Moving Targets:  
the role of web preservation in supporting sustainable citation  

Richard M. Davis  
Kevin Ashley  

Innovations in Reference Management (TELSTAR)  
Open University  
14th January 2010
Moving Targets:
web preservation and sustainable citation

- Why do we cite?
- Web preservation
- Referencing web archives
- Blogs and blog archives
- Citation Repositories
- Institutional approaches
- Information skills
Why do we cite?

- To allow those reading the record of what you’ve done to read the sources you have read.
- To credit, and show you have read, the key relevant work and can use it to support your arguments and so indicate where your work has taken you further.
- Citing and referencing work you have read and used avoids plagiarism.

Citing References: Why and how to do it
University of Edinburgh Information Services
Why do we cite?

• *To allow those reading the record of what you’ve done to read the sources you have read.*
• *To credit, and show you have read, the key relevant work and can use it to support your arguments and so indicate where your work has taken you further.*
• *Citing and referencing work you have read and used avoids plagiarism.*
Web preservation

“a series of managed activities necessary to ensure continued access [...] for as long as necessary”

Digital Preservation Coalition, 2002
Web preservation

Cool URIs don't change

What makes a cool URI?
A cool URI is one which does not change.
What sorts of URI change?

URIs don't change: people change them.

There are no reasons at all in theory for people to change URIs (or stop maintaining documents), but millions of reasons in practice.

In theory, the domain name space owner owns the domain name space and therefore all URIs in it. Except insolvency, nothing prevents the domain name owner from keeping the name. And in theory the URI space under your domain name is totally under your control, so you can make it as stable as you like. Pretty much the only good reason for a document to disappear from the Web is that the company which owned the domain name went out of business or can no longer afford to keep the server running. Then why are there so many dangling links in the world? Part of it is just lack of forethought. Here are some reasons you hear out there:

We just reorganized our website to make it better.

Do you really feel that the old URIs cannot be kept running? If so, you chose them very badly. Think of your new ones so that you will be able to keep them running after the next redesign.
Web archiving

Welcome to the UK Web Archive

Thousands of UK websites have been collected since 2004 and the Archive is growing fast.

Here you can see how sites have changed over time, locate information no longer available on the live Web and observe the unfolding history of a spectrum of UK activities represented online. Sites that no longer exist elsewhere are found here and those yet to be archived can be saved for the future by nominating them.

The Archive contains sites that reflect the rich diversity of lives and interests throughout the UK. Search is by Title of Website, Full Text or URL, or browse by Subject, Special Collection or Alphabetical List.

Explore the Special Collections

Special Collections are groups of websites brought together on a particular theme by librarians, curators and other specialists, often working in collaboration with key organisations in the field. They can be events-based (e.g. The Olympic & Paralympic Games 2012), topical (e.g. The Credit Crunch Collection) or subject-oriented (e.g. The British Countryside Collections).
“Sure, you use the Internet all the time, but you need to wise up to the web when you use it for your university or college work.”

USE THIS FREE INTERNET TUTORIAL TO LEARN TO DISCERN THE GOOD, THE BAD AND THE UGLY FOR YOUR ONLINE RESEARCH.

Launched 13th June 2006 - read the press release
Beacon cited in fog

Beacon cited in fog

Beacon cited in fog

Auricle: The Case Of The Disappearing E-learning Blog
Posted by Brian Kelly on September 1st, 2008

The Auricle E-Learning Blog

The e-learning team at the University of Bath was one of the early adopters of blog technologies to provide a forum for reflecting on e-learning in a Web context. The blog was set up by Derek Morrison when he was head of the e-learning unit. Derek had an interest in exploring the potential of new technologies, with one example of this being the series of podcast interviews he recording and made available on the blog back in 2005. This included an interview with John Dale about the innovative blogging service developed at the University of Warwick (the first large scale student blogging service in the UK) and, perhaps not as noteworthy, an interview with me on my reflections of the WWW 2005 conference.

