The National Archives

Preserving the content and the network: An innovative approach to web archiving

Amanda Spencer, Web Continuity Manager, The National Archives, Sept 2008

Introduction

- Since the early 1990s the Government has been using websites to present information: official reports, papers, transcripts of speeches, guidance, announcements, press statements, regulations and advice.
- Benefits offered by new technologies means that services are increasingly delivered online.
- Majority of all interaction between government and the public now happens online.
- Web links cited in everything from Hansard through to academic research and PR campaigns
- Government web estate, just like the rest of the web, vulnerable to technological problems
- Broken Web links are increasingly common – and frustrating for us all

Web continuity matters

- Integrity of Web links crucial to the business of government. Without it:
- Public – impaired access
- Westminster village – impaired access
- The Press – impaired access
- Academics – impaired analysis
- Parliament – impaired scrutiny
- Impacts on reputation of government – public and abroad

Government is taking action

- April 2007 - Raised as a serious issue by Jack Straw, as Leader of the House of Commons
- May 2007 – Hilary Armstrong commits government to a package of measures that will provide a long-lasting solution to the problem
- Working Group established in May 2007 to investigate the problem and come up with a solution

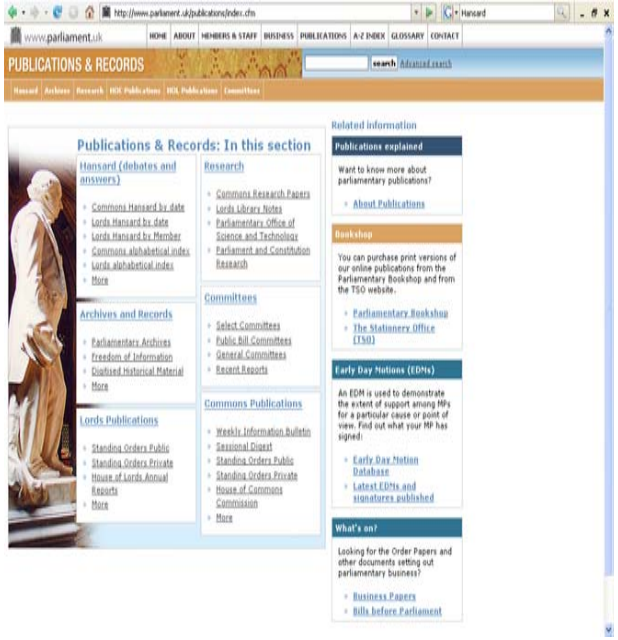
What the working group found…

- A longitudinal survey of URLs cited in response to Parliamentary Questions and recorded in Hansard revealed that 60% of URLs, cited in Hansard between 1997-2006 are broken
- Disconnect between those who view government websites as ephemeral and those who expected important information to be available in perpetuity
- Prevalence of e-only publication; issues with legal deposit in the electronic world
- Varied practice - PDF vs HTML
- Devolved system of web publication
- Concerns also about the impact of ‘Website Rationalisation’

Agreement

- Working group agreed that all web-based information should be treated as an important contribution to the body of government information

- Online information which has been cited elsewhere should remain available and accessible in its original form



Publications & Records: In this section

Hansard (debates and answers)

- Commons Hansard by date
- Lords Hansard by date
- Lords Hansard by Member
- Commons alphabetical index
- Lords alphabetical index
- More

Archives and Records

- Parliamentary Archives
- Freedom of Information
- Digitised Historical Material
- More

Lords Publications

- Standing Orders Public
- Standing Orders Private
- House of Lords Annual Reports
- More

Research

- Commons Research Papers
- Lords Library Notes
- Parliamentary Office of Science and Technology
- Parliament and Constitution Research

Committees

- Select Committees
- Public Bill Committees
- General Committees
- Recent Reports

Commons Publications

- Weekly Information Bulletin
- Sessional Digest
- Standing Orders Public
- Standing Orders Private
- House of Commons Commission
- More

Related information

Publications explained

Want to know more about parliamentary publications?

- About Publications

Bookshop

You can purchase print versions of our online publications from the Parliamentary Bookshop and from the TSO website.

