

Archiving the UK Web

Learning on the Job

UK WEB ARCHIVE

preserving UK websites

LIBRARY
HSLIBR

Provided by

Victoria and Albert Museum (Jun. 2006)



Route Masters (Apr. 2006)



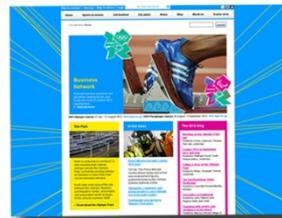
BBC - WW2 People's War (May 2006)

Preserving UK's national memory

Egg Bacon Chips and Beans (Feb. 2007)



London 2012 Olympic and Paralympic Games (Feb. 2008)



Helen Hockx-Yu
Head of Web Archiving
The British Library

July 2011

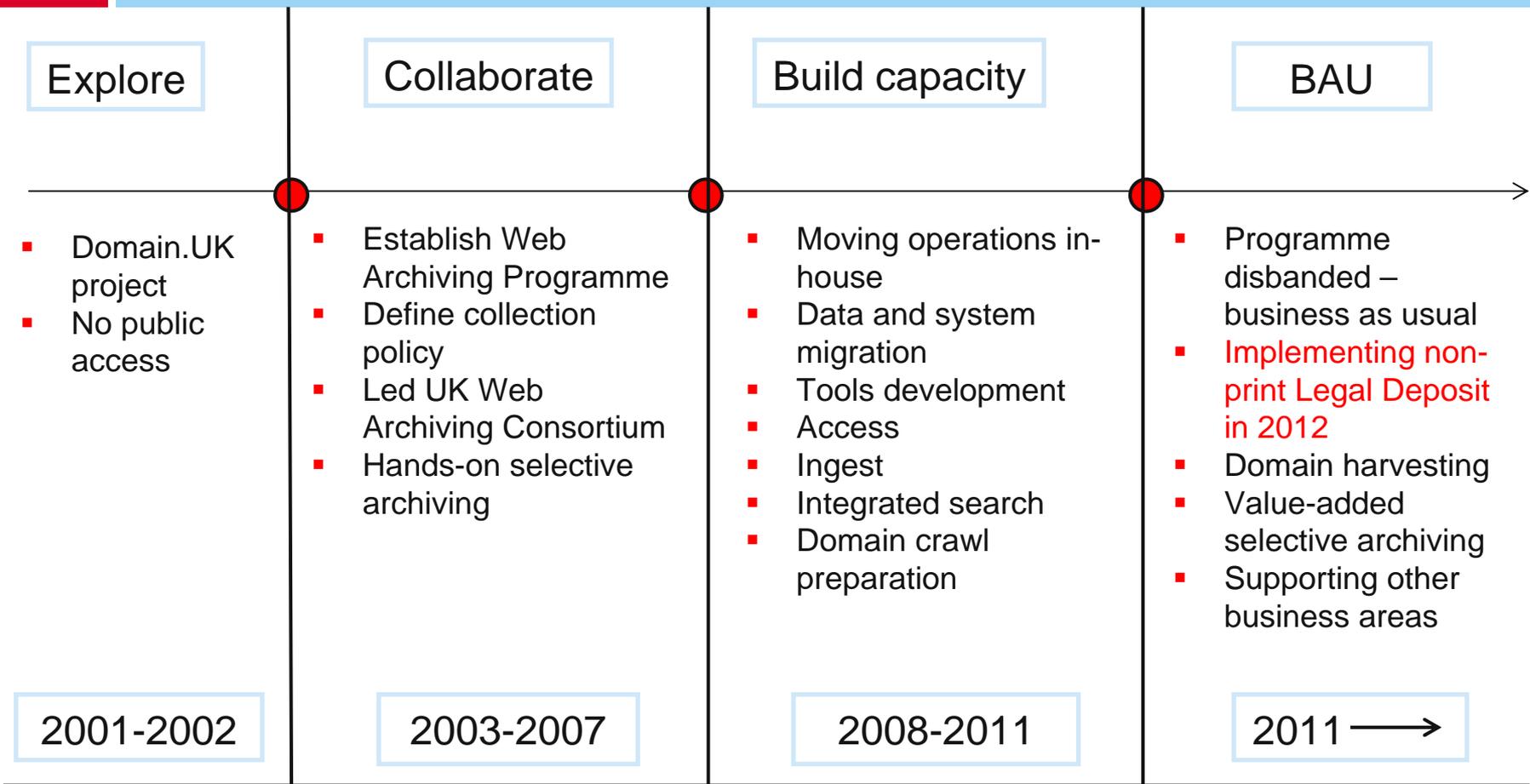
The UK context

- 9 million .uk domain names registered with Nominet in December 2011
- ~ another 1 million using other domain names, e.g. .com
- Growing between 12%-17% per year

