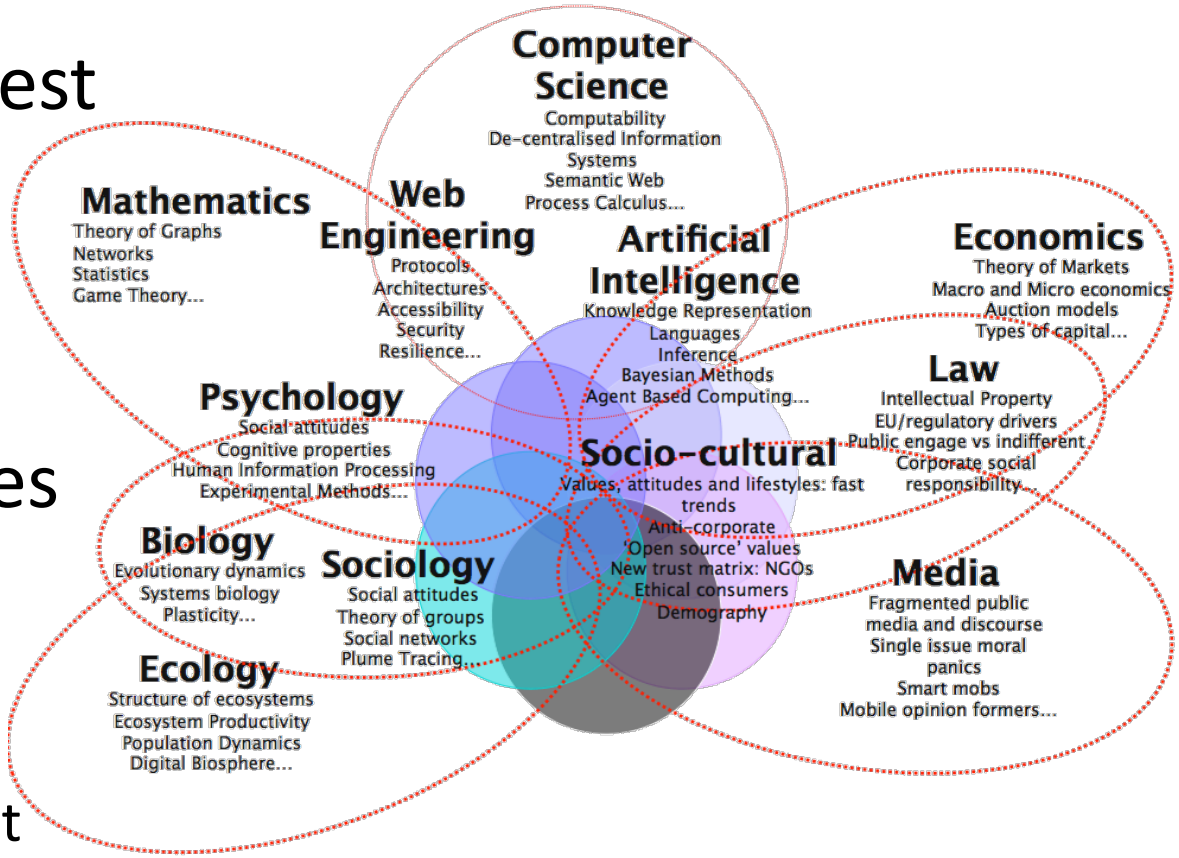
Decades of Achievement

Internet (packets)	Web (documents)	Repository (preservation)

What have we achieved with this multilayered platform in the last decade?
What do we need to consolidate?

