Continuity and Capture: Broken links and the UK Central Government Web presence
Amanda Spencer and Tom Storror, The National Archives, July 2009
Originally we had a simple plan

• UK Government Web Archive launched 2003
• Contracted out
  - Internet Archive 2003-2005
  - European Archive 2005 onwards
• Based around 7 core functions – sample of c50 major government websites monthly or six-monthly
• Also ad hoc collections relating to current events
Website Rationalisation

- Part of Transformational Government strategy
- Aims to better manage the government web estate through convergence of much of information on 3 super sites:
  - Stricter control over creation of new .gov.uk sites
  - The National Archives to archive all affected sites
  - As many as 2000 sites
  - Programme completed by 2011
...But the world is changing fast

- Massive growth of government web estate
- Web is used by Government for all sorts of purposes – transactions and interaction with citizens
- Even to answer PQs!

and this creates problems....

- Government’s use of the web varies
Government’s use of the web

- Ephemeral vs. Official record
- Prevalence of e-only publications
- Good and bad website management
- Problems with Hansard and the Jack Straw letter
Why it matters

• Integrity of Web links crucial to the business of government. Without it:
  • Public – impaired access – crucial given that the majority of all interaction between government and the public now happens online.
  • 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
The working group – research and recommendations

• 60% of URLs quoted in Hansard between 1997-2006 are broken
• Prevalence of e-only publication; issues with legal deposit in the electronic world
• Varied practice - PDF vs HTML
• Devolved system of web publication

• Project of comprehensive archiving and link management to be taken forward by The National Archives
The Web Continuity project

Part of a package of projects that are addressing the issue of sustainable information across government
Comprehensively archiving government’s websites

- In partnership with the European Archive
- Whole of UK central government
- Incorporating Transformational Government Website Rationalisation Programme
- 3 x yearly; around 1200 sites
- Website managers can request additional crawls
Making sure web links work in perpetuity

- 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
Web Continuity Community of Practice
Greater capture of content

• Greater stakeholder expectation

• Promoting use of XML sitemaps through guidance and training in partnership with Third Party provider

• Working directly with content creators to encourage best practice in website development

• Working with the European Archive to improve technical capture; benefitting from their experience in Research and Development
Single, centrally hosted government website database
Improved presentation in Web Archive
Improved presentation in Web Archive: Search
Improved presentation in Web Archive
Challenges and Influencing

Challenges:
- Identifying people to influence
- Reaching the target audience
- Complex IT arrangements in Government departments

Lessons learned:
- Crucial to influence policy makers e.g. COI and Web Standards
- Crucial to influence decision makers e.g. Heads of e-Comms across government
- Helpful to have levers – for example the tool developed to monitor and report on broken links on UK Central Government websites
Project Milestones 2008

• **Spring 2008** Brochure and briefings
• **September 2008** technical briefings to government; Sitemaps guidance and training
• **October 2008** Solution piloted with MoJ, software and technical guidance available
• **November 2008** Stripe implemented; Comprehensive archiving implemented
• **December 2008** Web Archive pages re-launched
Project Milestones 2009

- **January 2009** Archiving and Managing URLs guidance available
- **June 2009** Monitoring begins; Database available
- **September 2009** Web Archive Search available
- **October 2009** Official launch of Web Continuity