

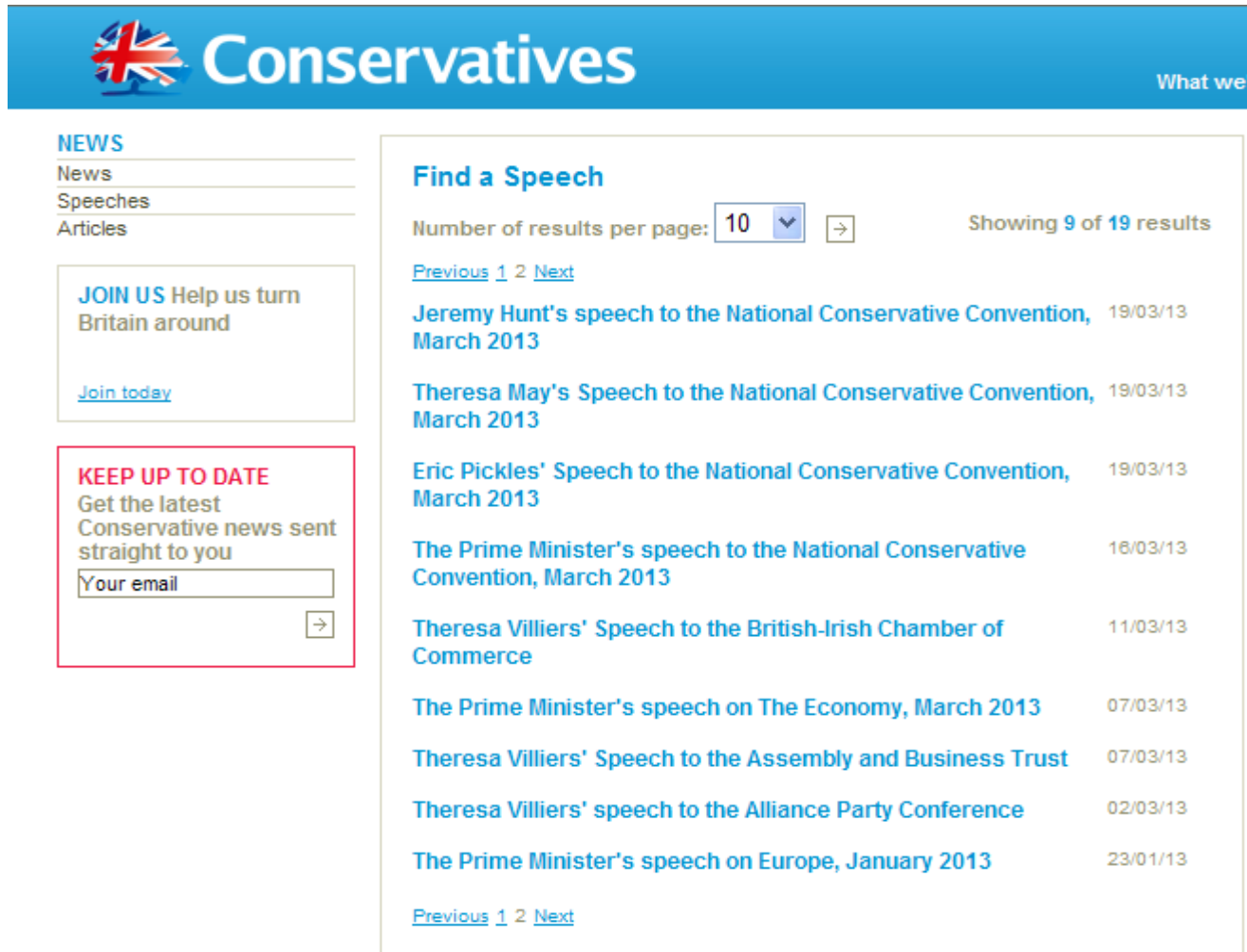
Web archiving at the British Library

Peter Webster (British Library)

@pj_webster / @UKWebArchive

webarchive.org.uk

The missing web ?



The screenshot shows the Conservative Party website. The header is blue with the Conservative Party logo (a Union Jack) and the word "Conservatives" in white. To the right of the header is the text "What we".

On the left side, there is a "NEWS" section with links to "News", "Speeches", and "Articles". Below this is a "JOIN US" section with the text "Help us turn Britain around" and a "Join today" link. Further down is a "KEEP UP TO DATE" section with the text "Get the latest Conservative news sent straight to you" and an email subscription form.

The main content area is titled "Find a Speech". It includes a "Number of results per page:" dropdown menu set to "10" and a "Showing 9 of 19 results" indicator. Below this are navigation links "Previous", "1", "2", and "Next".

The list of speeches is as follows:

Speech Title	Date
Jeremy Hunt's speech to the National Conservative Convention, March 2013	19/03/13
Theresa May's Speech to the National Conservative Convention, March 2013	19/03/13
Eric Pickles' Speech to the National Conservative Convention, March 2013	19/03/13
The Prime Minister's speech to the National Conservative Convention, March 2013	16/03/13
Theresa Villiers' Speech to the British-Irish Chamber of Commerce	11/03/13
The Prime Minister's speech on The Economy, March 2013	07/03/13
Theresa Villiers' Speech to the Assembly and Business Trust	07/03/13
Theresa Villiers' speech to the Alliance Party Conference	02/03/13
The Prime Minister's speech on Europe, January 2013	23/01/13

At the bottom of the list are navigation links "Previous", "1", "2", and "Next".

<http://www.conservatives.com/News/SpeechList.aspx?>

The missing web ?



Polly Toynbee @pollytoynbee

19 Nov

[@Cormjah](#) Many thanks for finding this! Here is lost Tory archive, lest we forget days of 'compassionate' Cameron webarchive.org.uk/ukwa/target/10...