- Parliamentary Bookshop
- The Stationery Office (TSO)

Early Day Motions (EDMs)

An EDM is used to demonstrate the extent of support among MPs for a particular cause or point of view. Find out what your MP has signed:

- Early Day Motion Database
- Latest EDMs and signatures published

What's on?

Looking for the Order Papers and other documents setting out parliamentary business?

- Business Papers
- Bills before Parliament

Options

- Improvements to existing practices
- Encourage government departments to take more responsibility
- Use Digital Object Identifiers
- All of the above?

The National Archives and the Web Continuity solution – issues to address

1. How to capture significant levels of important Government information from possibly thousands of distributed, heterogeneous websites (including websites closing as part of the Website Rationalisation Programme);
2. How to ensure not only a greater capture of content, but also increase exposure of this content to the web harvesting crawler, from sites that vary hugely in nature;
3. How to ensure that links persist so that users will always find the last available version of the page, whether it is on a live site, or in the web archive.

Web Continuity – capture I

- Comprehensive - Whole of UK central government
- Incorporating Transformational Government Website Rationalisation Programme
- SQL Server Government Website Database developed as a single source of information for central government websites

  - allows tight control over web estate

  - gives responsibility to government

Directgov

Business Link

Web Continuity – capture II

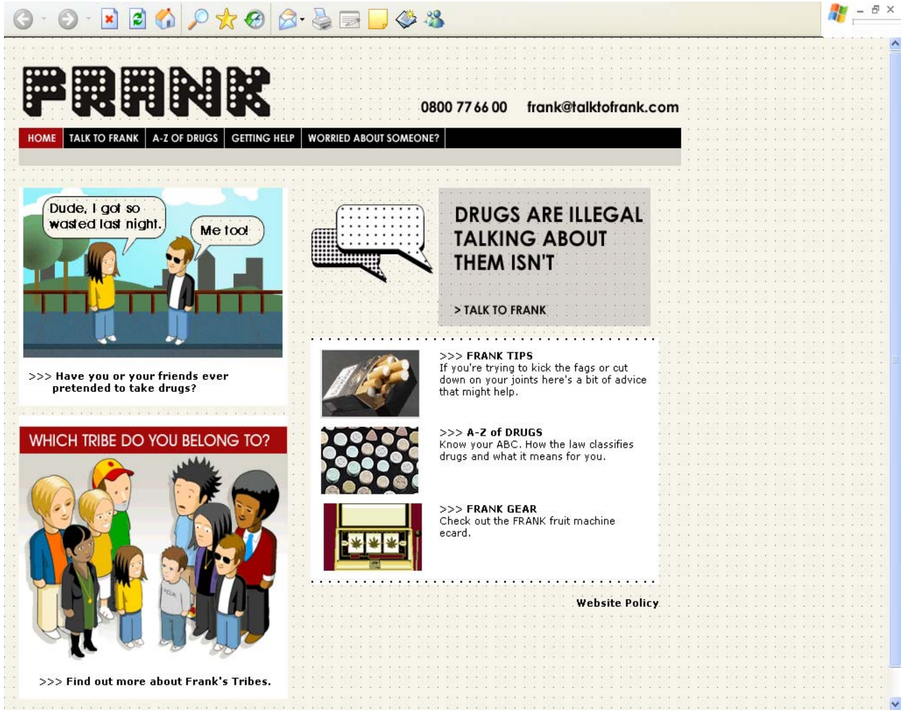
- Automated seeding of crawls and capture of preservation copies via two workflows:
- Harvesting workflow, crawls are seeded, and progress of harvesting and QA processes are recorded in the database through the exchange of a series of XML messages via FTP between TNA and the European Archive, who carry out web archiving under contract
- Preservation workflow, enables ARC files to be ingested into The National Archives Digital Object Store, once harvesting process is complete

Web Continuity – capture II

- Promoting use of XML sitemaps
- Benefits live website users as well as helping the archiving process; ministerial backing
- Training in partnership with Third Party provider
- Guidance provided
- Installation of sitemap generation software on government web servers – requires support from senior government
- Practical implementation – will be used to augment the initial gather rather than to drive the crawl

Web Continuity – links persistence

- Configuration of open source software
- Apache and Microsoft IIS Web Servers
- Enables users to be redirected to web archive instead of receiving a 404 Page Not Found error; does not replace existing redirects on live websites
- Will need to be installed on government web servers



>>> Have you or your friends ever pretended to take drugs?