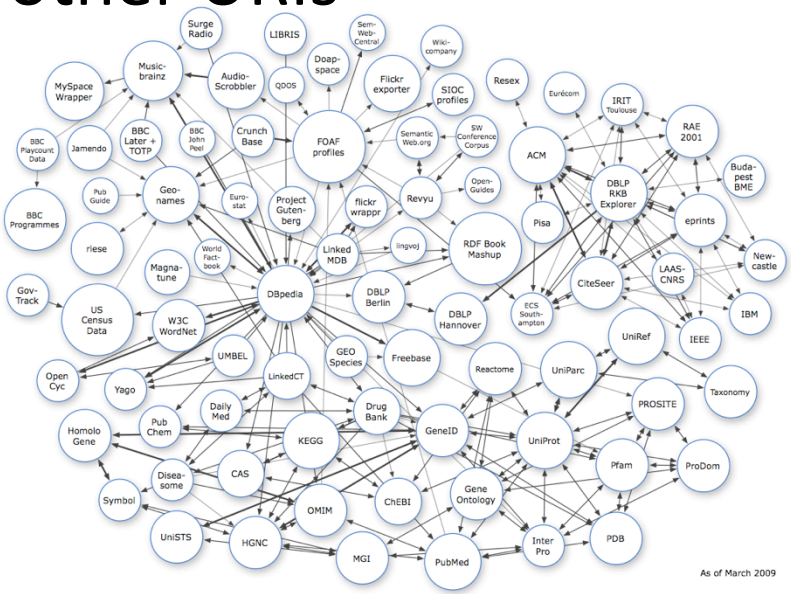
Web Science

- The Web is the largest human information construct in history
- Web Science enables us to
  - Understand what it is
  - Engineer its future
  - Ensure its social benefit
- What is the impact of the Web on society?



Four principles of Linked Data

- Four principles of linked data
  1. Every thing has a URID
  2. Publish information about that thing at that URI
  3. Use standard, open formats
  4. Mention other things using other URIs
- Result
  - An open web of open information



Four principles of scholarly comms

- Four principles of scholarly comms
  1. Every paper has a DOI
  2. Publish information at the DOI
  3. Use a standard format
  4. Reference other papers

- Result
  - The literature - a private web of licensed information that is not free to use
  - Even using the metadata is problematic

Aims of Citation Services

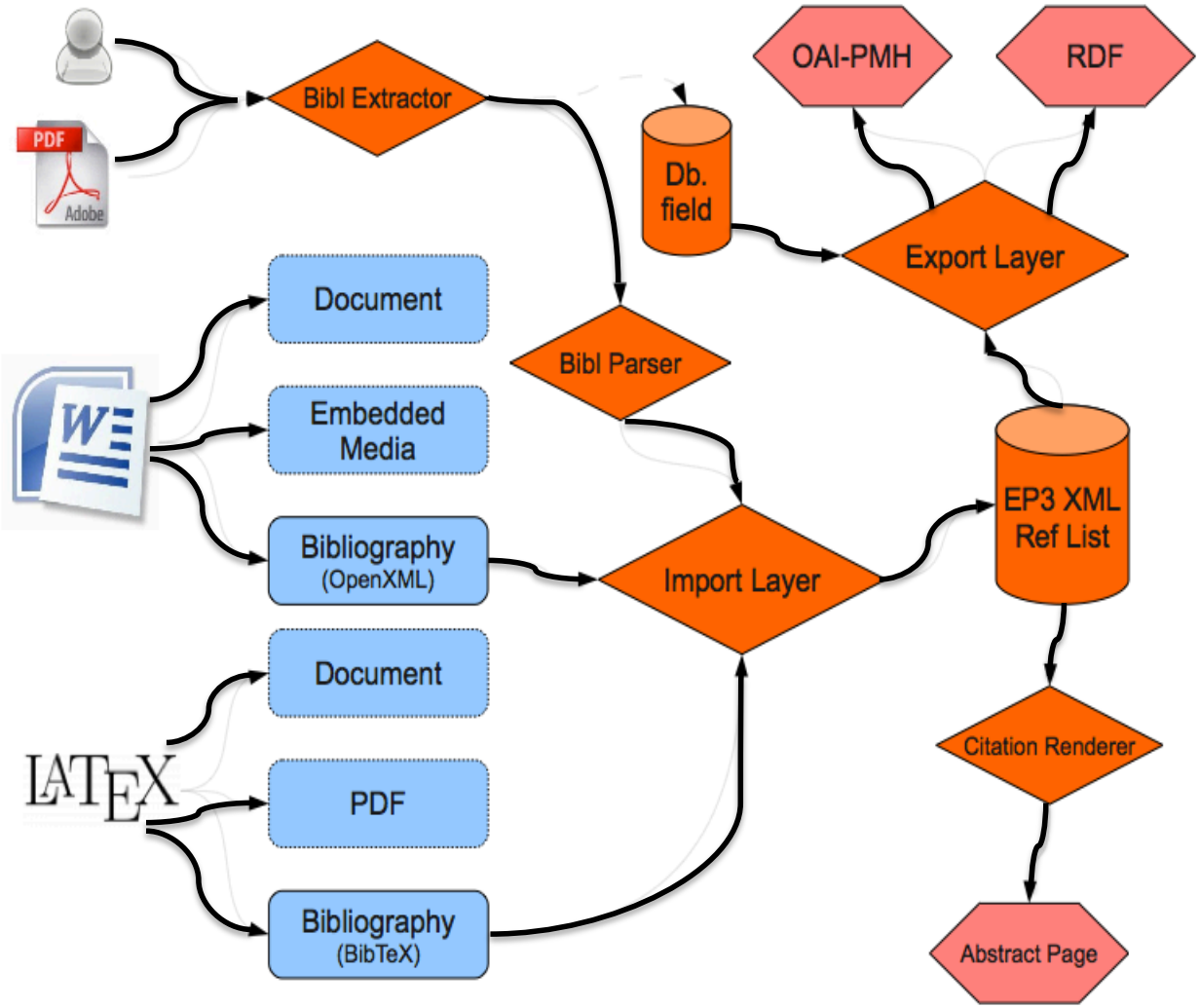
- improve the ways in which citation data relating to open access research documents are identified and shared
- current services and bibliographic research is restricted to a small number of commercial actors
  - insufficiently developed for the academic community
- citation information needs to be recognized as a part of the Commons
  - a freely shared citation infrastructure
  - format-agnostic
  - transdisciplinary,
  - international
- used by a wide variety of services