The name of the e-learning team's blog was Auricle, which has an advantage of being a very Google-friendly name, and a Google search for "Auricle Bath" finds links to the blog itself and various pages which refer to the blog. Unfortunately it seems that the blog no longer exists - following a link to the blog's home page gets a 404 error message:

The web address http://www.bath.ac.uk/dacs/cdnt/pMachine/mornblog.php was not found. It may have moved, or it may no longer be available.
Blogs
Blogs are evolving and being used for many valuable activities (here we highlightscholarship). 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/nmnrnx
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
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/y9yxtvj
Welcome to the ArchivePress website

ArchivePress is a blog-archiving project being undertaken by the University of London Computer Centre and the British Library Digital Preservation department, funded by the JISC Information Environment Programme under its Rapid Innovation Grants Call (03/09).

The project will explore practical issues around the archiving of weblog content, focusing on blogs as records of institutional activity and corporate memory. As an alternative to the web crawling/harvesting approach of the Internet Archive and the UK Web Archive, ArchivePress will test the viability of using RSS feeds and blog APIs to harvest blog content (including comments, embedded content and metadata). The archived content will be stored and managed using instances of Wordpress, thereby maintaining the blogs’ native data structures, formats and relationships.

We hope to develop tools and methodology that will enable organisations to use simple, free, open source blogging software to manage a central archive of designated institutional blog outputs, even if they are spread over different blog hosts and platforms. The benefits of this approach will include:

Recent Posts
- ArchivePress literature review
- No Comment/Make it Easy
- iPRES paper: ArchivePress project & progress
- Doing yet more with plugins
- Data and views and views of data

Recent Comments
- online stock trading guru on Data and views and views of data
- Jack Kingston on No Comment/Make it Easy
- Sarah on Doing yet more with plugins
- Dublin Core Plugin (and others) – WP and Friends on Doing yet more with plugins
- Owen Stephens on Doing yet more with plugins
ArchivePress

AP1 Blog Archive Demonstrator

Add new blog to archive

- Name
- Feed
- Elements to archive
  - Posts
  - Comments
  - Embedded images

Archived blogs

- Name
- Feed
- Last updated
  - The DCC Blawg
  - Digital Curation Blog
  - Research Data Management Forum

Update now
New issue of IJDC
Posted in Digital Curation Blog on October 21st, 2009 at 5:59pm by Chris Rusbridge

The latest issue (volume 4, issue 2) of the International Journal of Digital Curation is now available. It's a bumper issue, with two letters to the editor (a whiff of controversy there!), 8 peer-reviewed papers (originating from last year's International Digital Curation Conference), and 6 general articles (two of which came from last year's iPRES06 conference). I'm really pleased with this issue, which as always is extremely interesting.

[Cite this post: Harvard style]


Post metadata

<table>
<thead>
<tr>
<th>Post title</th>
<th>New issue of IJDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post author</td>
<td>Chris Rusbridge</td>
</tr>
<tr>
<td>Post date</td>
<td>20091021175900</td>
</tr>
<tr>
<td>Originating Blog</td>
<td>Digital Curation Blog</td>
</tr>
</tbody>
</table>

Blogs in this archive
- Digital Curation Blog
- Research Data Management Forum
- The DCC Blawg

Calendar

<table>
<thead>
<tr>
<th>November 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>21</td>
</tr>
<tr>
<td>27</td>
</tr>
</tbody>
</table>

< Oct

Authors
- Chris Rusbridge
- DataShare
- Graham Pryor
- Mags McGeever
Welcome to ArchivePrints

This is a repository of blog posts by researchers at this university.

- To support persistent citation of blog posts by our researchers (See Citation)
- To provide persistent identifier for blog content
- To encourage citation of this canonical and reliable source of blog posts
- To preserve this university's blogging activities as part of the institutional record

For further information is available on Citation, Copyright, Preservation.

Latest Additions
View items added to the repository in the past week.

Search Repository
Search the repository using a full range of fields. Use the search field at the top of the page for a quick search.

