

# The Digital Preservation Coalition Digital Preservation Awards 2005

Dave Thompson, John Tuck

The UK Web Archiving Consortium



# What is the problem?

New information is now being mainly produced in electronic format with the internet becoming the publishing medium of choice. The British Library suggests that 40% of UK research data will be available in electronic format only by 2020

Yet digital material has proved to be transitory, fragile and ephemeral and websites cited in the literature can be lost in just a few years or even months

Much material has already been lost. Devolved responsibility for dissemination means that many producers and creators don't have the resources & skills to keep material indefinitely and maybe don't even see future value in their material



# Isn't this just hype?

Of the website owners who have so far given Wellcome Library permission for their site to be archived as part of UKWAC

69% said they didn't archive their sites

24% said they did archive their sites

7% didn't answer the question



# UK Web Archiving Consortium

The issue of saving web-based material is too big, expensive and complex to be tackled by a single institution

UKWAC is a group of six institutions planning to archive and make permanently accessible selected '.uk' domain web sites, initially within a two-year pilot project

UKWAC believes that costs, risk and experience can be shared, thus minimising the burden on any single organisation whilst maximising the chances of preserving sites permanently



UK WEB ARCHIVING  
CONSORTIUM  
[www.webarchive.org.uk](http://www.webarchive.org.uk)

# Benefits of the project

The first successful selective archive of UK web space which imposes no charges for including material or for access. This is based on the National Library of Australia web archiving application 'PANDORA'

UKWAC works with the *explicit* permission of rights holders to archive selected UK websites. Such collaborative web archiving supports the viability and accessibility of web-based material into the future

Expanding the lifespan of web-based materials ensuring that scholarly, cultural and scientific resources remain available for future generations through the creation of an authoritative 'trusted' archive of web-based material



# The UKWAC archive

Currently 552 different websites archived since September 2004

A total of over 2,000 instances of these sites captured

Events-based 'collections' which centre around the 2004 Asian Tsunami, the 2005 UK General Election, the London bombings of 2005, and the G8 summit of 2005

Websites and their contents freely accessible and located either by subject-based browsing or via a search feature



UK WEB ARCHIVING  
CONSORTIUM  
[www.webarchive.org.uk](http://www.webarchive.org.uk)

# Innovative technology & practice

Centralised web-delivered core applications for both repository and public interface – the PANDAS module

Centralised data storage, outsourced to a commercial provider who provides a fully maintained, secure & managed environment

Federated collection management, collection policies and approaches to rights issues - but collective decision making

Shared cost = shared risk

The future of UKWAC is not dependent upon the PANDAS application



UK WEB ARCHIVING  
CONSORTIUM  
[www.webarchive.org.uk](http://www.webarchive.org.uk)

# Six practical benefits

1. Web-based material from the '.uk' domain is captured at a time when it is readily available and is at its most functional
2. An easy-to-use interface by which users can access information about the project and the contents of the archive itself
3. Each partner selects and archives material within its own field of expertise with clearly defined areas of activity, building a single coherent whole
4. Material can be cited in a familiar way, by URL, and accessed later by means of persistent identifiers
5. The archive clearly differentiates between 'archived' and 'live' versions of web sites whilst offering access to either
6. Access to the site and to the archive is free and open to all





# Who benefits from the archive?

1. The UKWAC archive benefits academics and subject specialists as well as 'lay users' and the simply curious through its growing collection of selected web sites which have been measured against collection criteria for their cultural and social importance as well as their intellectual content
2. Content creators and rights owners can be sure that their material is being preserved indefinitely in a trusted archive which respects their intellectual property and which strives to resolve issues through dialogue and co-operation
3. The archive supports professionals working with digital preservation by providing an example and a model for the UK that helps increase popular awareness of digital preservation to all members of society



# How the UKWAC project differs

The project is built on broad international co-operation and collaboration in web archiving tools and standards yet responsibility for the acquisition and management of UK material remains within the UK

UKWAC archives selected material from the UK web space more frequently, more deeply and with more reliability and accuracy than projects that archive the whole of domains or are done by external agents

It is built on as firm a basis of legal compliance and best practice as is possible

We build on shared expertise/learning and UKWAC contributes to a global network of archives built upon common experiences

The Consortium approach offers the best possible chance for the long-term survival of the websites it has decided to capture and maintain



UK WEB ARCHIVING  
CONSORTIUM  
[www.webarchive.org.uk](http://www.webarchive.org.uk)

# UKWAC: the future

Partners are currently evaluating the project and will soon be making recommendations for the future

It is our expectation that we will build on strong foundations

We will be looking ahead to improved software and system developments

We will be looking ahead towards secondary legislation under the Legal Deposit Libraries Act 2003

The future is bright



UK WEB ARCHIVING  
CONSORTIUM  
[www.webarchive.org.uk](http://www.webarchive.org.uk)

# In conclusion

Our project is different. It has delivered on its objectives and represents a collaborative, experience-based and responsive approach to web archiving

It benefits the archival community as well as end users & content creators

It has succeeded through co-operation and a shared process of learning based on technological innovation

It has achieved the solid foundations of a system able to rise to these challenges and the aggregation of people with the necessary skills and enthusiasm to embrace them



# Thank You

Dave Thompson, John Tuck  
The UK Web Archiving Consortium

The Digital Preservation Coalition Digital Preservation Awards 2005

