ARCHIVEPRESS

or
"Diamonds in the rough"

Richard M. Davis, ULCC
Maureen Pennock, BL

ARCHIVEPRESS

or
"Set a blog to catch a blog"

Richard M. Davis, ULCC
Maureen Pennock, BL

ARCHIVEPRESS

or
"**O**pen-source **B**log **A**rchive **M**anagement **A**pplication"

Richard M. Davis, ULCC
Maureen Pennock, BL

Blogs

- Simple diary/journal/log format

- Readers can (usually) comment

- Quick and easy to set up

- Available from many public hosts as free or freemium services (Blogger, WordPress, LiveJournal, Typepad...)

- Personal and institutional hosted installations

- *Different platforms, all support newsfeeds (Atom, RSS)*

Web 2.0 recap

Web 1.0	Web 2.0
Britannica Online	Wikipedia
personal websites	blogging
screen scraping	web services
publishing	participation
content management systems	wikis
directories (taxonomy)	tagging ("folksonomy")

Tim O'Reilly (2006)

Value of blogs

- Primary source, first-hand, personal account (diary, log, journal)

- Research, discussion and outputs (notebook, lab-book)

- E-learning, reflection and discussion (notes, essays, exercise books)

- Project activities - centralised, distributed (reports)

Blogs are evolving and being used for many valuable activities (here we highlight scholarship). Some bloggers spend hours or more on a post. Bill Hooker has an incredible set of statistics about the cost of Open Access and Toll Access publications, page charges, etc. Normally that would get published in a journal no-one reads [...] So I tend to work out my half-baked ideas in public. Some people do their early science in the Open. Some are activists. Some review the current landscape, etc.

Peter Murray Rust (2009)

New genres of publications are becoming increasingly important to participants. For example, blogs are cited as a good window into what expert practitioners are doing. This material is not duplicated in traditional sources, yet it is important to consult:

"This guy has a fantastic blog. He's actually a software architect at Microsoft… and he writes about a lot of issues in data centers…"

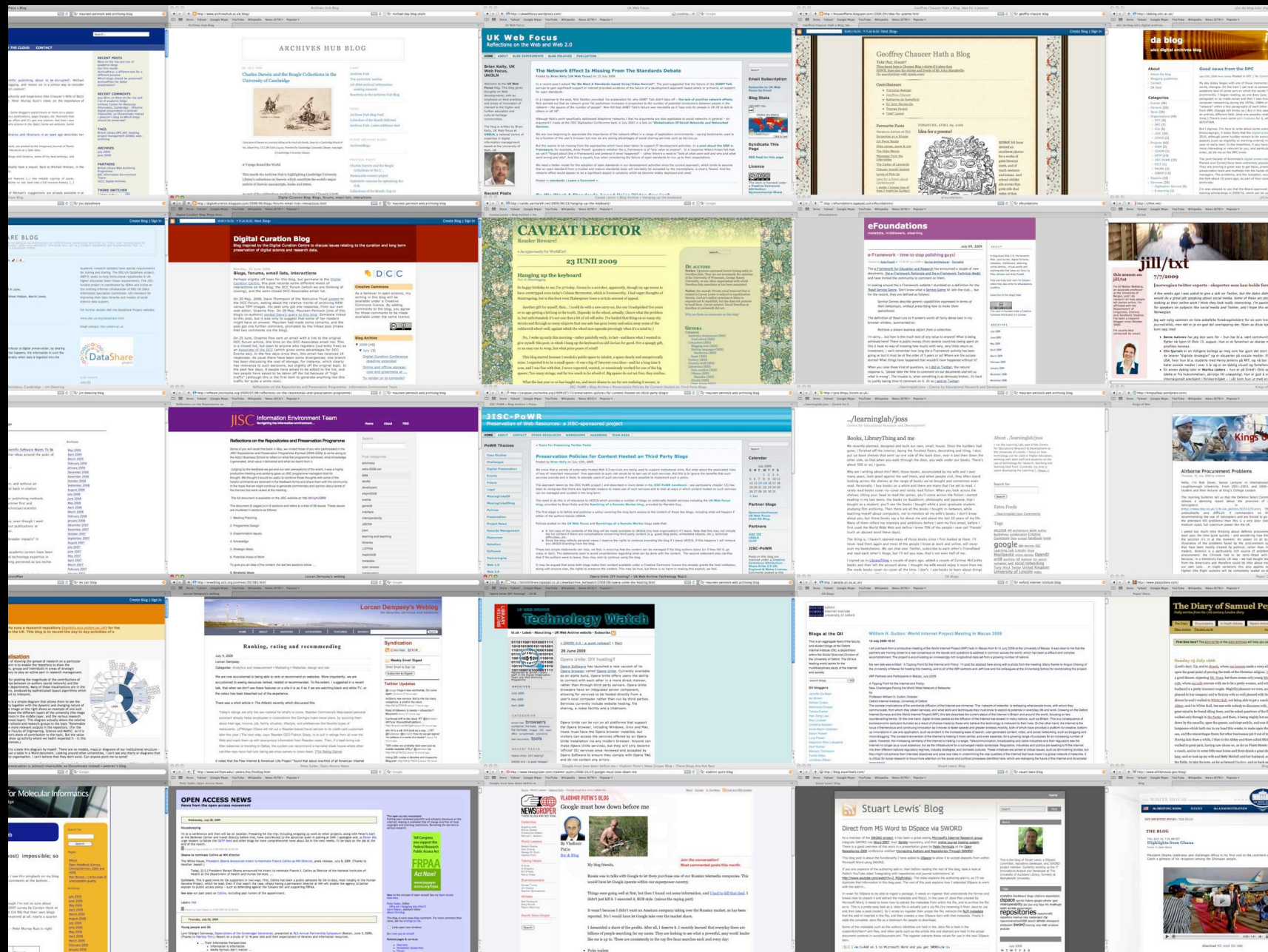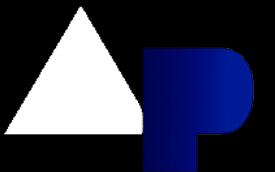Catherine Marshall (2008)