>>> Find out more about Frank's Tribes.

DRUGS ARE ILLEGAL TALKING ABOUT THEM ISN'T

>>> FRANK TIPS
If you're trying to kick the fags or cut down on your joints here's a bit of advice that might help.

>>> A-Z of DRUGS
Know your ABC. How the law classifies drugs and what it means for you.

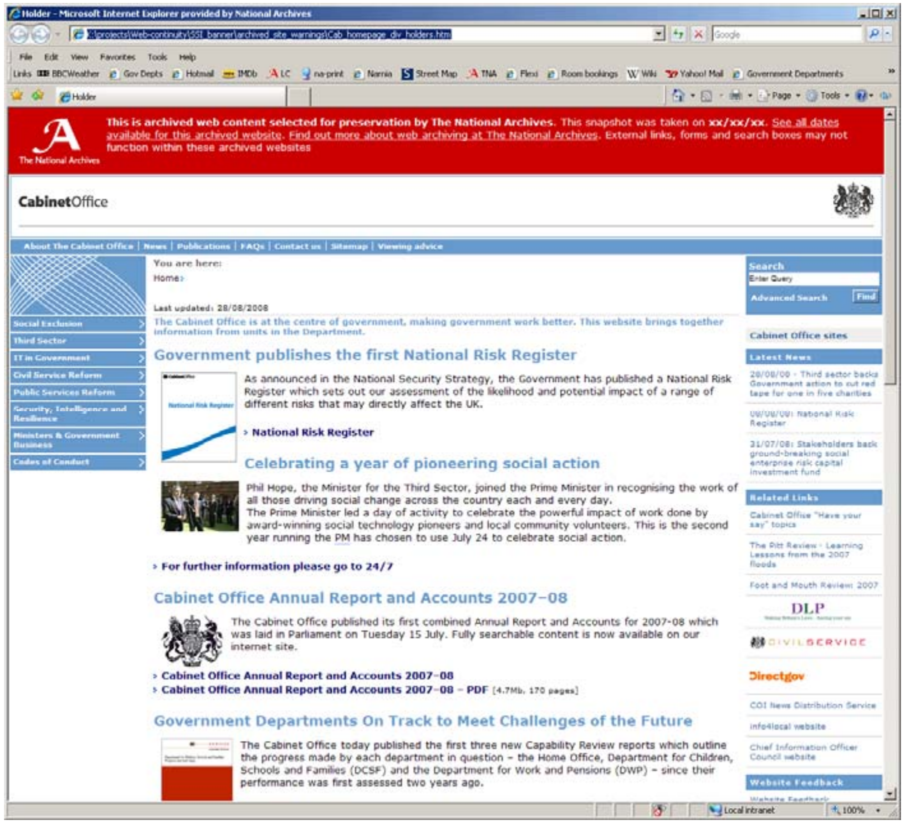
>>> FRANK GEAR
Check out the FRANK fruit machine ecard.

Web Continuity – links persistence - implications

- Brings together different audiences
- Archived websites have: historical value

  current value

  value in preserving integrity of network as a whole
- Introduces temporal dimension – brings longevity to the web, but also needs to be clearly signposted – new presentation and branding, including a TNA URL



Holder - Microsoft Internet Explorer provided by National Archives

X:\projects\Web-continuity\SSI_banner\archived_site_warnings\Cab_homepage_div_holders.htm

The National Archives

This is archived web content selected for preservation by The National Archives. This snapshot was taken on **xx/xx/xx**. See all dates available for this archived website. Find out more about web archiving at The National Archives. External links, forms and search boxes may not function within these archived websites

CabinetOffice

About The Cabinet Office | News | Publications | FAQs | Contact us | Sitemap | Viewing advice

Social Exclusion
Third Sector
IT in Government
Civil Service Reform
Public Services Reform
Security, Intelligence and Resilience
Ministers & Government Business
Codes of Conduct

You are here:
Home›

Last updated: 28/08/2008

The Cabinet Office is at the centre of government, making government work better. This website brings together information from units in the Department.

