Scientific communication: evolution or paradigm shift?

Allan Foster

INFORUM, Prague, 23rd May 2006

allan@allanfoster.co.uk
www.allanfoster.co.uk

Communications occasionally go awry …

Source: BBC24 television channel, May 2006

Structure of presentation

- Strains in the existing pattern of scientific communication

- Data management challenges

- Open access: where are we?

- The new information companies

- Social technologies: are they changing the information world?

Scholarly communication: 1970s-

- increasing reliance on journals as main channel for dissemination of scientific knowledge, growth paralleling that of research produced

- dominance of 'reader-pay' or 'library pay', as opposed to 'author-pay' model of journal dissemination

- existence of many publishers in market, with two big groups of publishers, For-profits (FP) and Not-for-profits (NFP - including learned societies and university presses)

- very fast growth of some big FP publishers, through new journal introduction, administering journals for learned societies, and through mergers.

Tensions in traditional pattern of scholarly communications

- Economic & financial change – for buyers, contributors and providers
  - Pressures on universities & funding bodies, incl. selectivity, re-focusing, costing, static library budgets
  - Pressures on 'researcher as author'
  - Pressures on publishers – costs, markets, margins, technology & voluntarism (editing & reviewing)
- Technological change
- Social change

Scholarly communications market now

- Core STM publishing: $7-11B
- 1975-1995 journal prices increased by 300% over inflation rate
- Very imperfect market mechanism
- FP journals are 3 times more expensive than NFP titles
- Positive link between quality of journal and its price

http://digbig.com/4hrnh

INFORUM
2006

EU Survey of … scientific publication markets in Europe: recommendations

- Guarantee public access to publicly funded research results shortly after publication
- Aim at a level playing field in terms of publishing business models
- Extend quality rankings of scientific jnls
- Guarantee perennial access to scholarly journal digital archives

INFORUM 2006

EU Survey … policy recommendations (cont.)

- Foster interoperable tools to improve visibility, accessibility & dissemination

- Promote pro competitive pricing strategies

- Scrutinise future significant mergers

- Promote the development of electronic publications (VAT treatment; public/private partnerships)

Data management: some key issues

- How can the results of publicly funded scientific research be made accessible, optimally, internationally?

- How can we provide suitable integration and linkage between scientific publication and underlying data sets, incl. observational and experimental data?

- How can we encourage the proper curation of important data to ensure long term accessibility?

- How can we best work with partners in this endeavour, including funders, universities, researchers, users and publishers?

Data management issues (cont.)

- Humanities and social scientific data as well as STM
- Legal and IPR issues to be faced in these tasks
- Very mixed picture across European states on advice/good practice on ownership of data, restrictions on use…
- Natural interest by funders and researchers to move on to new areas without sufficient attention to the established research record
- Mainly cultural, sociological and financial factors rather than technical
- Highlight data management as a profession and career

Data curation

- Key to sustainability, 'reproduce-ability' and re-use of reliable and trusted digital resources

- Incentivising good practice: research funders requirements & career recognition of data management outputs

- Reproduce-ability sometimes requires a curation of a 'fixed' rather than dynamic dataset

Data curation (cont.)

- Software and versioning issues
- Need for new accounting model recognising long term costs of data curation and preservation
- European and international ongoing work such as by UK's Digital Curation Centre (DCC) to undertake research and disseminate good practice

Model of virtual infrastructure

The Open Access agenda

- Open access journals
- Self-archiving
- Repositories

BOAI definition of an open access journal

"**free availability** on the public internet, permitting any users to **read, download, copy, distribute, print, search**, or **link** to the full texts of these articles, **crawl them** for indexing, pass them as **data to software**, or use them for any other lawful purpose, **without financial, legal**, or **technical** barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give **authors control over the integrity of their work** and the right to be **properly acknowledged and cited**." (my highlighting)

Budapest Open Access Initiative FAQ: http://digbig.com/4hrbg

Source: DOAJ, www.doaj.org   Screen shot 18/5/2006

JISC's open access programme

- Limited funding to allow authors to experiment with OA publishing
- 6 participating publishers: PLoS, Institute of Physics, J. of Experimental Botany, Inter. Jnl of Crystallography, BMJ, OUP
- 124 authors; 78% would publish in an OA journal again. Factors include:
  - Free access for users
  - Wide exposure and consequential academic discourse
  - Prestige
  - Impact factor
  - Citations
- Publishers: performance positive; OA experience fruitful
- Authors 50:50 split on whether fees should be paid by govt agencies or from research grants

Source: Evaluation of the JISC's open access funding initiative. Key Perspectives Ltd, Spring 2006. See http://digbig.com/4hrfa

Repositories

- Research publications, data and learning/teaching resources
- Institutional *and* disciplinary
- Inter-relationships between institutional repositories and specialist data centres
- "Scientific data is best handled by scientists in data centres" (NERC - UK)
- Open access embraces self-archiving in repositories *and* new business models for publishers

Repositories (cont.)