Followed by [Helen Hockx](#)

[Collapse](#)
[Reply](#)
[Retweet](#)
[Favorite](#)
[Buffer](#)
[More](#)

2 RETWEETS	2 FAVORITES	  
---------------	----------------	--

11:02 AM - 19 Nov 13 · Details

<http://www.conservatives.com/News/SpeechList.aspx?>

The missing web saved



External links, forms and search boxes may not function within this archived website.

http://www.conservatives.com/News/Speeches/2009/07/David_Cameron_One_Wi [Go](#)

JUNJULDEC »

18

200820092010

[Close](#) [Cymraeg](#) [Help](#)

 **Conservatives**

[Policy](#) [People](#) [News](#) [Video](#) [Campaigns](#) [Where you live](#) [Get involved](#) [Donate](#) [Show your support](#)

[→](#)

NEWS

[News](#)
[The Blue Blog](#)
[Speeches](#)
[Articles](#)
[Photo Galleries](#)

Find a speech

[▼](#)

or

[▼](#)

or

[▼](#) [▼](#)

[→](#)

Speech

David Cameron: One World Conservatism

David Cameron, Monday, July 13 2009



Today Britain continues to fight a terrible recession. We're facing a debt crisis unprecedented in our peacetime history.

The country understands that public spending will have to be cut.

So I understand why people might say that at a time like this, in economic circumstances like these increasing the amount we spend on foreign aid is the last thing we should do.

But I want to explain today why I think - as I said at the start of the year - that economic difficulties at home should be the time for us to re-affirm our moral responsibilities, not reduce them.

Making big decisions is about doing what's right, not what is convenient.

And it's about ensuring that short-term pressures don't lead you to make long-term mistakes.

So yes - people at home are hurting in this recession.

But they understand that there are still billions who have the tiniest fraction of what we have, even in a recession, and that it's our social responsibility to help them.

And yes - we have a huge debt to pay off thanks to Labour's economic incompetence.

But paying down the debt must not mean pushing down the poor - at home or abroad.



David Cameron

David was elected Leader of the Conservatives in December 2005, on a mandate to change the Party and change the country.

[Read more](#)



Conservatives News

Conservative volunteers travel to Rwanda for Project Umubano

Friday, July 17 Andrew Mitchell will lead a group of around 100 volunteers for the third year of Project Umubano

<http://webarchive.org.uk>

The missing web: individuals

Vote For...

DAVID CAMERON

CONSERVATIVE CANDIDATE FOR THE WITNEY CONSTITUENCY

- About David Cameron
- At Home
- Message to Voters
- Policies
- Latest News



Working for West Oxfordshire

"I will always champion local causes, help local people and work for the good of West Oxfordshire."

Press Release - Council Tax Revaluation

[Click here to read on.](#)

What the Conservatives will do:

- 1 Lower taxes »
- 2 School discipline »
- 3 »

David has:

- Worked with our county council to do more for forces schools in Carterton
- Pressed the government to save special schools for the learning disabled
- Spoken up in Parliament for farmers markets and British food
- Organised public meetings and lobbied FGW for improved rail services
- Campaigned for improved rights for park home residents
- Continually spoken up for our community hospitals
- Taken delegations to Ministers about forces homes at RAF Brize Norton
- Raised awareness about Group B Strep, an infection that kills new born babies
- Campaigned for rural broadband, bringing BT to Witney for key meetings

Click to email...

If you would like to contact David, or help with his campaign, please get in touch:

E-mail: ask@votedavidcameron.com

Telephone: 01993 702 302

Fax: 01993 776 639

Address: Conservative Campaign Centre

votedavidcameron.org (archived 24/5/05) at UK Web Archive

The missing web: organisations

[Sitemap](#)

[Home](#)
[About us](#)
[What we do](#)
[News](#)
[Events](#)
[Committees and meetings](#)
[Police and Crime Commissioner](#)
[Corporate governance](#)
[Finance](#)
[Transparency agenda](#)
[Value for Money](#)
[Conduct and complaints](#)
[Working with you](#)
[Stop and Search](#)
[Custody visiting](#)
[Policing Plan](#)



Thames Valley Police Authority



An **independent body** of local people who **oversee the police**

[→ Search](#)

[What we do](#) [About us](#) [Working with the public](#)

Maintaining an efficient and effective police service.

As a Police Authority, our primary responsibility is to maintain an efficient and effective police service for the people of Thames Valley by monitoring, supporting and scrutinising the Force.

[→ More about what we do](#)

★ Latest updates

Welcome to the Police and Crime Commissioner

Thames Valley Police Authority welcomes the new Police and Crime Commissioner (PCC) for Thames Valley.
19 Nov 12

[→ View all updates](#)