Government publishes the first National Risk Register

As announced in the National Security Strategy, the Government has published a National Risk Register which sets out our assessment of the likelihood and potential impact of a range of different risks that may directly affect the UK.

› National Risk Register

Celebrating a year of pioneering social action

Phil Hope, the Minister for the Third Sector, joined the Prime Minister in recognising the work of all those driving social change across the country each and every day.
The Prime Minister led a day of activity to celebrate the powerful impact of work done by award-winning social technology pioneers and local community volunteers. This is the second year running the PM has chosen to use July 24 to celebrate social action.

› For further information please go to 24/7

Cabinet Office Annual Report and Accounts 2007-08

The Cabinet Office published its first combined Annual Report and Accounts for 2007-08 which was laid in Parliament on Tuesday 15 July. Fully searchable content is now available on our internet site.

› Cabinet Office Annual Report and Accounts 2007-08
› Cabinet Office Annual Report and Accounts 2007-08 - PDF [4.7Mb, 170 pages]

Government Departments On Track to Meet Challenges of the Future

The Cabinet Office today published the first three new Capability Review reports which outline the progress made by each department in question - the Home Office, Department for Children, Schools and Families (DCSF) and the Department for Work and Pensions (DWP) - since their performance was first assessed two years ago.

Search
Enter Query
Advanced Search Find

Cabinet Office sites

Latest News
28/08/08 - Third sector backs Government action to cut red tape for one in five charities
08/08/08: National Risk Register
31/07/08: Stakeholders back ground-breaking social enterprise risk capital investment fund

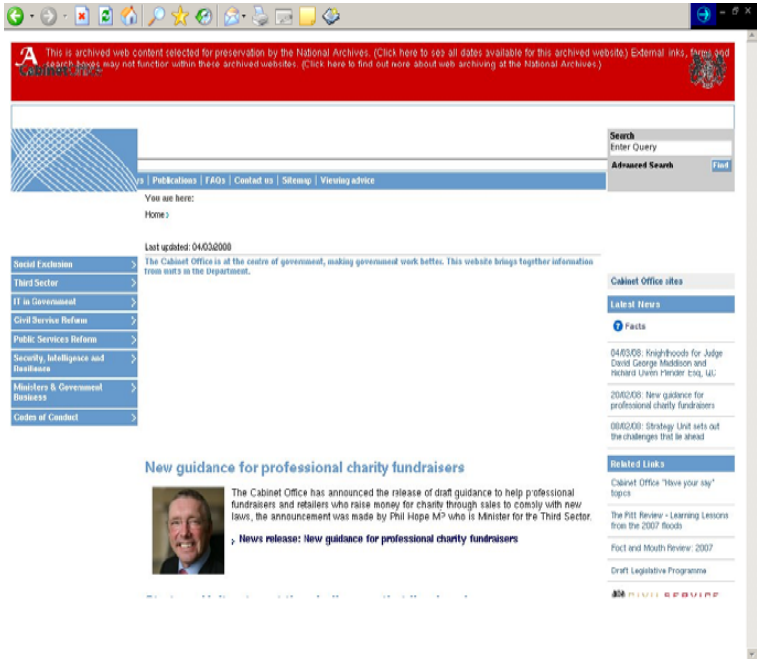
Related Links
Cabinet Office "Have your say" topics
The Pitt Review - Learning Lessons from the 2007 floods
Foot and Mouth Review: 2007
DLP
CIVIL SERVICE
Directgov
COI News Distribution Service
info4local website
Chief Information Officer Council website

Website Feedback
Website Feedback

Web Continuity – Presentation issues

- Several attempts at producing a banner:

Using Iframes:



This is archived web content selected for preservation by the National Archives. (Click here to see all dates available for this archived website.) External links, forms and search boxes may not function within these archived websites. (Click here to find out more about web archiving at the National Archives.)

You are here:

Home

Last updated: 04/03/2008

The Cabinet Office is at the centre of government, making government work better. This website brings together information from units in the Department.

