

ArchivePress

Posted on 3rd December 2009. Tags: [JISC](#), [ULCC](#), [BL](#), [iDCC](#), [blogs](#)

In the academic and scientific sectors, researchers increasingly use blogs to record their thoughts, activities and insights, and to share and discuss them with colleagues and peers.

Blogs therefore merit preservation as part of the institutional and scientific record. Even as a matter of record-keeping, an institution or research project should capture the content of its researchers' blogs, and consolidate it into one managed, reliable, searchable collection. Benefits of creating and maintaining institutional blog archives include:

- ▲ an authentic record of intellectual activity in an institution
- ▲ a reliable online resource for citation and reference
- ▲ persistently identified resources in the institutional domain
- ▲ a rich text-base for future research and analysis

The information content of blogs is the *sine qua non* of any preservation activity. Typical 'crawler based' web archiving focuses on the HTML rendering of blog pages, but they ignore the underlying rich data structure of blogs. What's more, as the JISC-PoWR project discovered, web crawlers and harvesters are little used in HE institutions, as they require considerable investment of skills and resources.

The ArchivePress project is developing a lightweight, off-the-shelf, blog archiving tool using the open source WordPress blog package.

Since all blogs publish their posts - and usually comments too - using XML-based RSS and Atom newsfeeds, ArchivePress will monitor designated feeds and import new content automatically, along with associated metadata.

The blog rendering is not captured. However many readers of blogs rarely visit blogs' websites, and read the content mainly through third-party RSS readers or aggregators.

The project has three key aims:

- ▲ Creating plugins for WordPress to work as a blog archive
- ▲ Creating demonstrator instances
- ▲ Assessing the effectiveness of the approach and promoting discussion and debate

The project's target group is HE institutions, but it will equally be useful to other organisations - for example, in the cultural heritage sector - wishing to create a collection of blog content around a given theme and to manage it as a single collection.

ArchivePress is funded by the JISC, and undertaken by the University of London Computer Centre, with assistance from the British Library Digital Preservation Department. Plugins will be available to download in early 2010.

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

You can follow any responses to this entry through the [RSS 2.0 feed](#). You can leave a response, or [trackback from your own site](#).

Leave a comment

It's easy to miss the impact of blogs on research, because most science blogs focus on outreach. But more and more blogs contain high quality research content. [...] What we're seeing here is a spectacular expansion in the range of the blog medium. By comparison, the journals are standing still.

Michael Nielsen
<http://tinyurl.com/otydah>

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.

Peter Murray Rust
<http://tinyurl.com/nmrrnx>

[D]esign may be an issue; on the other hand 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
<http://tinyurl.com/yemq2rx>

The discrete "item" — the book, the journal article — is becoming less and less relevant in today's interconnected world. [...] For the library, what this means is that collections work will gradually need to shift from a focus on discrete items, to a focus on comprehensive collections and links both within and outside of collections.

Heather Morrison
<http://tinyurl.com/y9yxytj>

[B]loggers are interested in the preservation of their content (but, not all blog content), and are generally willing participants in their blogs' persistent access and preservation.

Carolyn Hank et al.
<http://tinyurl.com/ykzbbgf>