Browse Repository
Browse the items in the repository by subject.

About this Repository
More information about this site.
ArchiveDSpace Blog Repository

This is a repository of blog posts by researchers at This University. Its purpose is:

- To support persistent citation of blog posts by our researchers
- To provide persistent identifier for blog content
- To encourage citation of this canonical and reliable source of blog posts
- To preserve This University’s blogging activities as part of the institutional record

Further information is available on Citation, Copyright, Preservation.

Search

Enter some text in the box below to search DSpace.

[Go]

Communities in ArchiveDSpace

Choose a community to browse its collections.
Digital Archive for Chinese Studies
Leiden Division

Citation Repository

2009

- Van Crevel/Tan/Hockx 2009
  Text, Performance, and Gender in Chinese Literature: Essays in Honor of Wilt Idema (Leiden/Boston: Brill, 2009)

2008

- Van Crevel 2008
  Chinese poetry in times of mind, mayhem and money (Leiden: Brill, 2008)

- Wielander 2008

Created: 12 May 2009  Last update: 23 Nov 2009
DACHS - Digital Archive for Chinese Studies

Our aim is archiving and making accessible Internet resources relevant for Chinese Studies, with special emphasis given on social and political discourse as reflected by articulations on the Chinese Internet.

About DACHS
For further information on DACHS, our collection policy, legal issues or the technical infrastructure you might want to take a look at the Introduction.

Access
To access the collection use the box on the right or get to the Metadata search (simple, advanced or recent downloads). For a short overview of the collection take a look at the table of contents. The login process does not work correctly with Mozilla browsers. Therefore, until we can solve this problem we recommend to use other browsers, i.e. Internet Explorer. More information on how to access the collection is given here.

Newsletter
Read ChinaResourceNews for descriptions of the latest acquisitions in DACHS and ChinaResource.org.

Publications
DACHS was last presented at the JISC, DPC and UK Web Archiving Consortium Workshop "missing links: the enduring web", London, 21 Jul 2009. Take a look at the
### Search result

**You searched for "ughur"**

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Material</td>
<td>Computer file</td>
</tr>
<tr>
<td>Author</td>
<td>Chung Chien-peng</td>
</tr>
<tr>
<td><strong>Title</strong></td>
<td>China's &quot;war on terror&quot;: September 11 and uighur separatism : author's postscript, September 2002 / by Chien-peng Chung</td>
</tr>
<tr>
<td>Source</td>
<td>Foreign Affairs (Sep. 2002)</td>
</tr>
<tr>
<td><strong>Orig. URL</strong></td>
<td><a href="http://www.foreignaffairs.org/articles/chung.html">http://www.foreignaffairs.org/articles/chung.html</a></td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>1 HTML file</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td>eng.</td>
</tr>
<tr>
<td><strong>Loc. note</strong></td>
<td>Full text of Chung's article 'China's 'war on terror': September 11 and Uighur separatism&quot; available at: <a href="http://www.sino.uni-heidelberg.de/archive/documents/terrorism/chienpengchung030203a.htm">http://www.sino.uni-heidelberg.de/archive/documents/terrorism/chienpengchung030203a.htm</a>.</td>
</tr>
<tr>
<td>Subjects</td>
<td>Terrorism - United States - 2001</td>
</tr>
<tr>
<td></td>
<td>Terrorism - China - 2001</td>
</tr>
<tr>
<td><strong>Material</strong></td>
<td>Computer file</td>
</tr>
<tr>
<td><strong>Author</strong></td>
<td>Mori, Yesuhiro</td>
</tr>
</tbody>
</table>
Information literacy

WHAT YOU NEED TO KNOW

bandits on the information superhighway

- Protect your privacy
- Avoid scams
- Buy & sell wisely
- Meet strangers safely
- Know your rights

Daniel J. Barrett
O'Reilly & Associates, Inc.
New kinds of collection
Still need convincing?
Still need convincing?
"The time to debate the role of blogs in legal scholarship has passed. As we approach the end of the first decade of the 21st century, one of our oldest and most conservative disciplines has clearly embraced the era of electronic publishing. Blogging has indeed transformed legal scholarship. Now it's time to move the dialogue forward."