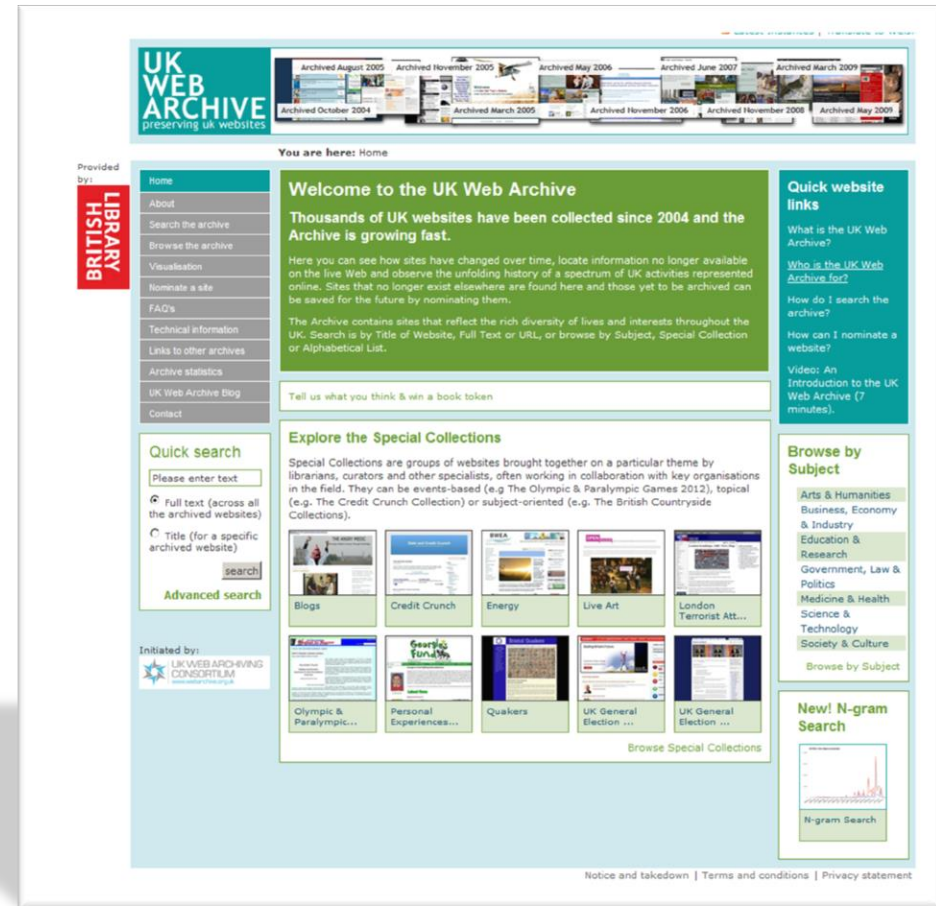
TVPA meeting calendar

[→ View](#) all meetings in which the TVPA is involved

tpa.police.uk (archived 21/11/12) at UK Web Archive

UK Web Archive

- Selective archiving since 2004
- Sites of cultural or scholarly importance for the UK
- 13,400 sites, 61,000 instances, 20TB of data
- British Library, National Library of Wales, JISC
- Plus many collaborators: Women's Library, Live Art Development Agency, NHS
- <http://webarchive.org.uk>



Web archiving: the basics

What

- Selecting, capturing, storing, preserving and managing access to snapshots of websites over time

How

- Use crawler software to download websites automatically
- Selective or domain archiving
- Provide access in a Web Archive

When

- Since mid 1990s

Who

- Heritage and memory organisations, eg BL, The National Archives
- University libraries
- Not-for-profit and commercial organisations, eg Internet Archive
- Individual researchers

Why

- Global information resource
- Artefact of cultural and technology change
- Representative sample of the web: historical and sociological data that may not be found elsewhere
- Part of national digital heritage - legal requirements

A lost website, saved

UKWA External links, forms and search boxes may not function within this archived website.

http://www.votedavidcameron.com/index.html Go

2 captures 24 May 05 - 11 May 10

APR MAY MAY »

2004 2005 2010 »

Close X

Cymraeg

Help ?

Vote For...

DAVID CAMERON

CONSERVATIVE CANDIDATE FOR THE WITNEY CONSTITUENCY

- About David Cameron
- At Home
- Message to Voters
- Policies
- Latest News



Working for West Oxfordshire

"I will always champion local causes, help local people and work for the good of West Oxfordshire."

Press Release - Council Tax Revaluation
[Click here to read on.](#)

What the Conservatives will do:

David has:

- Worked with our county council to do more for forces schools in Carterton
- Pressed the government to save special schools for the learning disabled
- Spoken up in Parliament for farmers markets and British food
- Organised public meetings and lobbied FGW for improved rail services
- Campaigned for improved rights for park home residents

ask@votedavidcameron.com

If you would like to contact David, or help with his campaign, please get in touch:

