

# Preservation of Web Resources: The JISC PoWR Project

Brian Kelly  
UKOLN  
University of Bath  
Bath, UK

## Acceptable Use Policy

Recording/broadcasting of this talk, taking photographs, discussing the content using email, instant messaging, blogs, etc. is permitted providing distractions to others is minimised.


**Co-authors:** Marieke Guy, Kevin Ashley, Richard Jones, Ed Pinsent and Jordan Hatcher

Resources bookmarked using 'ipres-2008' tag 



UKOLN is supported by:



This work is licensed under a  
Attribution-NonCommercial-ShareAlike  
2.0 licence (but note caveat) 

# The JISC PoWR Project

The JISC PoWR (Preservation of Web Resources) project:

- Funded by the JISC
- Short term project (April – Sept 2008)
- Funded to run workshops and provide handbook giving advice to institutions on institutional Web site preservation
- Provided by UKOLN and ULCC
- Uses a blog as its user engagement & dissemination channel. See: <http://jiscpowr.jiscinvolve.org/>



# “Am I Bovvered?”

Risks identified in joint UKOLN/ULCC’s submission for the JISC PoWR project:

- Institutions wouldn’t be sufficiently interested in the preservation of Web resources
- The complexities (technical, policy, resourcing, legal, ...) would be sufficient to de-motivate institutions



~~SOME RIGHTS RESERVED~~





University Archivist, Records  
Manager and FOI Co-ordinator  
Lizzie Richmond

- Archivist
- Background in collection cataloguing and archival administration and conservation
- Paper environment
- Responsible to the archives - keep them safe and accessible for the future

Head of Web Services  
Alison Wildish



- Web specialist
- Background in information technology, web design and development, communication and marketing
- Digital environment
- Responsible to the user - keep things up to date and useful

Acknowledgements to Wildish and Richmond:  
<<http://www.slideshare.net/jiscpowr/jiscpowr-wildish>>

Marieke Guy and Brian Kelly (UKOLN):

“ We’re doing these workshops on **Web Preservation** and wondered if you’d be willing to give us a case study about the approach from the University of Bath... ”



University Archivist, Records Manager and FOI Co-ordinator

Oh no... not this again!

Why me? This sounds technical... I'm a paper person

I have enough trouble trying to preserve hard copy records without having to worry about the web

I can see the value in theory, but in practice it's too huge

I guess it might be a good idea, but no one much cares what I think

I am interested though...

# Now and the future



Head of Web Services

EEEEEEEEEEEEKKKKKKKK!!!

In all honesty it isn't interesting to me...

We struggle to keep the site current - never mind thinking about preserving the old stuff

I am future watching... need to know what to bring in not how to keep hold of the past

Why is it something I should think about now?

I'm not really that interested

# Motivating The Techies

**Lizzie Richmond used the University of Bath prospectus to illustrate the importance of preservation an institution's memory:**



