

 **The National Archives**



The National Archives

**Enhancing Access to Web
Archives: Web Continuity at The
National Archives**

Amanda Spencer and Tom Storrar

7 October 2011

Enhancing Access to Web Archives

- UK Government Web Archive
- Web Continuity
- Trends in traffic statistics
- User survey work
- Improving findability through the Semantic Knowledge base project

The UK Government Web Archive

The collage displays several archived government websites:

- DFID (Department for International Development):** Includes an 'EVALUATION REPORT EV687' from April 2008 and an 'EVALUATION OF DFID COUNTRY PROGRAMMES' page.
- Number10.gov.uk:** The official site of the Prime Minister's Office, featuring a 'Top Story' section with a video of Gordon Brown's speech on constitutional reform.
- BERR (Department for Business, Enterprise & Regulatory Reform):** Shows a 'Meeting the Energy Challenge A White Paper on Nuclear Power January 2008' document.
- CABE (Commission for Architecture and the Built Environment):** Features an article titled 'Expert advice to help create better places'.
- The National Archives:** Multiple instances showing search results and metadata for various archived pages, including 'Rediscovering Frank Caza' and 'Stop, Look and Listen'.

	2007	2006
Notes	£000	£000
2	903	1,009
3	137	26
	<u>1,040</u>	<u>1,035</u>
4	(450)	(449)
	(6)	(6)
Other operating costs	(588)	(575)
	<u>(1,044)</u>	<u>(1,032)</u>
OPERATING (DEFICIT) / SURPLUS ON OPERATING ACTIVITIES	(4)	3
Interest receivable	6	1

Why?

- Change in the way government publishes information
- Danger information could be lost
- Valuable historical and cultural record

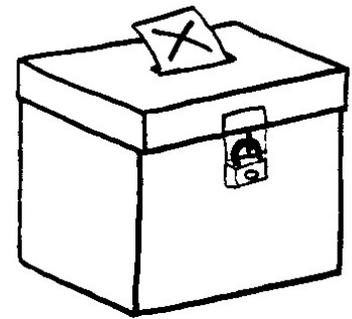
Web Continuity

Provides a redirection solution, using the UK Government Web Archive, to make sure information remains accessible.