Project Outline

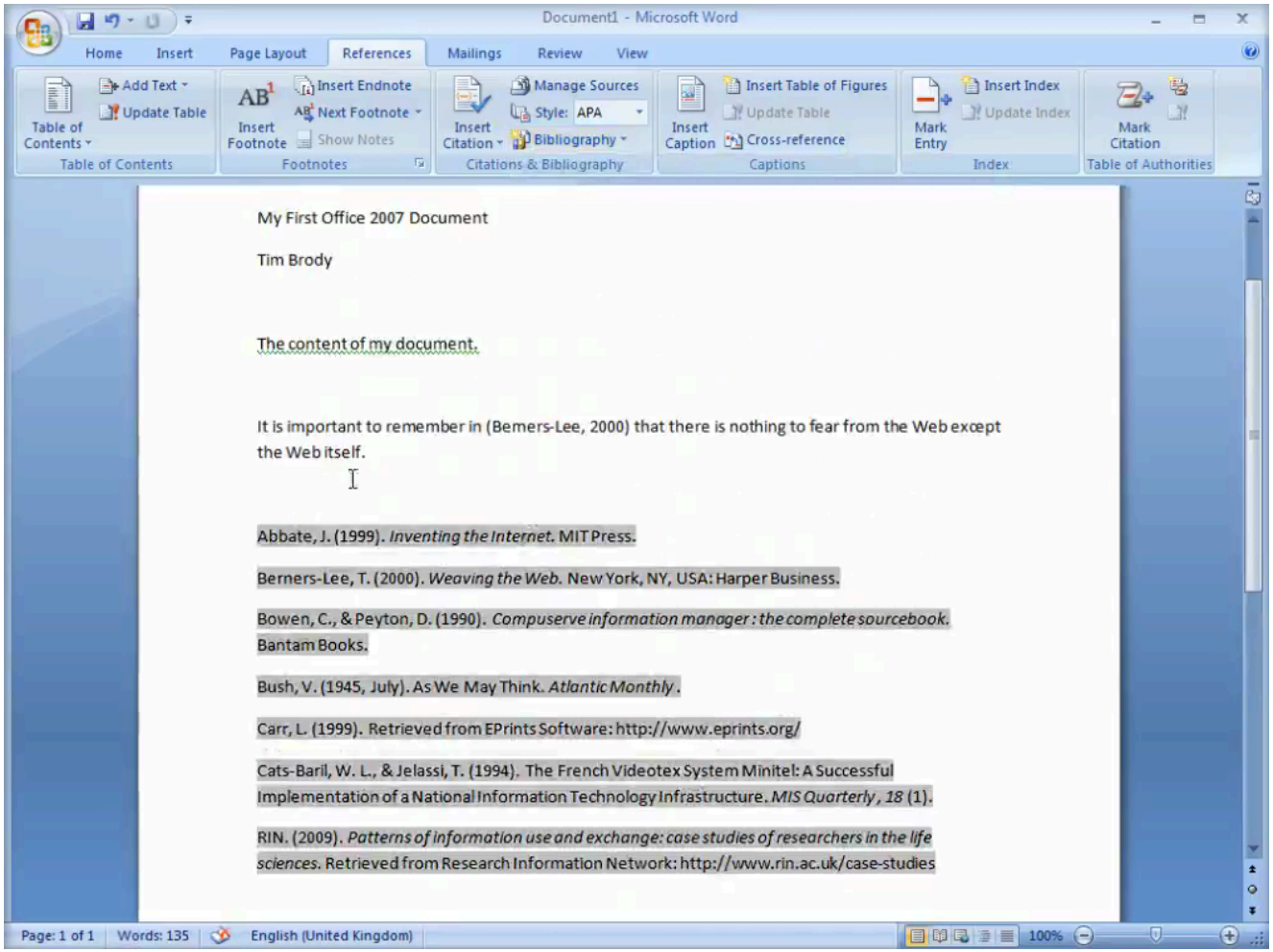
1) Author support tools: for Microsoft Word and BibTeX
2) Reference list deposit in repository metadata for general repository deposit workflow
3) OAI-PMH citation schema and protocol extension
4) Reference extractor (Perl)
   - extracts bibliographies from a research work (PDF, HTML, Office)
5) Reference deconstructor
   - breaks down a reference list into basic constituent parts for later analysis.
6) Large testbed of representative documents
7) One or more basic services (citation databases) which can collect, combine and disambiguate (recognise eliminate duplicate) citations
   - building on citebase
8) Exemplar advanced value-added services: examples of such services might be citation graph visualisation, network visualisation and trackback (track back through the literature using citation links) services
9) Infrastructure for auditing and quality assurance services