votedavidcameron.org (archived 24/5/05) at UK Web Archive

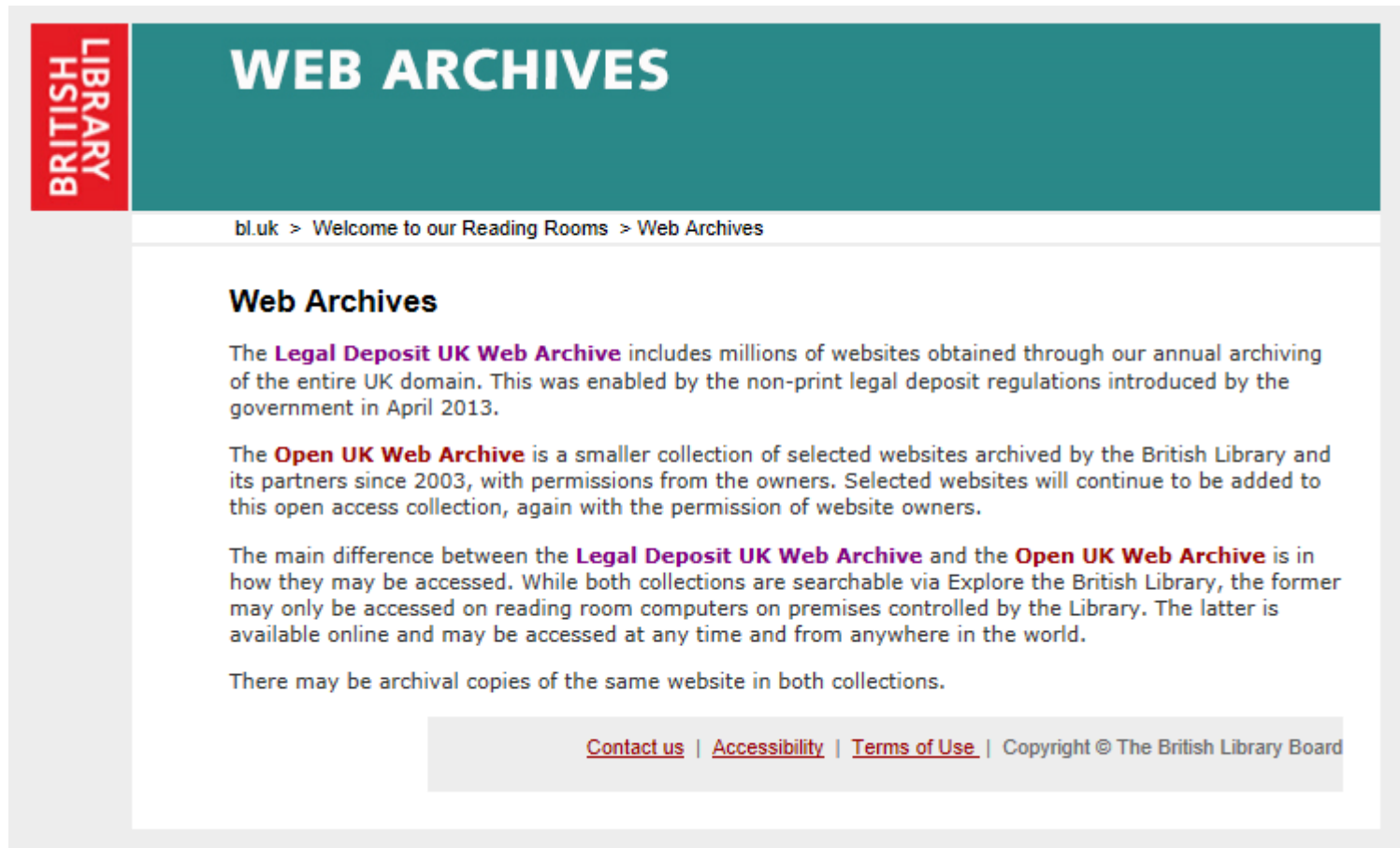
Non-print legal deposit, before and after: what has changed ?

	BEFORE	AFTER
Scale	14,000	4 – 5 million
Workflow (and tools)	Selection prior to harvesting	Selection / curation can happen <i>after</i> harvesting
Permission to archive	Required	Can collect in-scope material without permission
Access	Online	Reading rooms only (unless with direct permission for online access)
Ownership	British Library	Legal Deposit Libraries

Progress: domain crawl

- 1st Legal Deposit domain crawl, April – June 2013
 - Started with 3.8 million seeds
 - Ran between 8th April - 21st June and collected over 31TB data
 - 4.2 million hosts
 - c.1.2 billion resources

Access: via reading room pages

A screenshot of the British Library's 'Web Archives' page. The page has a teal header with the 'WEB ARCHIVES' title in white. On the left, there is a red vertical bar with the 'LIBRARY HSIIRB' logo. Below the header, a breadcrumb trail reads 'bl.uk > Welcome to our Reading Rooms > Web Archives'. The main content area is white and contains three paragraphs of text. The first paragraph describes the 'Legal Deposit UK Web Archive'. The second paragraph describes the 'Open UK Web Archive'. The third paragraph compares the two archives. At the bottom, there is a grey footer bar with links for 'Contact us', 'Accessibility', and 'Terms of Use', followed by a copyright notice for 'The British Library Board'.

<http://www.bl.uk/rroomwelcome/webarchives.html>

LDUKWA access tool : search results

The screenshot shows the UK Web Archive Legal Deposit Web Archive Search Tool interface. The header includes the UK Web Archive logo and the text 'Legal Deposit Web Archive Search Tool'. Below the header are navigation links: Home, Browse Archive, About, FAQ, and Search Tips. The main content area shows the search results for the query 'influenza' on the domain 'southworceccg.nhs.uk'. The results are displayed in a list format, showing the title, crawl date, host, and content type for each result.

Legal Deposit Web Archive Search Tool

UK WEB ARCHIVE
preserving uk websites

Home Browse Archive About FAQ Search Tips

Home » "southworceccg.nhs.uk"

Refine by domain:

- (-) southworceccg.nhs.uk

Refine by domain suffix:

- nhs.uk (17265)

Refine by crawl year:

- 2013 (17265)

Enter search terms:

influenza Search

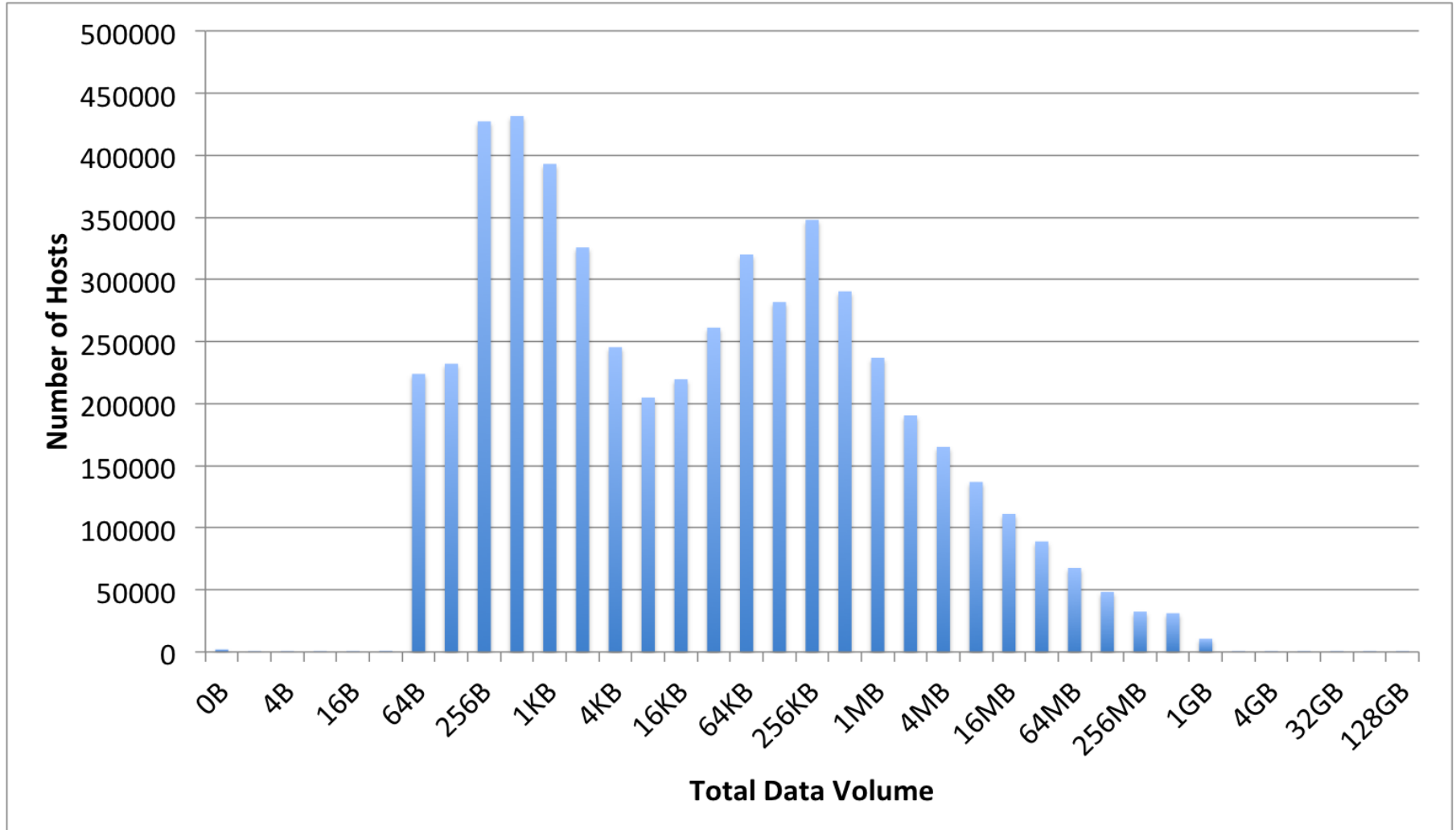
Displaying 1 - 10 of 17265

1) Title: South Worcestershire Clinical Commission...
Crawl Date: 2013-04-09T01:35:03Z [Other crawl dates]
Host: southworceccg.nhs.uk
Content Type: html

2) Title: South Worcestershire Clinical Commission...
Crawl Date: 2013-04-09T01:35:32Z [Other crawl dates]
Host: southworceccg.nhs.uk
Content Type: html

3) Title: South Worcestershire Clinical Commission...
Crawl Date: 2013-04-09T01:39:56Z [Other crawl dates]
Host: southworceccg.nhs.uk

What does the UK web look like ?



JISC UK Web Domain Dataset 1996-2013

- Funded by JISC to create a research collection of UK websites
- Collaboration between the Internet Archive, JISC and the British Library
- Copy of subset of the Internet Archive's web collection that relates to the UK
- c.300 million resources, 60TB in total
- No local access – possible through the Internet Archive
- Can be used to generate secondary datasets

Prototype search for UK Domain Dataset

**UK
WEB
ARCHIVE**
preserving uk websites

AADDA
Analytical Access to the Dark Domain Archive (1996-2010)

Home Browse

Home » Browse » "news.bbc.co.uk" » "www.stopwar.org.uk" » "2006"

Crawl year:

- (-) 2006

Host:

- (-) news.bbc.co.uk

Filter by private_suffix:

- bbc.co.uk (1)

Public suffix:

Browse

Filter by search term

"Blair iraq Galloway" ~100

Search

Search found 1 item

- (-) domain: "news.bbc.co.uk"
- (-) links_hosts: "www.stopwar.org.uk"
- (-) crawl_year: "2006"

Title: [BBC NEWS | UK | Thousands stage Iraq dem...](#)

Crawl Date: 2006-06-29T21:01:08Z [more]

Host: news.bbc.co.uk

Content Type: html

Links to host:

- (-) www.stopwar.org.uk
- news.bbc.co.uk (1)
- newsvote.bbc.co.uk (1)
- newswm.bbc.co.uk (1)
- ssl.bbc.co.uk (1)

[Show more](#)

Filter by
links_private_suffixes:

- bbc.co.uk (1)
- cnduk.org (1)
- mabonline.net (1)
- met.police.uk (1)
- stopwar.org.uk (1)

Archived site in Internet Archive

INTERNET ARCHIVE
wayBack Machine

http://news.bbc.co.uk/1/hi/uk/3143062.stm Go

16 captures
29 Sep 03 - 18 Sep 12

SEP JUN MAY
2004 29 2006 2008

bbc.co.uk Home TV Radio Talk Where I Live A-Z Index Search

BBC NEWS WATCH BBC NEWS IN VIDEO

UK version International version About the versions Low graphics Help Contact us

Last Updated: Saturday, 27 September, 2003, 19:49 GMT 20:49 UK

E-mail this to a friend Printable version

Thousands stage Iraq demo

Thousands of protesters gathered for an anti-war rally in London calling for an end to the US and UK occupation of Iraq.