- 2 Web Archives in the UK – covering ~1% of the UK domain
 - UK Web Archive provided by the British Library
 - UK Government Web Archive provided by the National Archives

- Legislation for non-print Legal Deposit not yet in place
 - Currently BL has to obtain permissions prior to archiving
 - LD Expected April 2012
 - Access expected to be restricted to Legal Deposit Libraries' reading rooms only – 6 LDLs in the UK

10 years of web archiving – a learning process



The UK Web Archive

- 10,027 websites, 37,879 instances by June 2011
- Streaming media solution (eg [One and Other](#))
- Earliest archived website instance in the world: British Library website archived on 18 April 1995
- Comprehensive search and browse functions including full-text search
 - full-text search at archive, collection and website levels
 - browse by subject, special collection, title
 - Alternative access methods, eg tags clouds, 3D wall and N-gram search
- Increasing usage of the archive
 - 167,726 visits and 683,966 page views between 1st April 2010 - 31st March 2011 (~50% increase)

And much more

- Capability and know-how
- Unified file format (including migration from legacy format)
- Tools development
 - Web Curator Tool (WCT)
 - Selection and Permission Tool (SPT)
- Integration with Primo for resource discovery
- Ingest of web archive into Digital Library System for long term preservation
- Preparations for non-print Legal Deposit
 - Domain crawl
 - Deposited websites
- Develop access and understand researcher needs
 - UK Web Archive and Researchers Project

Building web archiving capability

In 2020, the British Library will be a leading hub in the global information network, advancing knowledge through our collections, expertise and partnerships, for the benefit of the economy and society and the enrichment of cultural life.

- Capability is people, systems and processes
- Hire good people with wider set of skills (eg industrial experience)
- Make sure good people remain good – training
 - Linux, Java, Hadoop, SOLR
 - Visits to other web archiving teams
- Bridging the “digital divide“
- Know how things work - in possession of core expertise
- Workflows, policies & guidelines, roles and responsibilities
- Not something one-off, need to respond to the evolving web continuously

Placing web archive in the Library context

British Library Search - Robin Cook - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://search.bl.uk/primo_library/libweb/action/search.do?dscnt=0&vl(174399379UI0)=any&scp.s

British Library Search - ...

Guest My workspace Personal settings Log in

SEARCH OUR CATALOGUE

Home About Feedback Tags Find eJournals My Reading Room Requests Help

Robin Cook All Content Search Advanced search

Anywhere

rss

Show only
Online resources (6)

Refine my results

Material type
Audio Visual (134)
Books (101)
Articles (45)
Scores (2)
Theses (2)
Archived Websites (1)

Author/Contributor
Cook, Robin, 1940- (31)
Cook, Robin (22)

285 results for All Content Sorted by: relevance 1-10 Next

1 **Robin Cook / John Kampfner.** John Kampfner London : Phoenix, 1999, c1998. Book Details I want this Notes & Tags

2 **Robin Cook** 08/08/2005- Archived Website Details I want this Notes & Tags

3 **Robin Cook : a critical companion / Lorena Laura Stookey.** Lorena Laura Stookey Westport, Conn. ; London : Greenwood, 1996.

There are 2 versions of this item

Remember who we are doing this for

RESEARCHERS & UK WEB ARCHIVE BLOG

bl.uk Main About this blog UK Web Archive website Subscribe

UK WEB ARCHIVE
preserving uk websites

Scholars from different disciplines conduct individual investigations into the potential of the UK Web Archive for new research.