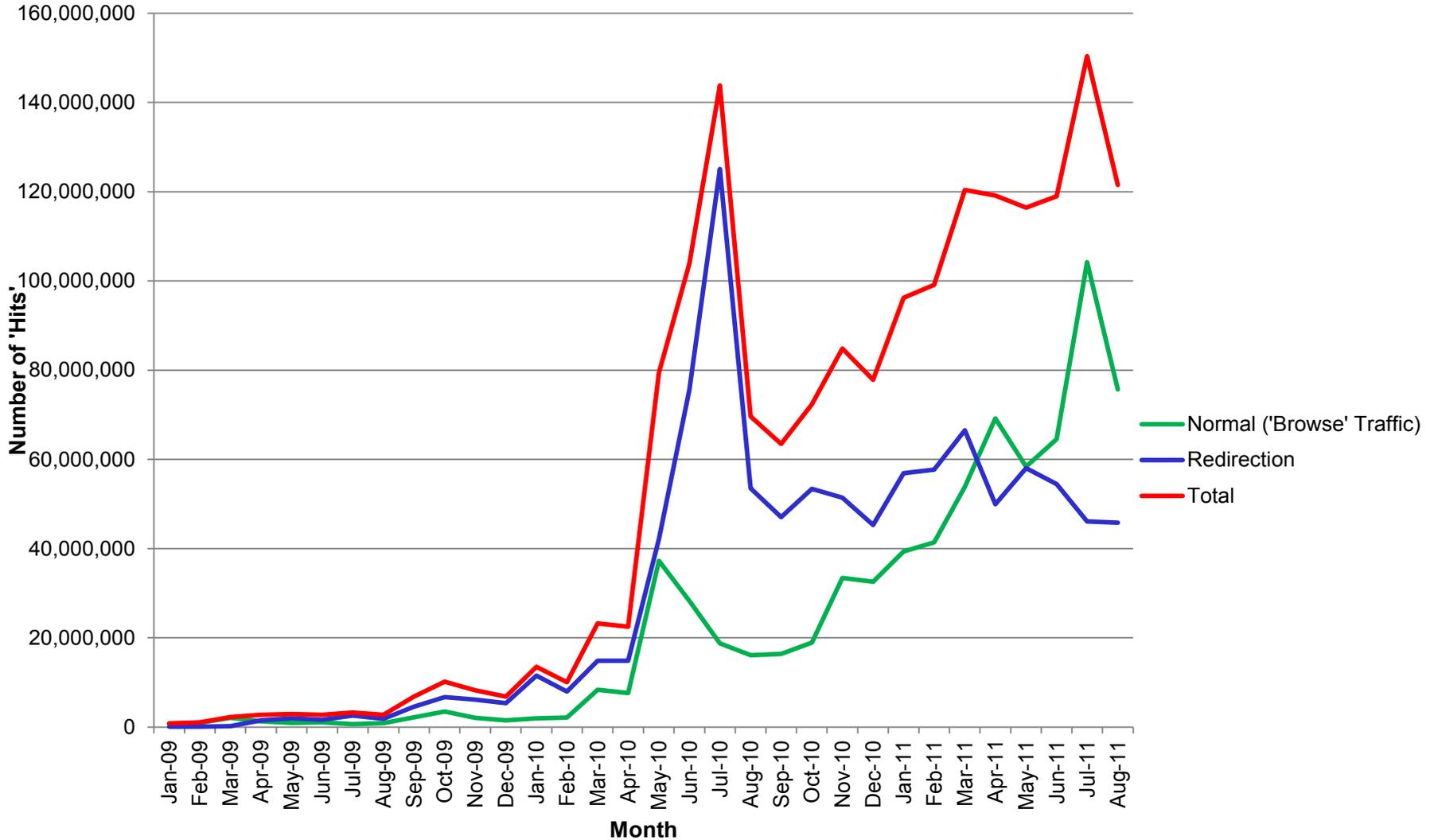


Machinery of government changes

Websites are at risk during Machinery of Government changes, such as might happen with a general election.



Usage



Traffic trends

- Over the past 6 months, more users are entering the web archive through ‘conventional’ browsing routes.
- Since we redesigned our web archive pages, we publish our popular websites.
- Analysis has shown that *Environment, Home affairs, public order, justice and rights, and Health, well-being and care* are consistently heavily used.

Other Current Web Archive Projects

Secure Web Archiving Project

Research into social media capture

Datasets work to support Transparency

Semantic Knowledge Base

Who are our users and what do they want?

- We conducted a user survey in May 2010.
- Showed that the web archive was used by a real variety of people.
- The biggest issue around the customer experience of using the web archive was found to be the existing search functions.