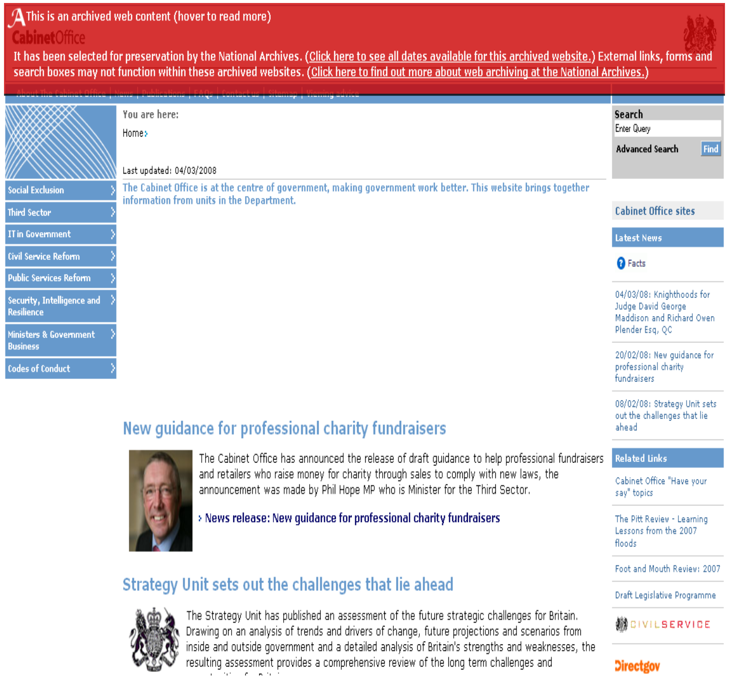
New guidance for professional charity fundraisers

The Cabinet Office has announced the release of draft guidance to help professional fundraisers and retailers who raise money for charity through sales to comply with new laws, the announcement was made by Phil Hope MP who is Minister for the Third Sector.

› News release: New guidance for professional charity fundraisers

CabinetOffice
s | Publications | FAQs | Contact us | Sitemap | Viewing advice
Search
Enter Query
Advanced Search
Find
Social Exclusion
Third Sector
IT in Government
Civil Service Reform
Public Services Reform
Security, Intelligence and Resilience
Ministers & Government Business
Codes of Conduct
Cabinet Office sites
Latest News
Facts
04/03/08: Knighthoods for Judge David George Maddison and Richard Owen Plender Esq, QC
20/02/08: New guidance for professional charity fundraisers
08/02/08: Strategy Unit sets out the challenges that lie ahead
Related Links
Cabinet Office "Have your say" topics
The Pitt Review - Learning Lessons from the 2007 floods
Foot and Mouth Review: 2007
Draft Legislative Programme
CIVIL SERVICE

Using CSS overlay:



This is an archived web content (hover to read more)

It has been selected for preservation by the National Archives. (Click here to see all dates available for this archived website.) External links, forms and search boxes may not function within these archived websites. (Click here to find out more about web archiving at the National Archives.)

You are here:

Home

Last updated: 04/03/2008

The Cabinet Office is at the centre of government, making government work better. This website brings together information from units in the Department.

New guidance for professional charity fundraisers

The Cabinet Office has announced the release of draft guidance to help professional fundraisers and retailers who raise money for charity through sales to comply with new laws, the announcement was made by Phil Hope MP who is Minister for the Third Sector.