Ashley Down { 21 Gas Co  
381 Old Market

THE COLLEGE OF TECHNOLOGY  
BRISTOL

GENERAL INFORMATION

RELATING TO

Full-Time and Part-Time  
Courses of Study

FOR DETAILED INFORMATION  
SEE SEPARATE DEPARTMENTAL  
PROSPECTUSES

---

SESSION 1953-54

1953



**BRISTOL**

**COLLEGE OF SCIENCE & TECHNOLOGY**

**Prospectus**

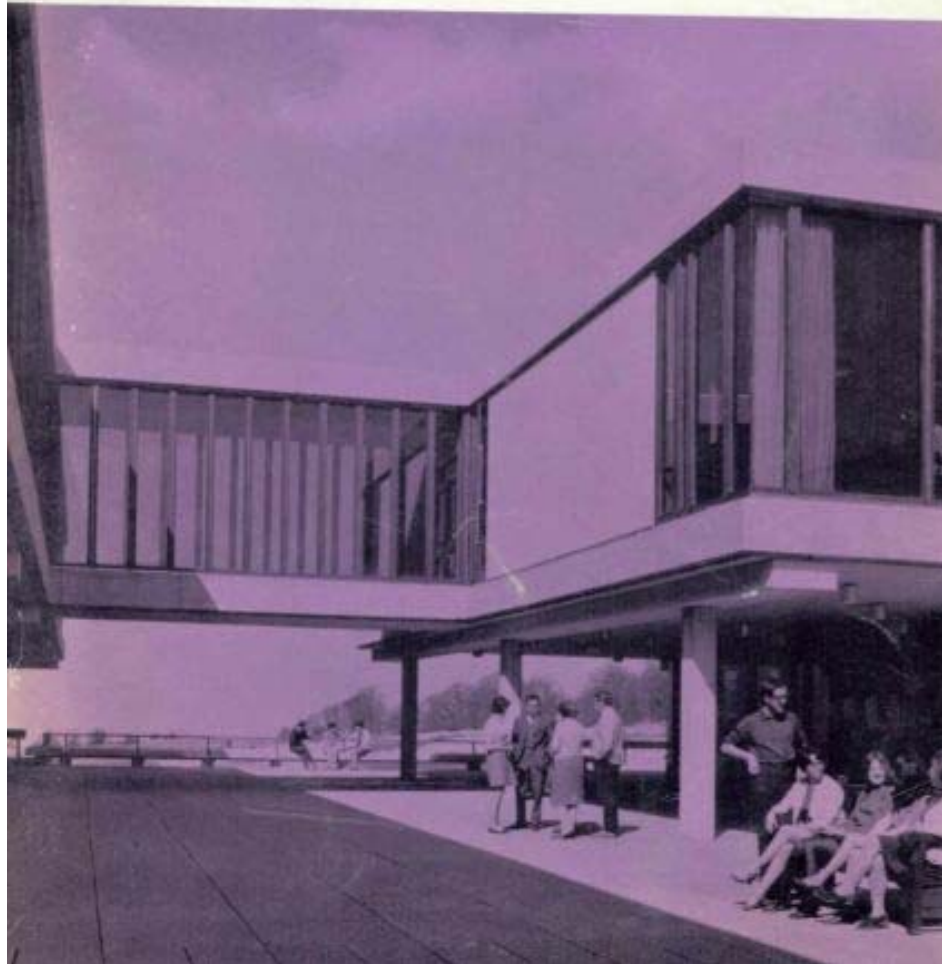
**1960-1961**

**1960**

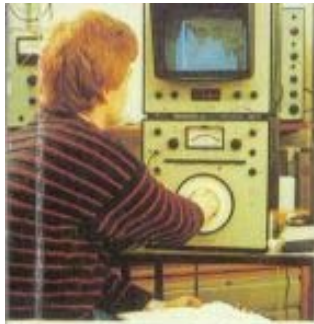


Bath University of Technology  
**Prospectus 1970/71**

990 Lee



1970



**University  
of Bath  
Prospectus  
for entry  
October 1985**

UCCA Code: Bath 05



1985

# University of Bath



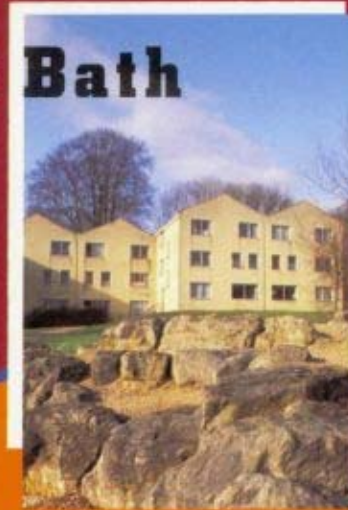
Prospectus for entry October 1991

UCCA Code Bath 05

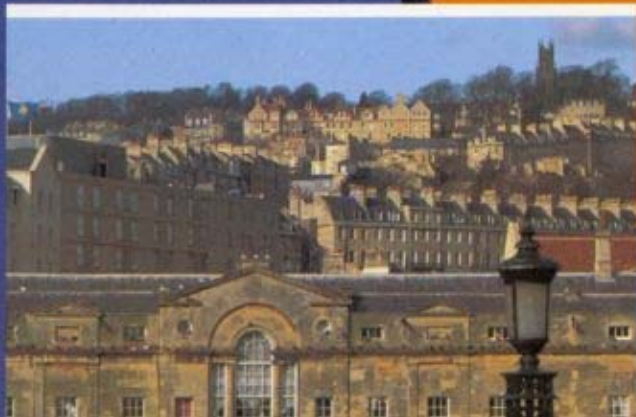
1991



# University of Bath



**1994 ENTRY**



**UNDERGRADUATE PROSPECTUS**

**1994**



Undergraduate  
Prospectus 1999



UNIVERSITY OF  
**BATH**

1999

# University of Bath



Undergraduate Prospectus 2001



UNIVERSITY OF  
**BATH**

2001



UNIVERSITY OF BATH  
Undergraduate Prospectus 2004



UNIVERSITY OF  
**BATH**

2004

The  
**Bath** Prospectus  
Undergraduate

08



[www.bath.ac.uk](http://www.bath.ac.uk)



UNIVERSITY OF  
**BATH**

2008

# The Web Equivalent

**What is the Web equivalent of the history of the printed prospectus?**

**Perhaps the history of the University's home page**

### A History of The University Of Bath Home Page

Produced by the JISC funded PACT project, which is one of a series of best practices for Web site preservation.