Georgetown University, Dec 2008

http://www.ll.georgetown.edu/ftls/
“Unlike e.g. a journal, a blog is often (and certainly in this case) the effort of an individual. Journals on the other hand can last for centuries (see for example this link to the ToC of the world’s oldest scientific journal that has been in continuous publication for 355 years!). So how long should/can a blog last?

The last question leads on to whether blogs should be archived or curated in a larger sense?”

Henry Rzepa muses on his blog

http://www.ch.ic.ac.uk/rzepa/blog/?p=1459
A reminder

- Citation: let others read your sources
- Author achieves this by publishing in the right place
  - publishers, libraries make the rest happen
- Citer uses recognised location-independent mechanism
- Want something as simple for blogs
Even the ephemeral is citable

Although tables were set up to record details of reports on schools by the Office for Standards in Education (the Ofsted_visit table), and the types of activities which speakers on GM issues were prepared to undertake (the Speaker_activity table), no data was entered into these tables, which were empty when the dataset was transferred to NDAD. [9]

From GM Schools: http://www.ndad.nationalarchives.gov.uk/CRDA/36/detail.html
Even the ephemeral is citable

Although tables were set up to record details of reports on schools by the Office for Standards in Education (the Ofsted_visit table), and the types of activities which speakers on GM issues were prepared to undertake (the Speaker_activity table), no data was entered into these tables, which were empty when the dataset was transferred to NDAD.

[9]

From GM Schools: http://www.ndad.nationalarchives.gov.uk/CRDA/36/detail.html
Even the ephemeral is citable

Although tables were set up to record details of reports on schools by the Office for Standards in Education (the Ofsted_visit table), and the types of activities which speakers on GM issues were prepared to undertake (the Speaker_activity table), no data was entered into these tables, which were empty when the dataset was transferred to NDAD.

[9]

From GM Schools: http://www.ndad.nationalarchives.gov.uk/CRDA/36/detail.html

9. Dataset Documentation Catalogue, reference CRDA/36/DD/1/2/1, pp. 5, 12, 17, 22; email of 22 March 2000 from NDAD to the DfEE; compliments slip accompanying floppy disk received from the DfEE on 27 March 2000.
Although tables were set up to record details of reports on schools by the Office for Standards in Education (the Ofsted_visit table), and the types of activities which speakers on GM issues were prepared to undertake (the Speaker_activity table), no data was entered into these tables, which were empty when the dataset was transferred to NDAD.

9. Dataset Documentation Catalogue, reference CRDA/36/DD/1/2/1, pp. 5, 12, 17, 22; email of 22 March 2000 from NDAD to the DfEE; compliments slip accompanying floppy disk received from the DfEE on 27 March 2000.
Solutions

- Institutions can provide mechanisms
- Authors need to use them
- Blogs need to automatically expose citable link
- “Permalinks” with a bit more “perm”
Solutions

• Institutions can provide mechanisms
• Authors need to use them
• Blogs need to automatically expose citable link
• “Permalinks” with a bit more “perm”
Credits

Screenshots

• UK Web Archive: http://www.webarchive.org.uk
• Intute Internet Detective: http://www.vts.intute.ac.uk/detective/
• ArchivePress: http://archivepress.ulcc.ac.uk
• DACHS at Leiden University: http://leiden.dachs-archive.org/citrep/
• DACHS at Heidelberg University: http://www.sino.uni-heidelberg.de/dachs/

Photographs

• It’s elemental by Desiree Delgado on Flickr. http://www.flickr.com/photos/desireedelgado/3273760287 CC:BY-NC-ND
• geneperm by CounterBreak on Flickr http://www.flickr.com/photos/gene-aquino/3446137574/ CC:BY-NC-ND