ARCHIVES

January 2011
December 2010
November 2010
October 2010

CATEGORIES

Web/Tech

RECENT POSTS

25 January 2011

'Invisible women' in sport as reflected by websites

Kath Woodward, Professor of Sociology at the Open University, is a member of the Researchers and UK web archive project. She is writing here about a specific issue for sports researchers in finding information – the invisibility of women. (For Kath Woodward's profile, see http://www.cresc.ac.uk/people/prof-kath-woodward_and http://www.open.ac.uk/socialsciences/staff/people-profile.php?name=Kath_Woodward)

Demonstrating the invisibility of women in the public arena of representation has a long history on feminist critiques. Women have been 'hidden from history' to quote the title of Sheila Rowbotham's famous 1975 book, or they have been an 'absent presence', as claimed in the work of French psychoanalytic feminists. History has been written as if only white men from the middle and upper classes were agents and protagonists who matter. We know the stories which feature men, because they are recorded, but women, if mentioned

- (Academic) researchers are the (targeted or perceived) main user group of the UK Web Archive
- Researchers and the UK Web Archive Project
 - 8 selected researchers take part in the project, all from UK HEIs
- Objectives
 - Engage researchers in use of the UK Web Archive as a resource for research
 - Develop the UK web archive in line with the requirements of its key user group
 - Develop special collections in collaboration with the researchers
 - Apply data analysis tools to support research
 - Disseminate findings of the project nationally and internationally
- Work with JISC and IA on the next phase, involving wider range of disciplines, eg web science, digital humanities

Web archives as historical documents for reference

Translate to Welsh

UK WEB ARCHIVE
preserving uk websites

You are here: [Home](#) > [Search](#) > British Library, The

British Library, The

This site is part of the following subject(s):
Education & Research > Libraries, Archives and Museums

This site was archived for preservation by the British Library.
The live site may provide more information.

Text Search

Search all instances by text

Instances

 Archived 18 Apr 1995	 Archived 07 Dec 2004	 Archived 16 Jul 2005	 Archived 29 Jul 2005	 Archived 12 Aug 2005	 Archived 09 Sep 2005
 Archived 23 Sep 2005	 Archived 07 Oct 2005	 Archived 21 Oct 2005	 Archived 07 Jan 2006	 Archived 20 Apr 2006	 Archived 12 Jun 2006
 Archived 21 Feb 2007	 Archived 17 Oct 2007	 Archived 19 Nov 2007	 Archived 02 Sep 2008	 Archived 09 Dec 2008	 Archived 24 Jul 2009
 Archived 23 Oct 2009	 Archived 27 Apr 2010	Sorry, no thumbnail yet	Sorry, no thumbnail yet		

Quick search

Please enter text

Title (for a specific archived website)

Full text (across all the archived websites)

Advanced search

Your comments

Please send your comments and suggestions about sites archived by British Library to web-archivist@bl.uk

Provided by:
BRITISH LIBRARY

PORTICO - online information about THE BRITISH LIBRARY

Welcome to [Portico](#), The British Library's Online Information Server.

[Current Portico Highlights](#)

Portico currently features the following:

- A preview of some forthcoming [exhibitions](#) at The British Library
- [Initiatives for Access](#) - An overview of The Library's programme of digitisation and networking projects
- News of a Major British Library Acquisition - [The Archive of John Evelyn](#)
- The British Library and the [St Pancras Building](#)
- [Science Technology and Innovation](#) - A Review of Recent Policy Developments
- [The Portico Gopher](#) - A guide to British Library events, services and collections
- A Guide to Further [World Wide Web Resources](#)

[More information about Portico](#)

We welcome your [comments and suggestions](#) on the development of this prototype.

Copyright © 1995, The British Library Board

portico@bl.uk

THE BRITISH LIBRARY
Explore the world's knowledge

We hold 14 million books, 920,000 journal and newspaper titles, 58 million patents, 3 million sound recordings, and so much more. Start exploring here.

SEARCH

Search tips and advanced searching

- British Library**
10,000 pages on our main website
- Online gallery**
30,000 treasures from our collection
- Catalogue records**
14 million items in our collections
- Journal articles**
9 million articles from 20,000 journals

Quick links | **What's on** | **Site highlights** | **Your library**

Magnificent Maps

- Opening times, maps
- Reader Registration
- Reading Rooms
- Help for researchers
- Online catalogues
- Information in foreign languages
- For higher education
- For entrepreneurs
- For librarians
- For publishers; legal deposit etc.
- Collection Care
- Press Room
- Contact us

News

- 26 Apr 2010 Magnificent Maps: latest
- 17 Apr 2010 Event: Stem Cells - Panacea?
- 8 Apr 2010 Guardian: Mervyn Peake archive

Opens Fri 30 April

- Preview it online
- Read Curators' blog

British Library websites | Please choose...