› News release: New guidance for professional charity fundraisers

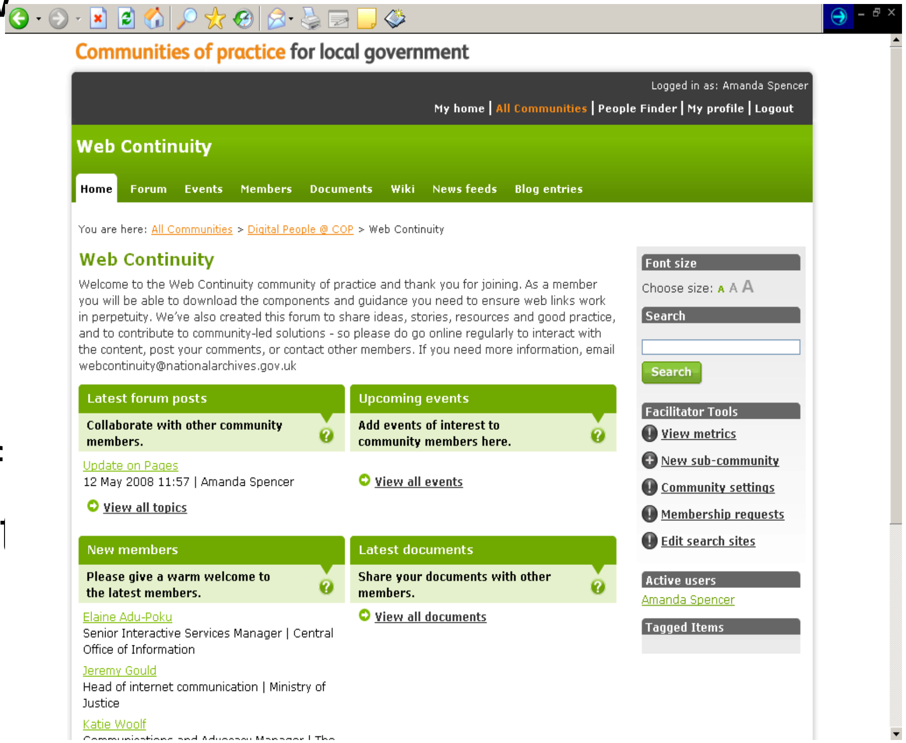
Strategy Unit sets out the challenges that lie ahead

The Strategy Unit has published an assessment of the future strategic challenges for Britain. Drawing on an analysis of trends and drivers of change, future projections and scenarios from inside and outside government and a detailed analysis of Britain's strengths and weaknesses, the resulting assessment provides a comprehensive review of the long term challenges and

CabinetOffice
About The Cabinet Office | News | Publications | FAQs | Contact us | Sitemap | Viewing advice
Search
Enter Query
Advanced Search
Find
Social Exclusion
Third Sector
IT in Government
Civil Service Reform
Public Services Reform
Security, Intelligence and Resilience
Ministers & Government Business
Codes of Conduct
Cabinet Office sites
Latest News
Facts
04/03/08: Knighthoods for Judge David George Maddison and Richard Owen Plender Esq, QC
20/02/08: New guidance for professional charity fundraisers
08/02/08: Strategy Unit sets out the challenges that lie ahead
Related Links
Cabinet Office "Have your say" topics
The Pitt Review - Learning Lessons from the 2007 floods
Foot and Mouth Review: 2007
Draft Legislative Programme
CIVILSERVICE
Directgov

Practical Implementation

- Implementation across government has required new approach to sharing information and different technical support model
- Collaborative platform established
- Guidance for web and e-comms communities – part of Transformational Government Web standards
- Regular monitoring
- Different groups of stakeholders involved at The National Archives



You are here: All Communities > Digital People @ COP > Web Continuity

Web Continuity

Welcome to the Web Continuity community of practice and thank you for joining. As a member you will be able to download the components and guidance you need to ensure web links work in perpetuity. We've also created this forum to share ideas, stories, resources and good practice, and to contribute to community-led solutions - so please do go online regularly to interact with the content, post your comments, or contact other members. If you need more information, email webcontinuity@nationalarchives.gov.uk

Latest forum posts

Collaborate with other community members.

Update on Pages
12 May 2008 11:57 | Amanda Spencer

View all topics

Upcoming events

Add events of interest to community members here.

View all events

New members

Please give a warm welcome to the latest members.

Elaine Adu-Poku
Senior Interactive Services Manager | Central Office of Information

Jeremy Gould
Head of internet communication | Ministry of Justice

Katie Woolf
Communications and Advocacy Manager | The

Latest documents

Share your documents with other members.

View all documents

Timelines

- November 2007 – April 2008 Feasibility studies
- May – November 2008 Development work at The National Archives, website archiving programme scaled up
- November 2008 Solution delivered to government, software and guidance available, monitoring in place

Any Questions?

The National Archives

nationalarchives.gov.uk