

Web Archiving at TNA

David Ryan

Head of Archive Services

Selective Approach

- Only suitable in certain circumstances
- Resource intensive
- Requires detailed documentation
- Provides most authentic version of website

Harvesting

- Fast and effective for large numbers of sites
- Easy to automate
- Requires detailed testing
- Limitations of what can be collected

TNA Strategy

- Evaluating a number of technical approaches
- Developing a selection policy for websites
- Working with Government websites to develop appropriate guidance
- Developing long-term preservation strategies

Why?

"The *Modernising Government* White Paper sets a target that all Government services to the citizen and to business should be available online by 2005. The World Wide Web is increasingly becoming the principle means of interaction between Government, citizens and business, and the National Archives has a responsibility to collect and preserve websites as evidence of the changing nature of this interaction."



Issues

- Size of Government web domain (circa 2500 sites)
- Increasingly dynamic content
- Copyright
- Legal deposit

Selection Policy

Fifty websites were selected by the National Archives, in accordance with criteria designed to reflect the overall functions of government. A number of departments and agencies were then chosen which are representative of each of these functions. This provides a broad cross-section across UK Central Government.

Launch of the UK Central Government Web Archive



<http://www.pro.gov.uk/webarchive>

Method

- Sites collected weekly and six monthly
- Using modified Internet archive crawler



Delivery

- Archived sites available from TNA websites
- Also available via wayback machine
- Copy of all data supplied to TNA for long term preservation
- Investigation acquisition of Internet archive back catalogue

Issues relating to Internet Archiving

- Need to ensure long term security of existing collections
- Internet Archive approach emphasises breadth over depth
- Detailed Quality Assurance procedures are essential