About this information: The 12-month course is a snapshot of the University of Bath Home Page from 1996 to 2007. Each archive is available to view on the PACT website.

[See PACT project website](#)

Bath, 2001



PicLens™ view for RSS  
 About This Presentation  
 This presentation provides a snapshot of the University of Bath home page, from 1997 to 2007 based on data held in the Internet Archive's WayBack machine

re.org/

Discover

Researching Accessibility Resources

Open Day

zotero

The University of Bath Home Page - Mozilla Firefox

File Edit View History Simpy Bookmarks Tools Help del.icio.us

http://web.archive.org/web/20000301/

University of Bath  
 BATH  
 BA2 7AY  
 United Kingdom

Telephone +44 (0)1225 826826

UNIVERSITY OF BATH

PROSPECTUSES SEARCH A-Z INDEX

NEWS

- University wins ?1.1 million for Reach-out Activities
- New Year's Honours for Bath
- NEW! Creative Arts - What's On? to June 2000

Information about

- The University & Bath Area
- Research & Study
- Departments & Faculties
- Library & Computing
- Services and Facilities
- Open Days
- Prospective Students
- International Students
- Business & Industry

Information for

- Contacting People
- Accommodation
- News & Events
- Expertise in the University
- Conferences
- Graduates & Alumni
- Prospective Staff
- Visitors & the community

Maintained by web support. Last update: 21 February 2000. Photos ? University of Bath IPU

http://web.archive.org/web/20000302153600/http://www.bath.ac.uk/internal/staff/

zotero

2 Mar 2000

UNIVERSITY OF BATH World Heritage City

City & Bath. Open days.

Study & Research

Computing Services

Campus Life & Student Services

Creative Arts, Sports & Recreation

Hoburne Museum

News & What's On

Bath, 2000

UNIVERSITY OF BATH

UNIVERSITY OF BATH

Prospectuses Search A-Z Index

New Year Honours of the University of Bath

University of Bath wins Queen's Prize

21.1 Million research grants to University of Bath 2000

Welcome

University of Bath BATH BA2 7AY

Bath academics help ch Pharmacy & Pharmacol

Are you a prospective student?

Research

People & Departments

Conferences & Events

Holiday Opening Times

Search

Information for

- Prospective Students
- International Students
- Business & Industry
- The University & Bath Area
- Research & Study
- Departments & Faculties
- Services, Facilities, Arts & Sport

Information about

- Home Page
- Research
- People & Departments
- About us

20012004045http://www.bath.ac.uk/interdept/

200124700122http://web.nal.bath.ac.uk/

# What We Did

We:

- Captured screen images from the Internet Archive of the home page since 1997
- Used FireFox Piclens extension to produce an interactive gallery of the images
- Created a video with commentary providing reflections on the changes to the home page
- See <http://www.ukoln.ac.uk/web-focus/experiments/experiment-20080612/>

Why?

- To draw parallels with the real world example
- As a scenario for first workshop (it's the University's anniversary)
- To illustrate one approach – use of a third party service (Internet Archive)
- To illustrate preservation of the user experience (as opposed to the underlying data)



# Feedback From First Workshop

“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 but now grows increasingly urgent.”

*Preservation of Web Resources: Making a Start,*  
Stephen Emmott, Ariadne (56) Jul 2008 →



# Preservation And Web 2.0

## JISC PoWR Workshop 2: Preservation and Web 2.0

Posted by [Brian Kelly](#) on 31st July 2008

The **second JISC PoWR workshop** was held on 23<sup>rd</sup> July 2008 as part of UKOLN's annual institutional Web management workshop, **IWMW 2008**.

This workshop provided an opportunity to review the outcomes of the **first workshop**, in which members of the JISC PoWR team and the 30+ participants identified some of the challenges to be faced in preserving content held on institutional Web services and explored some of the ways in which these challenges can be addressed. The slides for this review are **available on Slideshare** and are embedded below.