In my opinion, if Peter Suber's focus were the peer–reviewed article rather than blogging, we would be waiting longer for less knowledge about open access. It is hard to keep up with Open Access News; however, without this blog, it would be even harder ...
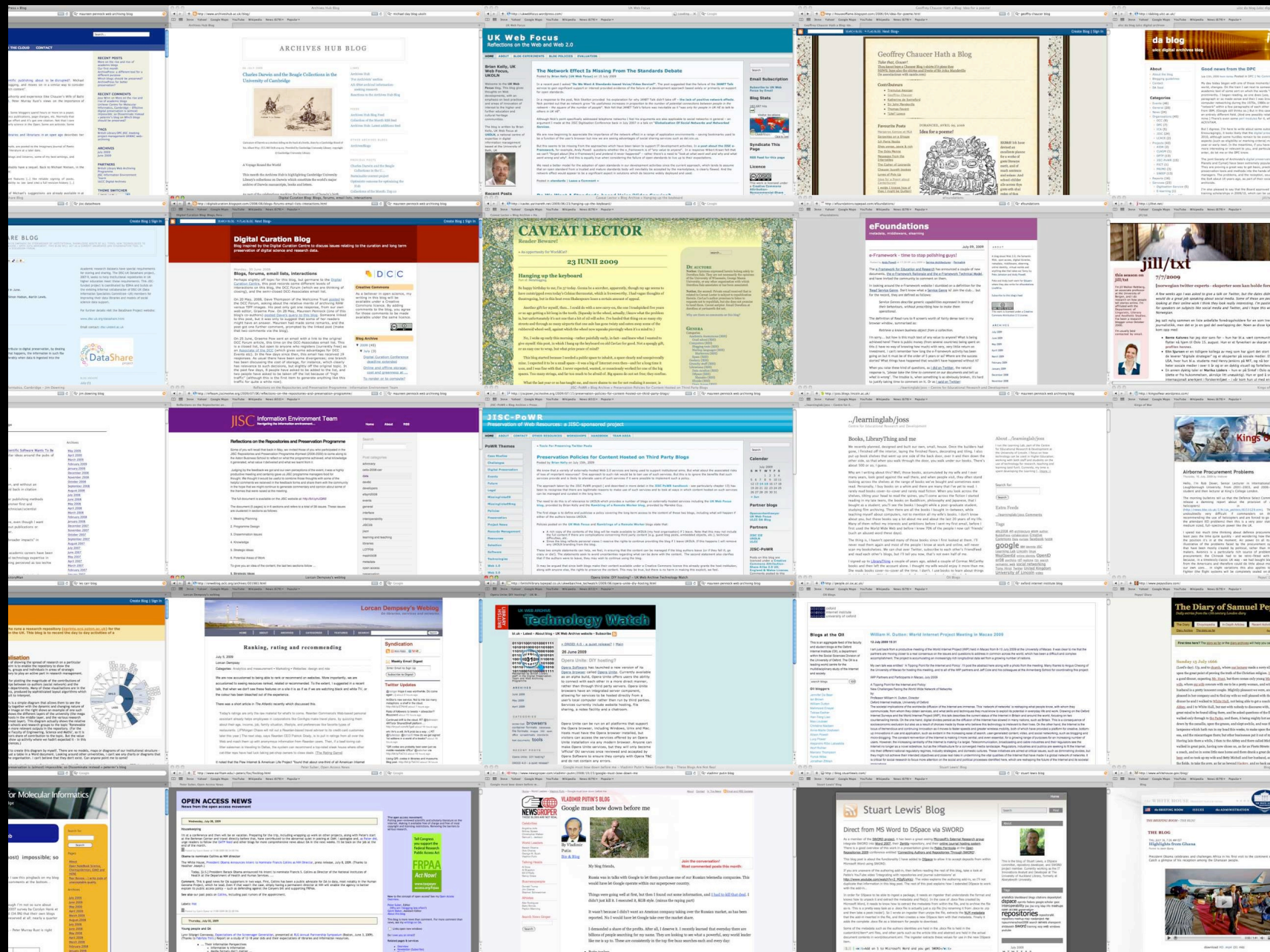
Heather Morrison (2007)