The problem with conventional search

- Scale
- Structure
- Domain knowledge
- Duplication

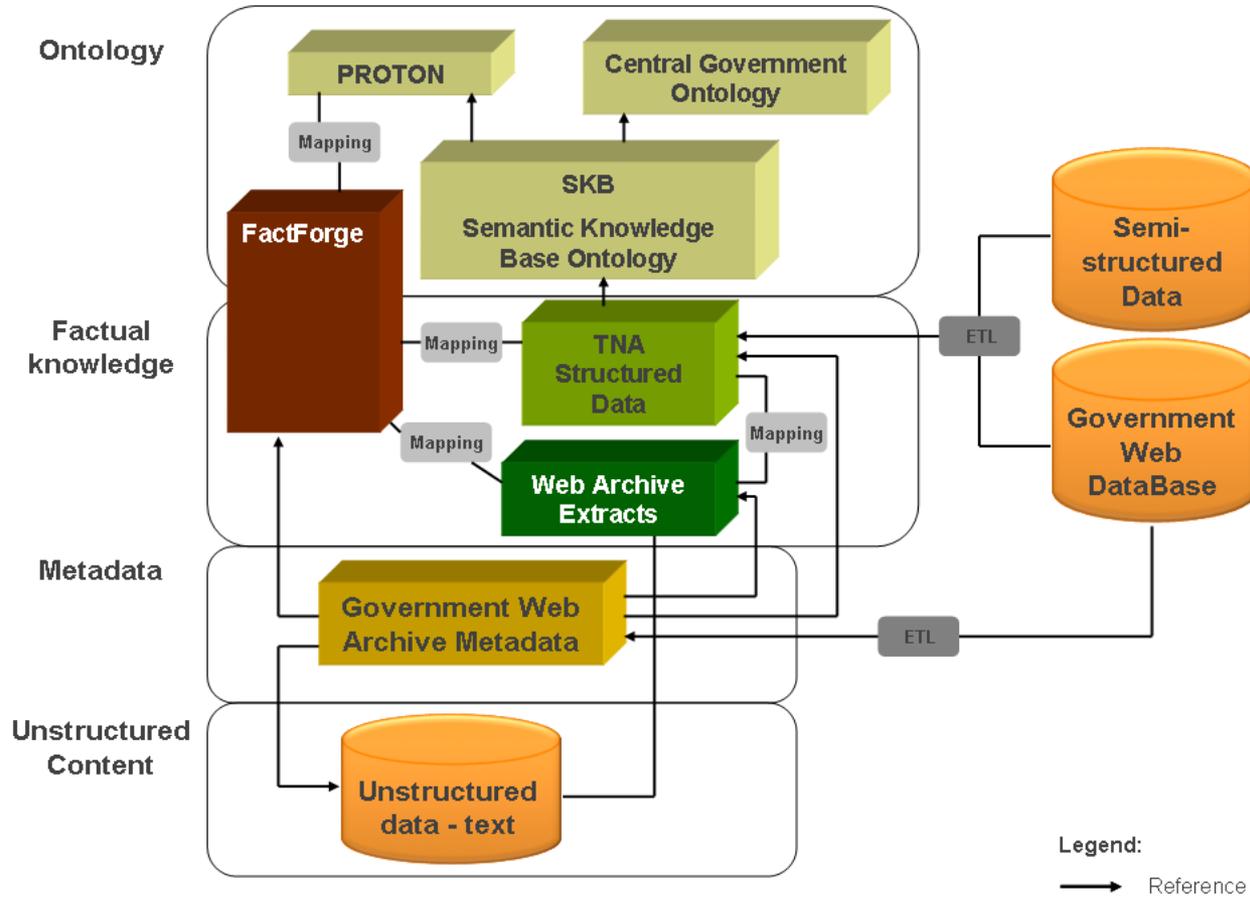
The Semantic Knowledge Base

- Our mandate was to improve access to the information contained in the web archive by providing far richer and more sophisticated information about what it contains
- We worked with a consortium led by Ontotext AD, together with the GATE Team from the University of Sheffield and System Simulation.