### [Web Preservation in a Web 2.0 Environment](#)

The screenshot shows a presentation slide with the following content:

- URL: <http://www.ukoln.ac.uk/web-focus/events/workshops/webmaster-2008/>
- Title: **Web Preservation in a Web 2.0 Environment**
- Author: Brian Kelly, UKOLN, University of Bath, Bath, UK. Email: [B.Kelly@ukoln.ac.uk](mailto:B.Kelly@ukoln.ac.uk)
- Boxed text: **About This Talk**  
Will use of Web 2.0 services lead to new preservation concerns? And how should we respond to these new challenges?
- Resources bookmarked using 'iwmw2008' tag
- UKOLN is supported by: JISC, MLA, and University of Bath.
- Licensing: This work is licensed under a Attribution-NonCommercial-ShareAlike 2.5 licence (but note caveat)
- Navigation: share, embed, back, forward, search, and user icons.
- Page: 1 / 20

The JISC PoWR project explicitly sought to engage with the preservation implications of Web 2.0

The project has used blogs and wikis to support its work



# Is Web 2.0 Different?

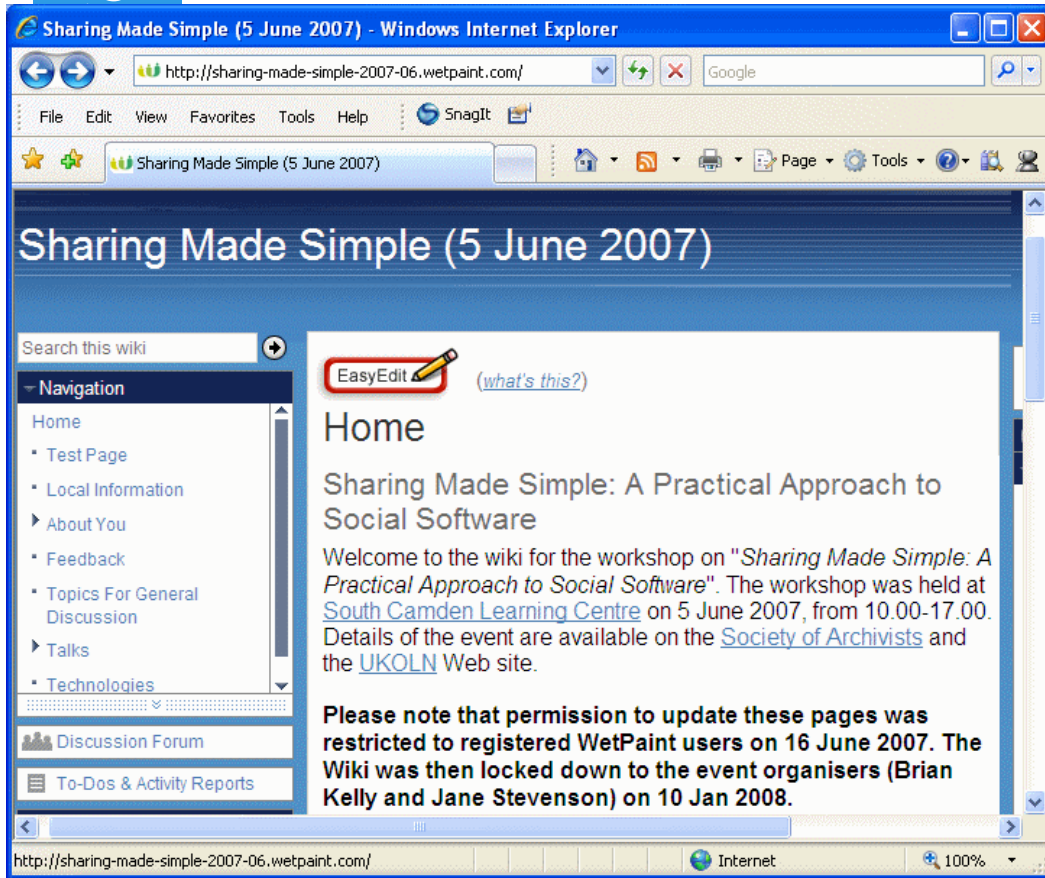
Implications of Web 2.0 for Web site preservation:

- Use of 3<sup>rd</sup> party services (*'network as platform'*)
- Content  $\Rightarrow$  collaboration and communication
- Richer diversity of services (not just a file on a filestore/CMS/database)
- More complex IPR issues

Let's look at:

- Case study 1 - Wikis
- Case study 2 – Blogs
- Case study 3 – Reusing data
- Case study 4 – Disposable data
- Case study 5 – Slideshare

# Case Study 1: A Public Wiki



WetPaint wiki used to support various workshops →

Approaches taken:

- Open access to all prior to & during event (to minimise barriers to creating content)
- Access restricted to WetPaint users after event
- Access later restricted to event organisers