The march started in Hyde Park and was followed by a rally in Trafalgar Square addressed by speakers including Tony Benn, filmmaker Ken Loach and Labour MP George Galloway.

The organisers said they wanted to make Labour respond to their anger over the situation in Iraq, ahead of the party's annual conference.

Police estimated there were about 20,000 marching, although organisers estimated up to 100,000 had attended the event.

Many of the protesters chanted anti-Bush slogans and carried banners with messages like "Blair must go" and "UK troops out of Iraq".



Saturday's march was addressed by leading anti-war activists

WATCH AND LISTEN
The BBC's Philippa Young
"The first major protest since the fall of Saddam Hussein"
VIDEO

THE STRUGGLE FOR IRAQ In Depth

KEY STORIES

- Iraq PM positive on amnesty plan
- Russia 'to kill Iraq kidnappers'
- Samarra shrine suspect held
- Bombing hits northern Iraqi city
- Violence in Iraq leaves 25 dead

ZARQAWI KILLED

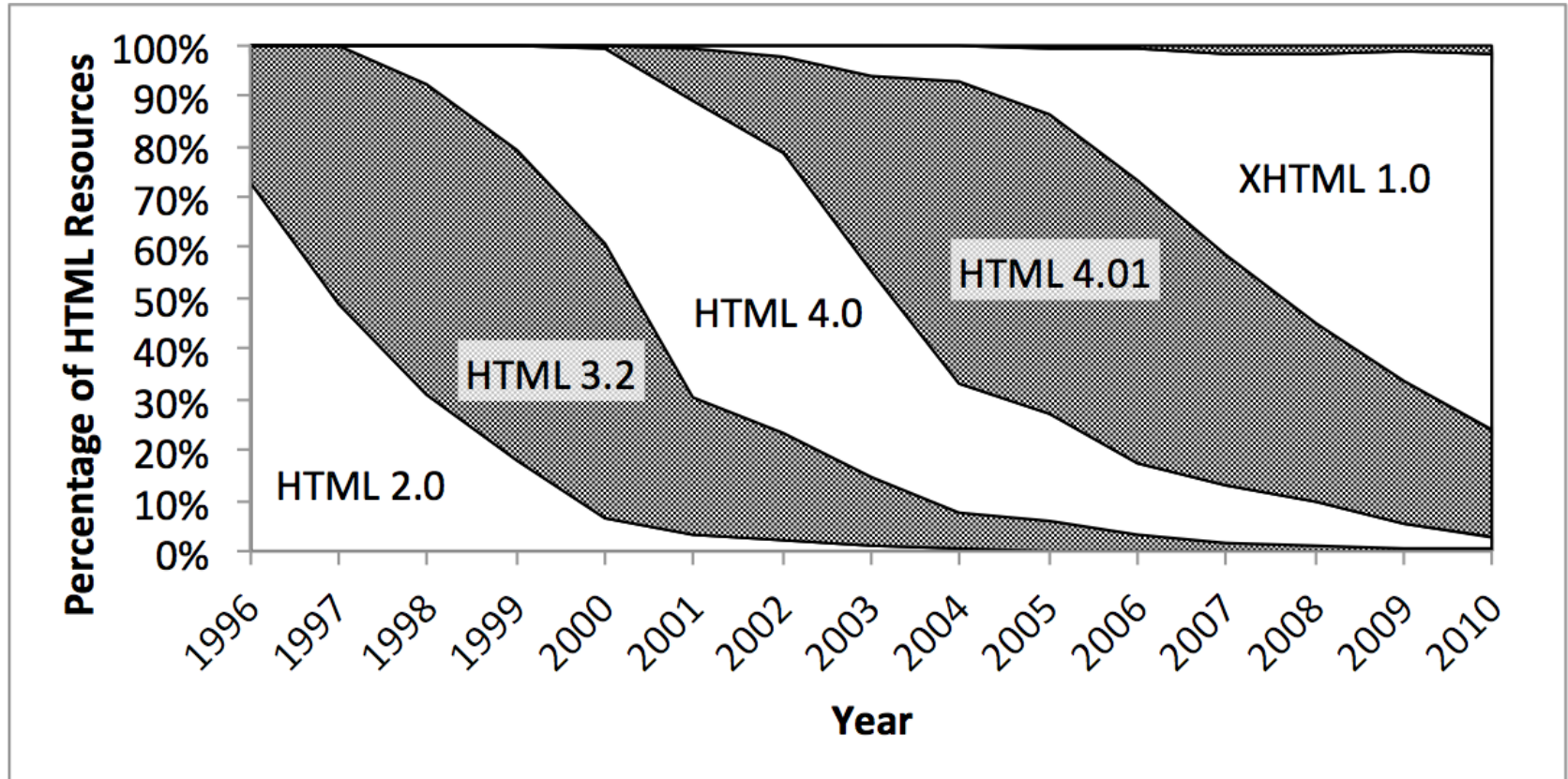
- Iraqis give their reaction
- Al-Qaeda may change tactics
- How Zarqawi was tracked down

BACKGROUND AND ANALYSIS
New dawn?
Security crackdown offers ray of hope to Baghdad residents