Citation Workflow

- Opportunities to obtain citation data
  - manually from author
  - explicit bibliography data (Office or BibTeX)
  - automated data scraping



Demo


Document1 - Microsoft Word
Home
Insert
Page Layout
References
Mailings
Review
View
Table of Contents
Add Text
Update Table
Table of Contents
Insert Footnote
Insert Endnote
Next Footnote
Show Notes
Footnotes
Insert Citation
Manage Sources
Style: APA
Bibliography
Citations & Bibliography
Insert Caption
Insert Table of Figures
Update Table
Cross-reference
Captions
Mark Entry
Insert Index
Update Index
Index
Mark Citation
Table of Authorities
My First Office 2007 Document
Tim Brody
The content of my document.
It is important to remember in (Berners-Lee, 2000) that there is nothing to fear from the Web except the Web itself.
Abbate, J. (1999). Inventing the Internet. MIT Press.
Berners-Lee, T. (2000). Weaving the Web. New York, NY, USA: Harper Business.
Bowen, C., & Peyton, D. (1990). Compuserve information manager : the complete sourcebook. Bantam Books.
Bush, V. (1945, July). As We May Think. Atlantic Monthly .
Carr, L. (1999). Retrieved from EPrints Software: http://www.eprints.org/
Cats-Baril, W. L., & Jelassi, T. (1994). The French Videotex System Minitel: A Successful Implementation of a National Information Technology Infrastructure. MIS Quarterly, 18 (1).
RIN. (2009). Patterns of information use and exchange: case studies of researchers in the life sciences. Retrieved from Research Information Network: http://www.rin.ac.uk/case-studies
Page: 1 of 1
Words: 135
English (United Kingdom)
100%

Next Stages

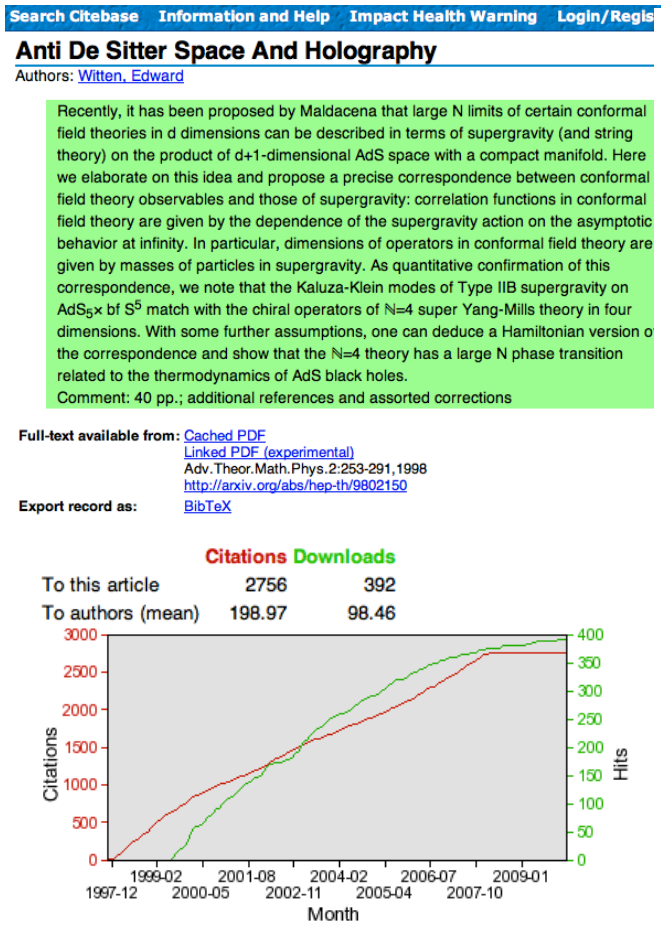
- Build testbed
- Collect citation extraction software
  - Open source projects
  - Citebase, Citeseer etc
- Evaluate software against testbed