Many aspects of Web site curation are to do with implementing such best practices, rather than implementing technical solutions

See JISC PoWR blog post →

# Case Study 2a: Blog Migration

A Backup Copy Of This Blog « UK Web Focus - Windows Internet Explorer

http://ukwebfocus.wordpress.com/2007/07/19/a-backup-copy-of-this-blog/ oclc lorcan dempsey

HOME ABOUT BLOG EXPERIMENTS BLOG POLICIES EVALUATION

**Brian Kelly, UK Web Focus**

Welcome to the **UK Web Focus** blog. This blog gives my thoughts on Web developments, with an emphasis on best practices and areas of innovation of interest to the higher and further education and cultural heritage communities.

**« Video Streaming of Plenary Talks at IWMW 2007 Talk on "Globalisation of Social Networks" »**

## A Backup Copy Of This Blog

Posted by **Brian Kelly (UK Web Focus)** on 19 July 2007

Casey Leaver has **raised an interesting issue** on her blog on the University of Warwick blog service:

In the middle of August I will be leaving Warwick (to be the new Corporate Communications Manager at the Open University). ... But, given that I will have to migrate my blog, where is the best place to go?

Blog migration is likely to be important not just for departing staff and students who make use of a blog service hosted at their institution, but also potentially users, such as myself, who use an externally hosted blog. What will I do if, for example, WordPress change the licence conditions for their hosted blog service?

John Dale suggested the **Vox** service - especially as it provides control over access to blog posts which appears to be a valuable aspect of the Warwick blog service. And as John is someone whose opinions I value I thought it would be worth trying out Vox.

65,720 hits

Visitor locations

ClustrMaps™ Click to see

Join the go-anywhere

How might you migrate the contents of a blog (e.g. you're leaving college)?

This question was raised by Casey Leaver, shortly before leaving Warwick University



# Case Study 2a: Blog Migration



She migrated her blog from blogs at Warwick Univ to Wordpress

Note, though, that not all data was transferred (e.g. title, but not contents) so there's a need to check transfer mechanisms



# Case Study 2b: The Individual's Blog (1)

The screenshot shows the Auricle blog homepage. At the top left, the word "AURICLE" is written in a large, stylized, hand-drawn font. To its right, the text "learning technologies in Higher Education" and "information, issues, perspectives" is visible. Below the header, there is a navigation menu with "Archives: January 2004" and "Fri Jan, 30 2004". The main content area features a post titled "The Virtual Resource Site for Teaching with Technology" by Lisa Williams, dated Jan 30, 04. The post text discusses enhancing teaching through technology and provides a link to the virtual resource site. On the right side, there are several utility boxes: "STATISTICS" showing total accesses (0001052930), total entries (305), and total comments (202); "ORIENTATION" with links for "About Auricle" and "Auricle - Home"; "CHANGE STYLE" with options like PDA, PDA2, Minimal, Plainwhite, Highcontrast, Bluesteel, and Whitesteel; "FIND" with a search input field, a "FIND" button, and a "Categories" dropdown menu; and "ARCHIVES" with links for "September 2008", "June 2006", and "April 2006".

Auricle blog:

- Launched Jan 2004 by head of e-learning team, Bath
- High profile & public visibility by early adopter & evangelist

Today:

- It's gone
- Lost after evangelist left, new staff arrive, new priorities, ...

# Case Study 2b: The Individual's Blog (2)



« Psst ... want to 'see' a good podcast?

(Return to home page)

## Auricle lives!

Over on the JISC-PoWR (Preservation of Web Resources) site there's a posting entitled [Auricle: The Case Of The Disappearing E-learning Blog](#) (1 September 2008). Yes, that's right **this** blog is the 'disappeared'. While it's a new experience becoming such a case study the author, Brian Kelly certainly raises some interesting issues. Although I've already contributed a comment on the PoWR site in today's post to Auricle I'm taking the opportunity to expand on the points I made there.

In his posting Brian describes how ghostly traces of *Auricle* continue to exist as a combination of past citations and dead hyperlinks to a now defunct University of Bath web site. Ironically, one such dead hyperlink in Brian's article was an extract from an old *Auricle* posting titled [The Weblog as the Model for a New Type of Virtual Learning Environment?](#) (*Auricle*, 25 February 2004). Readers of today's posting will find the hyperlink on this site is live. In my 2004 posting I posited that blogs could "form the basis of a distributed, not centralised, information and learning object system". It's always an interesting exercise to revisit what I've opined and cited in the past and then reflect on whether and how experiences since have impacted on what I opine and cite now.