RELATED BBC SITES
SPORT
WEATHER
CBBC NEWSROUND
ON THIS DAY

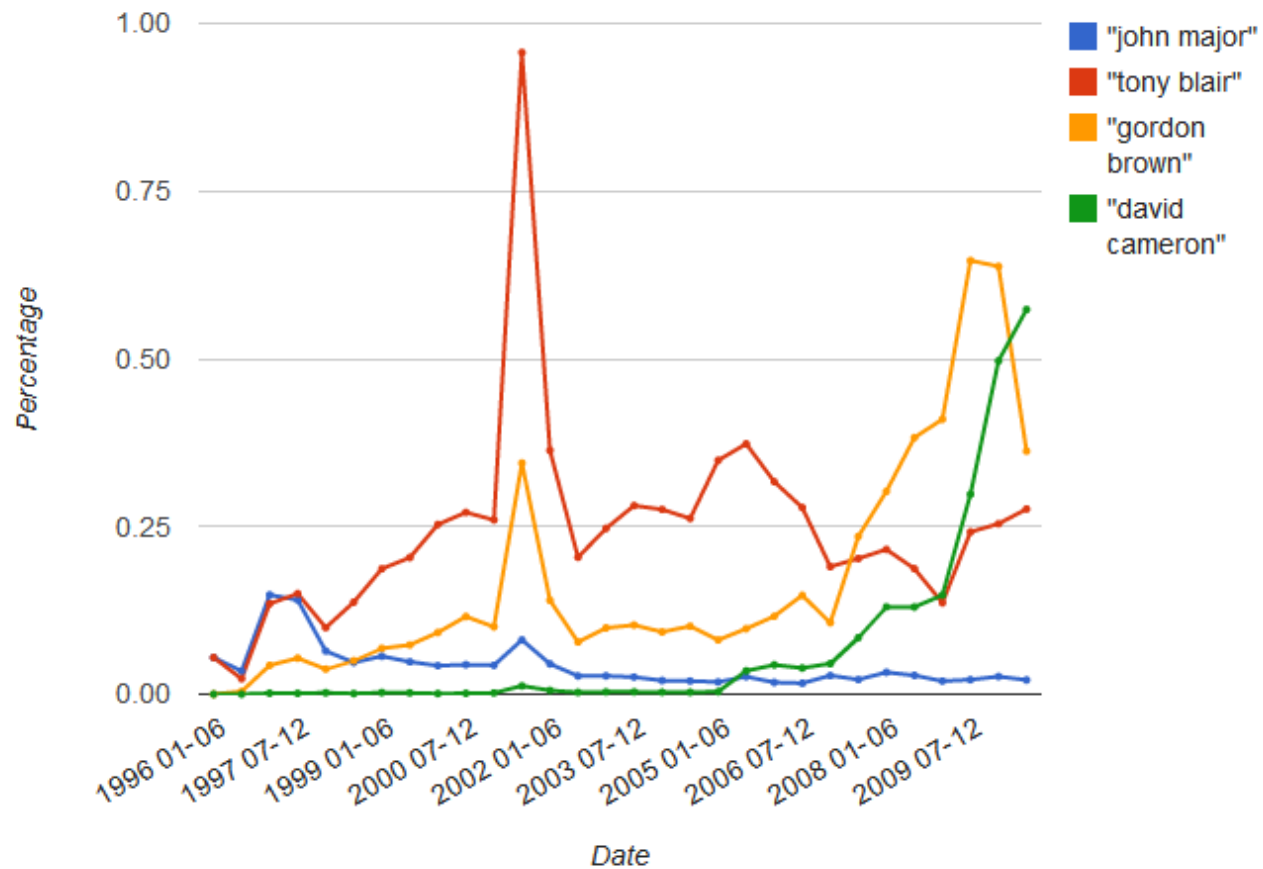
News Front Page
World
UK
England
Northern Ireland
Scotland
Wales
Business
Politics
Health
Education
Science/Nature
Technology
Entertainment
Have Your Say
Magazine
In Pictures
Country Profiles
In Depth
Programmes

HTML version analysis



<http://www.webarchive.org.uk/ukwa/visualisation/ukwa.ds.2/fmt>

Ngram: Prime Ministers



<http://www.webarchive.org.uk/ukwa/ngramia/>

Datasets available for download

The host link graph

1996 | appserver.ed.ac.uk | portico.bl.uk 1

1996 | art-www.acorn.co.uk | portico.bl.uk 1

1996 | astra.ich.ucl.ac.uk | portico.bl.uk 1

1996 | back.niss.ac.uk | portico.bl.uk 1

1996 | beta.bids.ac.uk | portico.bl.uk 2

19GB (130GB unzipped), at: <http://tinyurl.com/kon2eve>

An archbishop in hot water

BBC Home
Search
Explore the BBC

Low graphics | Accessibility help

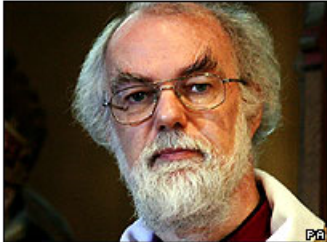
BBC NEWS
LIVE LT BBC NEWS CHANNEL

News services
Your news when you want it

Last Updated: Thursday, 7 February 2008, 16:55 GMT
E-mail this to a friend | Printable version

Sharia law in UK is 'unavoidable'

The Archbishop of Canterbury says the adoption of certain aspects of Sharia law in the UK "seems unavoidable".



Dr Williams says Muslims should have a choice in legal disputes
Listen Dr Williams interview

Dr Rowan Williams told Radio 4's World at One that the UK has to "face up to the fact" that some of its citizens do not relate to the British legal system.

Dr Williams argues that adopting parts of Islamic Sharia law would help maintain social cohesion.

For example, Muslims could choose to have marital disputes or financial matters dealt with in a Sharia court.

He says Muslims should not have to choose between "the stark alternatives of cultural loyalty or state loyalty".

'Sensational reporting'

In an exclusive interview with BBC correspondent Christopher Landau, ahead of a lecture to lawyers in London on Monday, Dr Williams argues this relies on Sharia law being better understood.

