PoWR: Explaining Web Preservation

Kevin Ashley, ULCC
What might be kept?

- Information content
- Information appearance
- Information behaviour
- Information relationships
- Change history
- Usage history
Content preservation

An ‘at the event’ report on the first JISC PoWR workshop held at Senate House Library, London on Friday 27th June 2008 has been published in the recent Ariadne Web Magazine (issue 56, July 2008). The piece, written by Stephen Emmott, concluded:

The challenges are significant, especially in terms of how to preserve Web resources. No doubt the institutional repository will play a role. Arguably, the absence of a solution to the preservation of Web resources leads to either retention or deletion, both of which carry risks. The workshop’s core message to practitioners was therefore to start building an internal network amongst relevant practitioners as advice and guidance emerge.

My thinking about this matter was certainly stimulated and I look forward to the next two workshops, and the handbook that will...
Preserving Appearance

**Ariadne Report on the JISC PoWR workshop**

Posted by Marieke Guy on September 2nd, 2008

An 'at the event' report on the first JISC PoWR workshop held at Senate House Library, London on Friday 27th June 2008 has been published in the recent *Ariadne Web Magazine* (issue 56, July 2008). The piece, written by Stephen Emmott, concluded:

> The challenges are significant, especially in terms of how to preserve Web resources. No doubt the institutional repository will play a role. Arguably, the absence of a solution to the preservation of Web resources leads to either retention or deletion, both of which carry risks. The workshop's core message to practitioners was therefore to start building an internal network amongst relevant practitioners as advice and guidance emerge.

> My thinking about this matter was certainly stimulated and I look forward to the next two workshops, and the handbook that will result. Web preservation is an issue which was always important.
Preserving Behaviour
Other things to preserve

• Relationships:
  – links behave
  – associated metadata survives
  – Styles and content stay related

• Usage/change logs: obvious what they are, but not whether they are needed
Techniques

- Save within the authoring system or server
- Save appearance at the browser
- Harvest content with crawlers
Capturing on the server

- Easy (?) if it’s your server
- Captures raw information, not presentation
- May be too dependent on authoring infrastructure or CMS
- Works in short to medium term, for internal purposes
- Not good for external access
Capture post-rendering

• You get what you see: but you don’t know why
• It’s relatively simple for well-contained sites
• Commercial tools exist
• Treats web content like a publication: frozen
• Loses behaviour and other attributes
Harvesting

- Most widely-used
- Presents many problems for capture – often don’t get everything (or too much)
- Defers some access issues:
  - Link re-writing
  - Embedded external content: from archive or live?
- Lots of work, tools and experience
When?

• What triggers things ?
• A regular schedule (yearly, monthly, termly....)
• When stuff changes (regular crawls, but throw away unchanged content)
• Manual initiation
• Intelligent agents
• Transactions