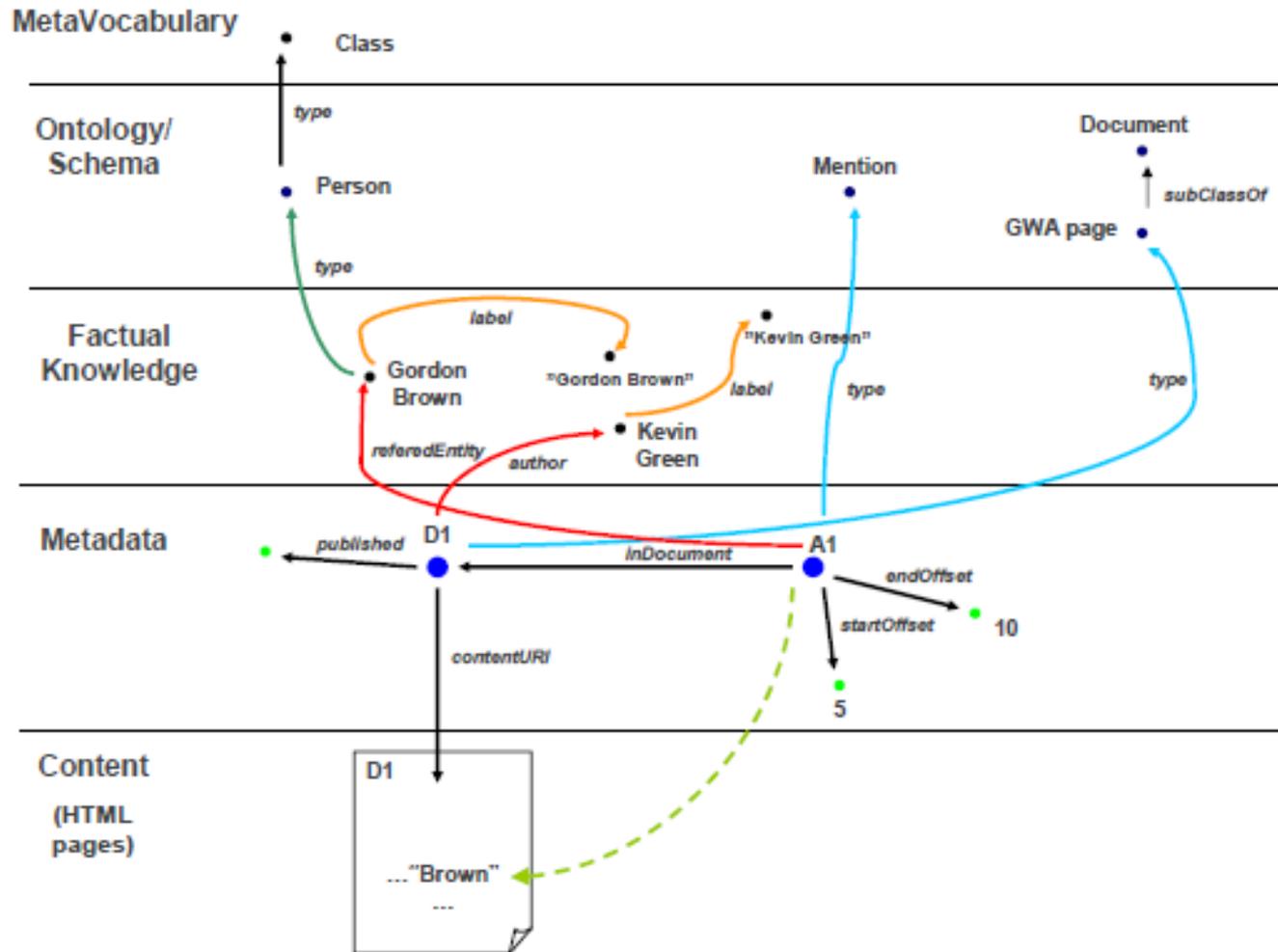
Semantic Search

- Structures unstructured content.
- Enables the system to understand the meaning of concepts and their relationships.
- Explicit facts can be attributed to otherwise loose terms.
- Technology used is open and machine-readable.

The Semantic Knowledge Base – Conceptual Architecture



Example – Gordon Brown



Sketching out User Stories

Key to project is to build something our users find useful.

Therefore, we built user stories based on survey results, experience of web archiving team and understanding of potential of technology.

Launching the System

- We will develop a user interface for non-technical use.
- We will also release an Application Programming Interface (API) for developers.

Thank you

Please visit www.nationalarchives.gov.uk/webarchive/ to see the web archive and for more information

webarchive@nationalarchives.gsi.gov.uk

Any Questions ?

Examples

Redirection:

<http://www.nationalarchives.gov.uk/news/stories/9.htm>

<http://www.dh.gov.uk/en/Healthcare/Primarycare/index.htm>

Treasury 1997 & 2010

<http://web.archive.org/web/19970707050911/http://www.hm-treasury.gov.uk/index.html>

<http://webarchive.nationalarchives.gov.uk/20100401161705/http://www.hm-treasury.gov.uk/>

General Election: Before & After

<http://webarchive.nationalarchives.gov.uk/20100505180303/http://www.number10.gov.uk/>

<http://webarchive.nationalarchives.gov.uk/20100512155632/http://www.number10.gov.uk/>