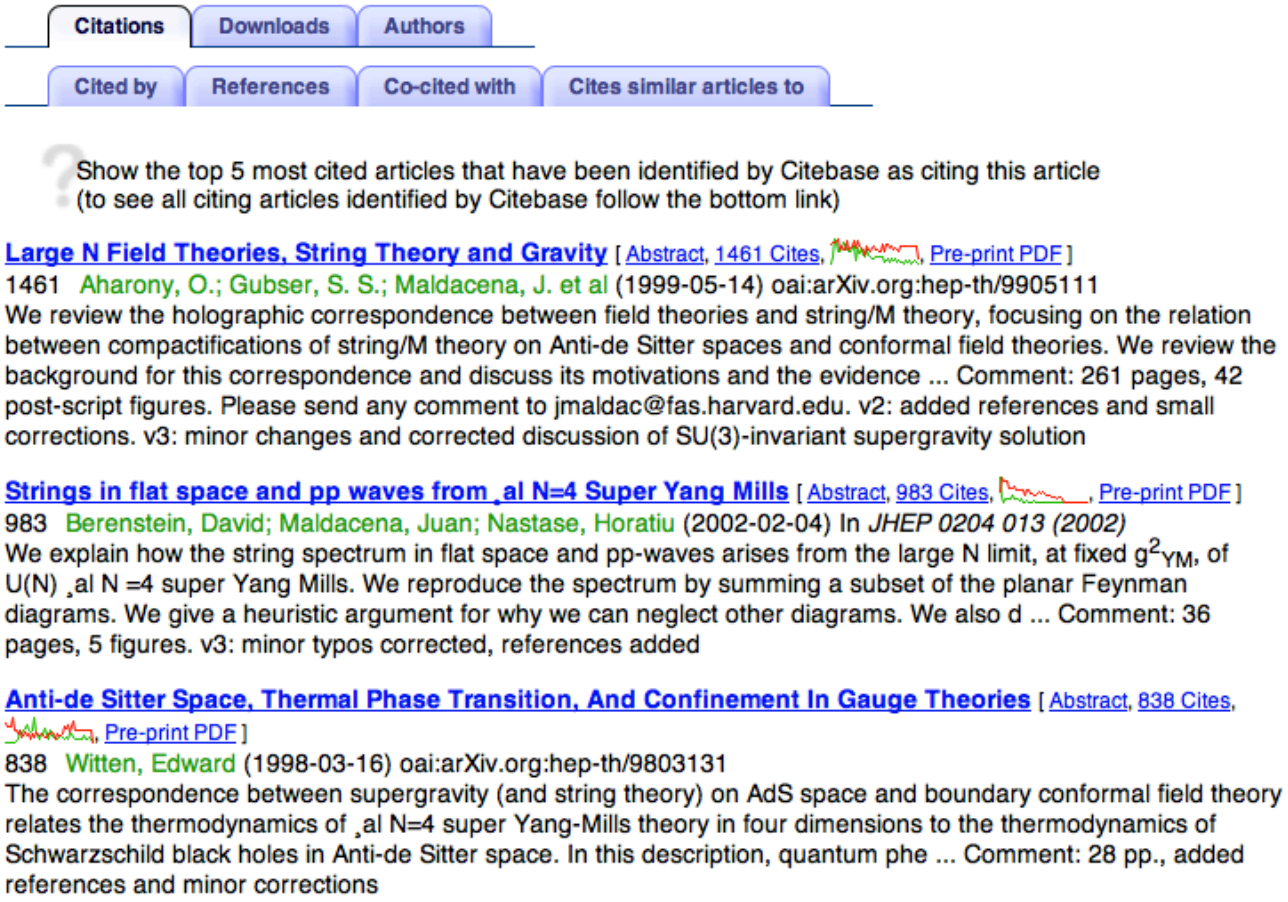
Building Testbed

- Previous solutions have not been general
- Reference Testbed
  - Interdisciplinary collection of documents
  - Wide-ranging exemplars
  - International scope
- A collection of material that is donated and curated by a community of people
  - ie a repository

Final Stages

- Build example services





Final Stages: Linked Citation Data