Blogs, scholarly or otherwise

"But the web is being archived isn't it?"

JISC: Preservation of Web Resources (PoWR)

- Web preservation awareness and activities in HE institutions - variable

- Institutional web archiving - perceived barriers:

  - Cost of implementation
  - Complexity of available tools

- Blog archiving

  - Backup and migration
  - Individual's responsibility

http://jiscpowr.jiscinvolve.org/

Institutional web archiving: a wish-list

- Application:

  - Quick and easy to install and maintain

  - Free, open-source, open-standards

  - Widely used, supported, documented, understood

  - Reliable

- Content:

  - Easy to ingest, manage, access, preserve
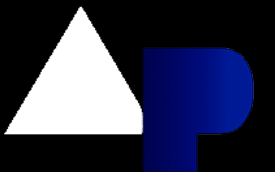
Institutional blog archives

- Part of the institutional record

- Cumulative, automated

- *Selection* - explicitly seeded with blogs of note

- Support *authenticity and fixity* through automation

- *Persistent* - within institutional domain

- *Citable* - through persistent URI for posts

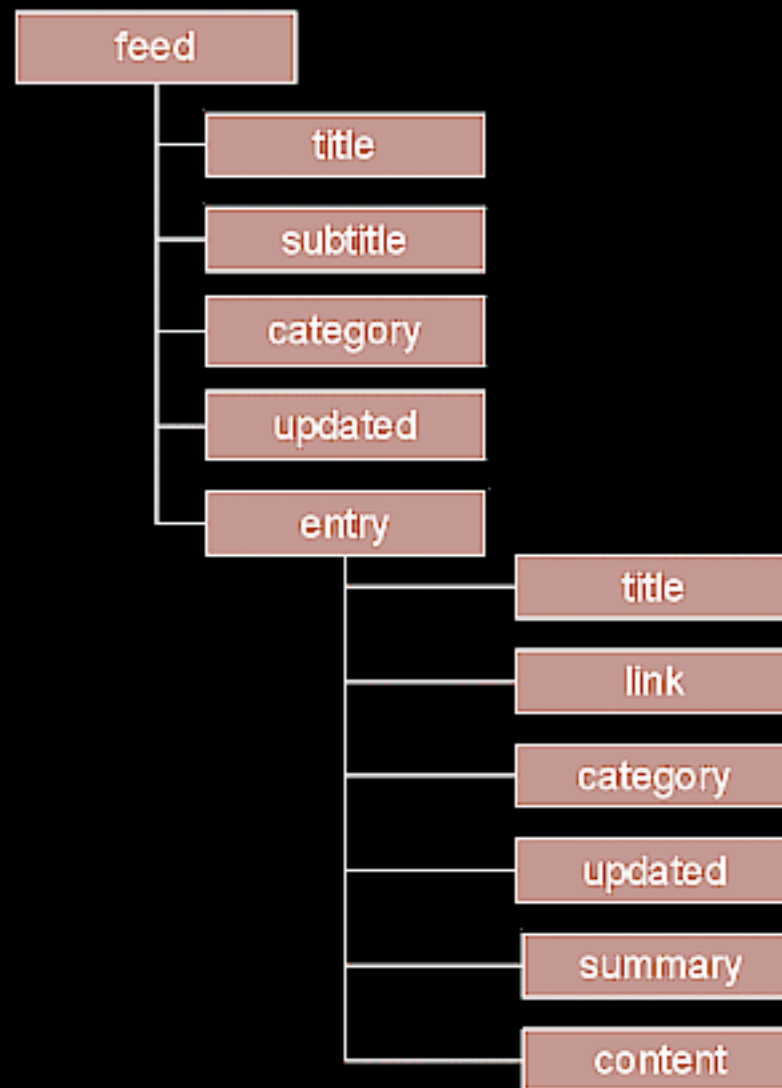- *Agnostic* about location or platform used by original blog

"Use the feeds…"

Blogs as data

Blogs as data

I rarely see any blog's design, since I read through NetNewsWire, so I'm inclined to think blogs represent an area where the content is primary and design secondary.

Chris Rusbridge (2009)

ArchivePress: aims

- JISC-funded, 6 months (June-November 2009)

- Enable easy creation of institutional (or thematic) blog post archives by using a WordPress database and installation to

  - monitor and gather feed content automatically

  - store gathered posts and comments

  - provide access (search, author, date, keyword, etc.)

  - be the focus of ongoing preservation activity

ArchivePress: approach

- Demonstrator/testbed using existing WordPress features and 3rd party plugins

- Test on existing corpora of institutional blogs (DCC, Lincoln University, UKOLN): different platforms, complexity and issues