In my 2004 posting I quote the following:

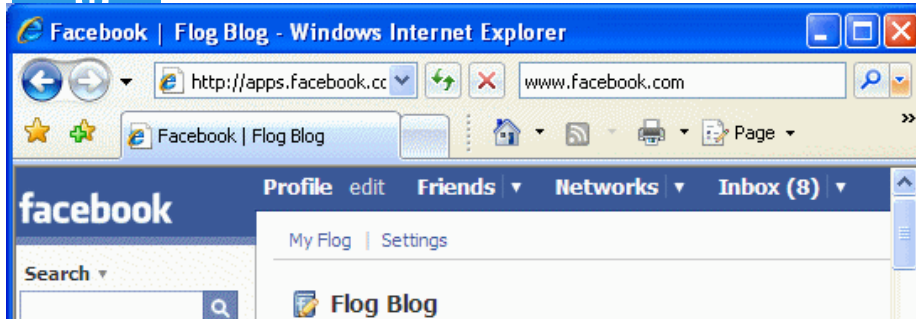
Auricle reborn:

- Further Google revealed the blog has been reborn
- New domain ([www.auricle.org/](http://www.auricle.org/))
- New engine (Wordpress) & look and feel (but old engine still available) →
- New content being added
- Old content still accessible

Preservation is helped by:

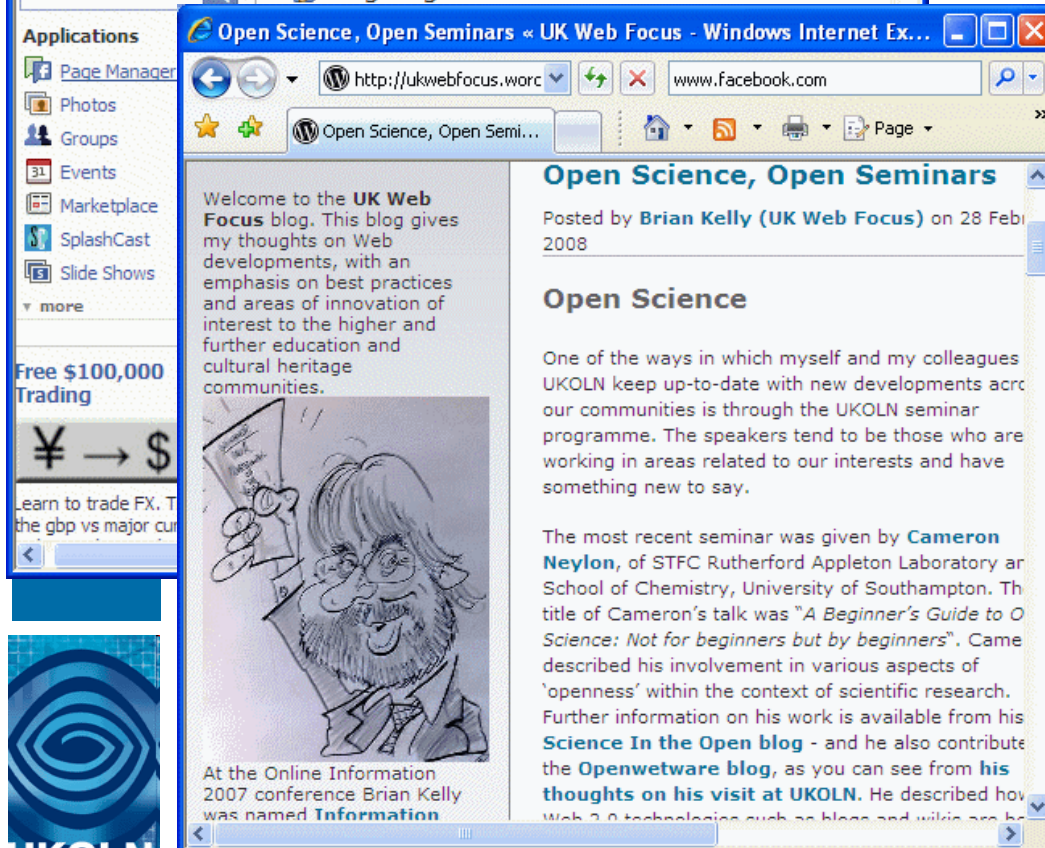
- Continued access
- Motivated & skilled owners

# Case Study 3: Reusing Data



Blog post in Facebook.  
Possible concerns:

- It's not sustainable
- You've given ownership to Facebook

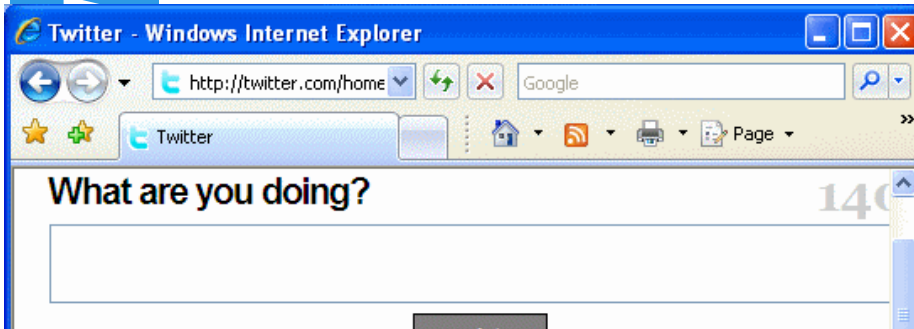


Response:

- The post is managed in WordPress; Fb displays copy (to new audience)
- Fb don't claim ownership – they claim rights to make money

It's not the service, it's how you use the service

# Case Study 4: Disposable Data



Twitter – example of a micro-blogging application →

Facebook status messages is another related example



Issues:

- Will Twitter be sustainable over a long period?
- What will happen to the data?
- What about the IPR for ‘tweets’?
- What about institutional uses? →



# Case Study 4: Disposable Data

@briankelly: Repeat after me: not everything needs to be preserved :-)

about 13 hours ago from im in reply to briankelly ☆



**PeteJ**

With Others

Previous

Cracking second goal by Murphy for Sunderland v Wigan 04:40 PM February 09, 2008 from im ☆

Coo. Vibert @ Trinity with SL linkup <http://www.xtremlab.net/dr...> 04:08 PM February 09, 2008 from im ☆

Delighted at the progress of the Keegan era 03:57 PM February 09, 2008 from im ☆

Lovely lunch in St Nick's market 03:18 PM February 09, 2008 from im ☆

Looking at Rum  
11:33 AM February

Testing XML Sc

BBC on MS, Go  
<http://news.bbc>



Many twitterers regard their tweets as disposal

I tend to use Twitter as a 'virtual water cooler' – sharing gossip, jokes and occasional work-related information with (mainly) people I know

You could make use of clients which manage your tweets (e.g. treat like email) But you should develop your policies first, prior to exploring technologies

# Case Study 4: Disposable Data



**Skype** (or your preferred VoIP application) are growing in popularity

Issues:

- Is the digital data (the call) preserved?
- What about the video and the IM chats?

Possible Responses:

- Am I bovered?
- I didn't bother with analogue phones, why should I worry now?

# Case Study 5: Slideshare

The screenshot shows a Slideshare presentation page. The title is "Web 1.0, Web 2.0 and Digital Preservation" by Brian Kelly, UKOLN, University of Bath, Bath, UK. The presentation is about a talk given at a workshop on behalf of MLA London on 14 July 2008. The presentation content includes an "About This Talk" section: "A recap of Web preservation challenges and approaches to the preservation of Web content. But will use of Web 2.0 services lead to new preservation concerns? How much of a concern is this? And what steps can be taken to minimise the risks of data loss?". The presentation is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 2.5 licence. The page also shows sharing options, a "More by user" section with a link to "An Introduction To RSS Readers: Google Reader and Netibes", and a "Privacy Info" section.

What happens to your slides if Slideshare disappears?  
Recommended approach:

- Master copy held on managed environment
- Info on master on title slide and metadata
- CC licence & download available – many copies

# Web Preservation Challenges

What do we wish to preserve:

- The underlying *resource*
- The user *experience*
- Ease of *access* to the resource

Examples of each addressed in project:

- Migration of blog posts (and comments?)
- Capturing images of home pages using Internet Archive
- Including address of master resource on Slideshare