SHARIA LAW AND THE UK

LATEST NEWS

- Sharia law 'could have UK role'
- Cameron steps into Sharia row
- Williams in synod Sharia address
- PM praises archbishop's integrity
- Carey weighs into Sharia law row
- Sharia law in UK is 'unavoidable'
- Reaction in quotes: Sharia row

BACKGROUND


Guilty of innocence
Analysis of what the Archbishop of Canterbury did wrong

- Q&A: Sharia law
- PM extends support to archbishop
- Archbishop's full speech
- How a British Sharia court works
- Sharia law around the world
- Bradford's Muslims puzzled
- Profile: Dr Rowan Williams
- Religious courts already in use

HAVE YOUR SAY

- Should UK adopt aspects of Sharia law?

VIDEO AND AUDIO

- Watch Lord Chief Justice
- Listen Muslims on UK Sharia

News Front Page
World
UK
England
Northern Ireland
Scotland
Wales
Business
Politics
Health
Education
Science & Environment
Technology
Entertainment
Also in the news
Video and Audio
Have Your Say
Magazine
In Pictures
Country Profiles
Special Reports
RELATED BBC SITES
SPORT
WEATHER
CBBC NEWSROUND
ON THIS DAY
EDITORS' BLOG

Inbound links to Canterbury site

The host link graph

2001 | itn.co.uk | archbishopofcanterbury.org 1

2006 | dioceseofyork.org.uk | archbishopofcanterbury.org 19

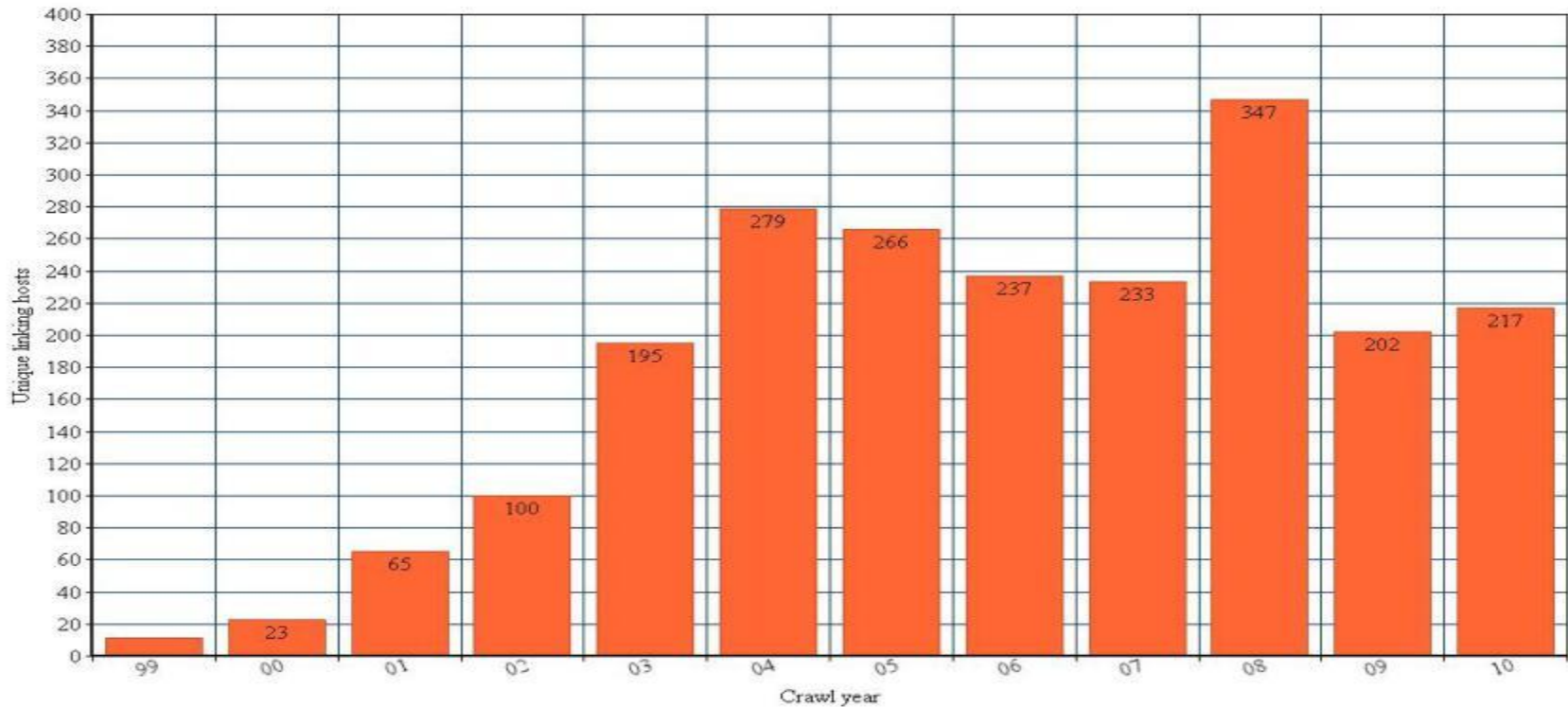
2008 | divinity.cam.ac.uk | archbishopofcanterbury.org 11

2004 | secularism.org.uk | archbishopofcanterbury.org 3

... and c.2.5k others

Watching the news from a distance

Hosts linking to archbishopofcanterbury.org



<http://peterwebster.me/category/web-archiving/>

Methodological challenges: what is in the archive ?

- National web archives: some selective, some legal deposit
- When is comprehensive not comprehensive ?
- Defining the national (<http://tinyurl.com/m9ue5gw>)

Methodological challenges: *when* was it in the archive ?

- Understanding the crawl profile
- Crawl date NOT publication date
- Citation standard: what, when archived

Thank you !

Peter.Webster@bl.uk

@pj_webster / @UKWebArchive / @netpreserve

britishlibrary.typepad.co.uk/webarchive