- Analyse results (metadata, content, sig. props, usability)

- Develop methodology and new plugins as necessary

ArchivePress: challenges

- Non-invasive

- Boot-strapping

- Comments

- Embedded content

- Versions

- Metadata

- Scalability

AP1 Demonstrator

ArchivePress: possibilities

- Thematic collections (e.g. local history projects)

- Other newsfeed-oriented content (e.g. Twitter)

- Islands of linked blog archives

- Incorporation of semantic metadata, linking, microformats

- Text-bases for text-mining

- E-anthologies of blogs, posts, discussions

- Integration with finding aids and discovery tools

References

- Marshall, C. (2008). *From writing and analysis to the repository: taking the scholars' perspective on scholarly archiving.* In: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries, June 16-20, 2008, Pittsburgh PA, USA [doi>10.1145/1378889.1378930].

- Morrison, H. (2007). *Rethinking collections — Libraries and librarians in an open age: A theoretical view.* First Monday, Volume 12 Number 10 - 1 October 2007. Available from: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1965/1841 [Accessed 20th July 2009].

- Murray Rust, P. (2009) *Effective digital preservation is (almost) impossible; so Disseminate instead.* Peter MR's Blog, Unilvever Cambridge Centre For Molecular Informatics, University of Cambridge. Available from: http://wwmm.ch.cam.ac.uk/blogs/murrayrust/?p=2159 [Accessed 20th July 2009].

- O'Reilly, T. (2005). *What Is Web 2.0?* Available from: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html [Accessed: December 18th, 2006].

- Rusbridge, C. (2009) Comment on *Preservation for scholarly blogs.* Gavin Baker: A Journal of Insignificant Inquiry. Available from: http://www.gavinbaker.com/2009/03/30/preservation-for-scholarly-blogs/ [Accessed 20th July 2009].

http://archivepress.ulcc.ac.uk/