- Currently, how interested are researchers and institutions in repositories??? (Some hostility)
- Should deposit be a requirement by funders? (NIH, RCUK & Welcome policy directions)
- Once again, cultural change is essential
- **D**igital **R**epository **I**nfrastructure **V**ision for **E**uropean **R**esearch (DRIVER)
    - Initially a common network of 51 existing repositories in 5 countries
    - Collective and enabling service layers
    - Using standards such as OAI-PMH, persistent identifiers and some technology standards (SOA, web services)

The new information behemoths

- Google
  - Google Scholar
  - Google Books
  - Google Co-op
  - Google Trends
  - AdSense: matching adverts to keywords on e-journal page. Payment by click-throughs
- Yahoo!
  - with Microsoft, RLG etc (Open Content Alliance): *Open Library* – cataloguing & digitisation
  - Social technologies – Flickr (photos), Del.icio.us (tags)

Google's superior searching

via a multi-algorithmic approach, recognising value of:

- **Full text**: matching keywords in the query and content

- **Information architecture:** analysing the internal link structure and hand-crafted metadata of each web site

- **Free tagging:** using the links between web sites

Google Trends: comparing searches

Social technologies

- Weblogs
- Tagging
- Wikis – public & private
- RSS & syndication

Weblogs

- as a marketplace of ideas
- as scholarship
- as democracy
- as a challenge to traditional authority
- for news
- as a corporate communications tool
- for understanding corporate reputation

Are information professionals taking blogs sufficiently seriously?

"Rethinking what is a 'local collection' is very necessary in an age in which less and less content starts out in book and journal format and more in raw forms such as weblogs.

In the meantime a growing army of webloggers are performing this function to some degree already themselves, cataloguing the Web from very personal and oftentimes local perspectives.

These people are creating their own 'sweet science' of content organization that's influencing more and more readers and writers worldwide."

John Blossom in Content Blogger   http://shore.com/commentary/weblogs

Peer review v social collaboration

- Are collaborative technologies antithetical to peer review? Can they co-exist when blogs challenge traditional notions of authority?

- Connotea from Nature Publishing: designed for scientists http://www.connotea.org

Connotea: organising, sharing & discovery service for scientists from Nature Publishing Group

What do we know about HE students' use of discovery services?   [INFORUM 2002]

- heavy reliance on generic search engines - the 'Google Factor'!
- want fulltext
- extremely 'quality tolerant'
- simple search strategies
- perplexing wealth of quality (paid-for) web resources, let alone free Internet stuff
- 'satisficing' behaviour ie 'good enough'
- and many academic staff behave in the same way!

All pretty much true in 2006

Source: JUSTEIS, JUBILEE & EDNER
evaluations

Searching and taxonomies

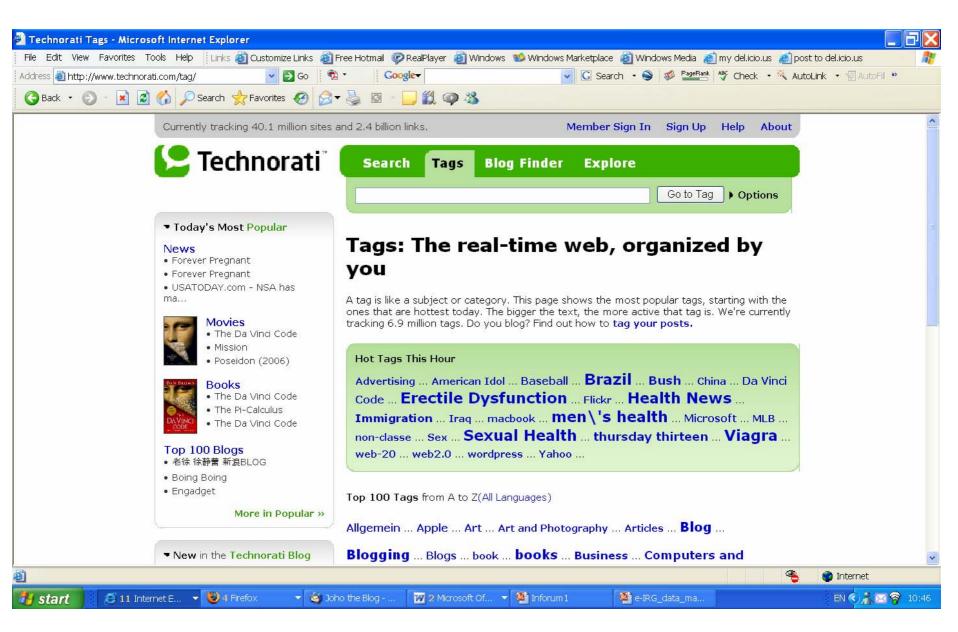Librarians' views of researchers' use of discovery tools:

- variable and idiosyncratic
- past experience constrains choice
- untroubled by different interfaces
- free text searching is norm except in a few areas such as chemistry
- little use of formal thesauri
- the Google inheritance - or has it always been so?
- some major exceptions eg systematic reviews

Source: Allan Foster's interviews with 40 UK academic librarians, May 2006

Searching and taxonomies

- 'Folksonomies': collaborative subject analysis
  - Tagging
  - Tag clouds
- Collabulary – a compromise between top down taxonomies and casual tagging
- A world of messiness – piles of leaves, not trees (David Weinberger)

  (see http://www.hyperorg.com/blogger)

Technorati's home page tag display: 19th May 2006

Factiva's use of tag displays

Summary: some key points

- Academics (and research funders) want more control over their IP (institutional & personal) yet still seek scientific status & recognition through high impact journal publishing

- Data management: more link ups between scholarly publication and original data; greater professionalism in data curation

- Richer & more varied business/publishing models, embracing high quality publishers

- New, powerful entrants to scholarly content access & distribution making their mark

Summary (cont.)

- Open access here to stay but with turbulence
- Rapid, uneven growth for OA journals & repositories (incl. links to scientific datasets) – and interoperable standards
- Social technologies are complementary to quality peer reviewed sources
- Taxonomies, 'folksonomies' and 'messiness' (complexity)

Thanks for listening!

Allan Foster

allan@allanfoster.co.uk

www.allanfoster.co.uk