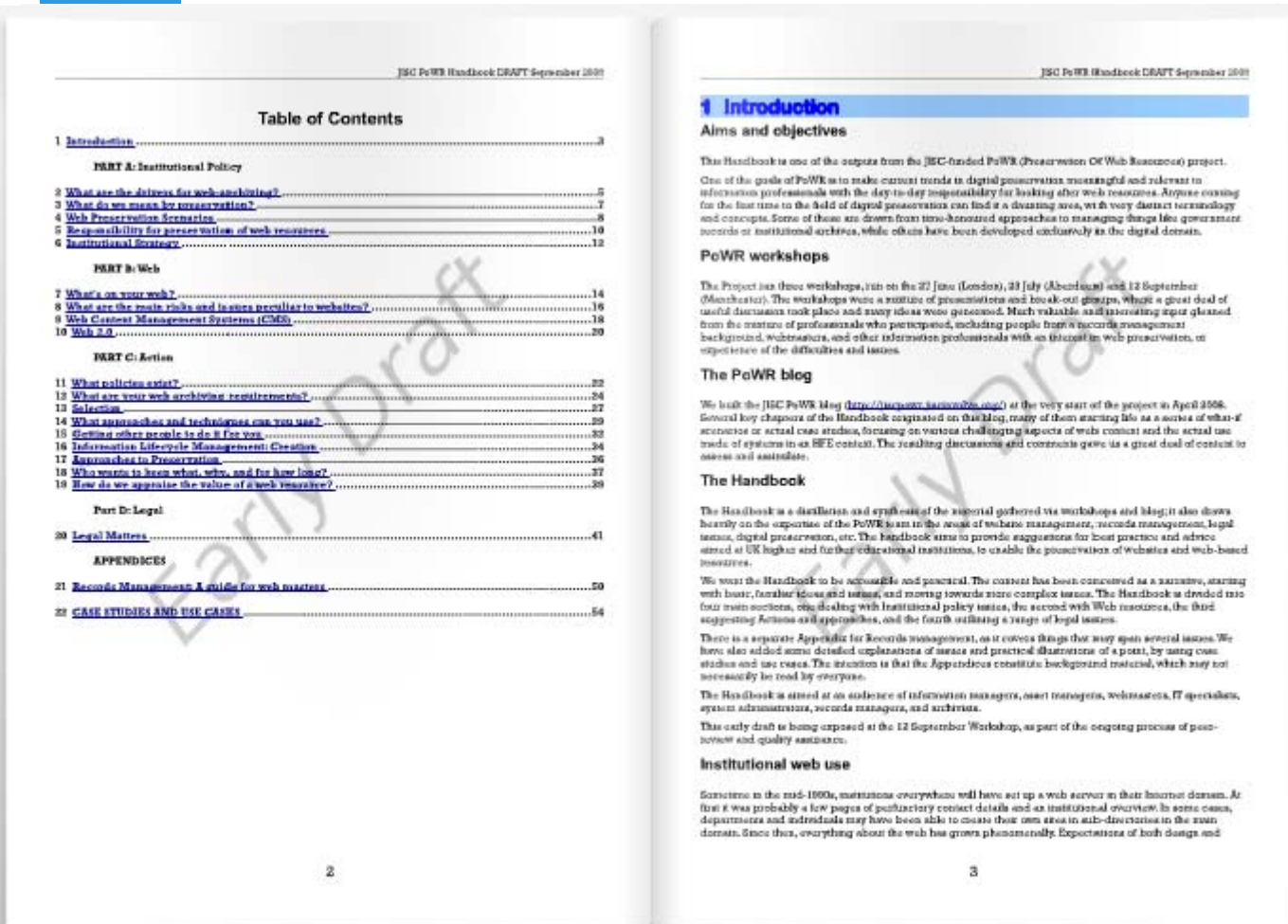


# The Handbook

Handbook:

- Currently being finalised
- Creative Commons licence

Your feedback will be appreciated



# Rethinking Web Preservation

Chris Rusbridge:

I would argue that outcome-related phrases like "long term accessibility" or "usability over time" are better than the process-oriented phrase "digital preservation" →

How does this relate to JISC PoWR work?

Consider *institutional*:

- Lack of interest in "digital preservation"
- Importance of *use* of services
- Importance of *reuse* of services

This needs to complement:

- National approaches to Web preservation and Web harvesting





University Archivist, Records  
Manager and FOI Co-ordinator  
Lizzie Richmond

Head of Web Services  
Alison Wildish



## What have we learned?

- Better informed about differences between printed and web records and their implications

- Recognition that web preservation should be addressed to avoid gap in University history

- This is worth doing
- There's a lot to think about
- We'll need to work together to succeed
- We need a strategy because:
  - its important at an institutional level
  - consistency of approach will be crucial
  - the line between publication and record is blurred

# Conclusions

JISC PoWR project has:

- Helped to begin process of raising awareness on Web preservation within institutions
- Facilitated engagement with key stakeholders in a small number of institutions
- Produced examples of pragmatic approaches to preservation of Web resources
- Received feedback on the approaches
- Produced draft handbook to share these approaches more widely

The challenges of Web site preservation are only just beginning