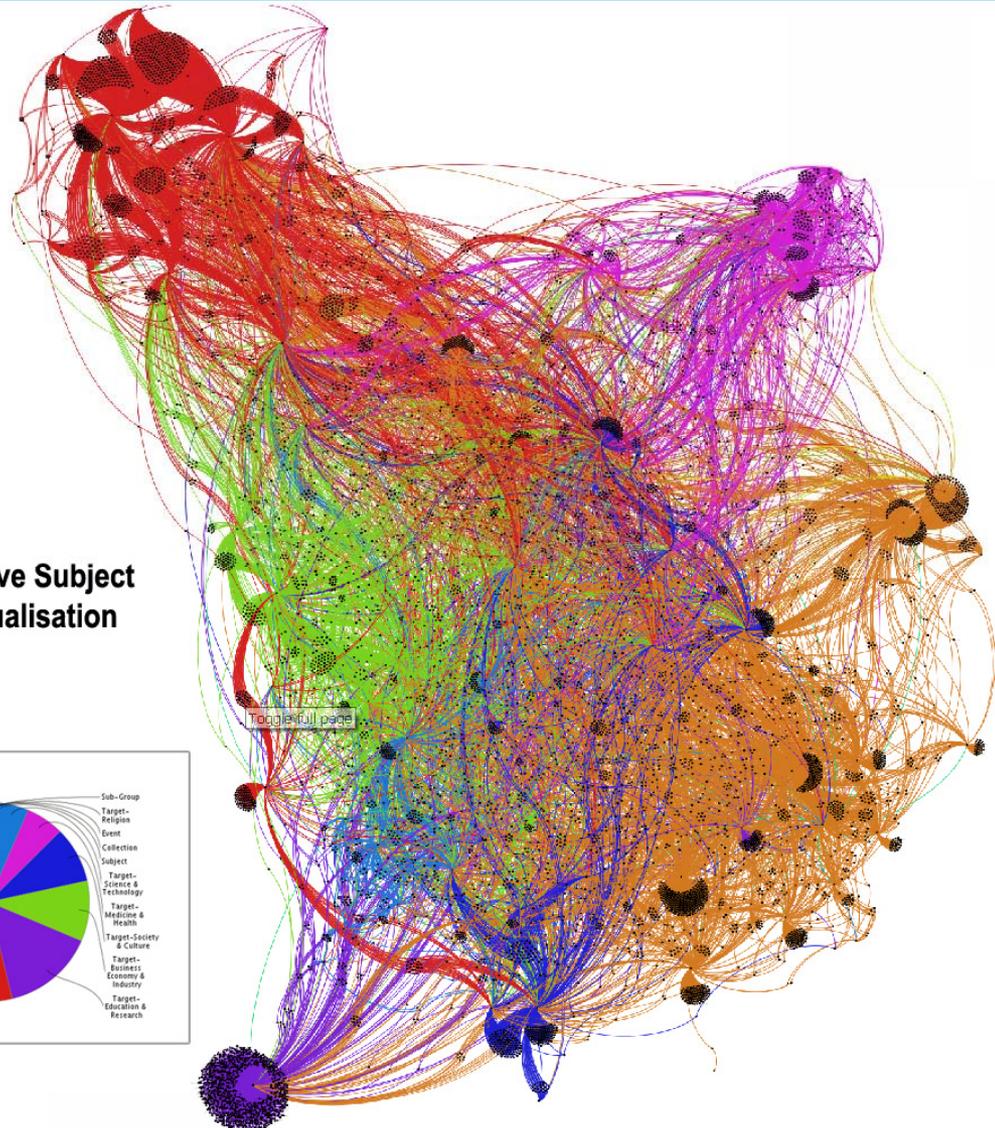
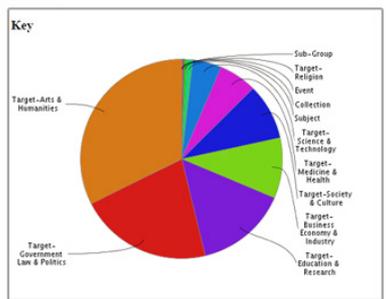
Showing the value of haystacks



- Web archives as datasets for analytics
- Focus on entirety of web archive collection (or subset) instead of individual websites
- Unlock embedded patterns, trends, context and relationships
- Use technology developed for the live web, eg visualisation
- Develop machine to machine access

Content visualisation

UK Web Archive Subject Hierarchy Visualisation



N-Gram search

[Translate to Welsh](#)



You are here: [Home](#) > [UK Web Archive N-gram](#)

Provided by:



- Home
- About
- Search the archive
- Browse the archive
- Visualisation
- Nominate a site
- FAQ's
- Technical information
- Links to other archives
- Archive statistics
- Contact

Quick search

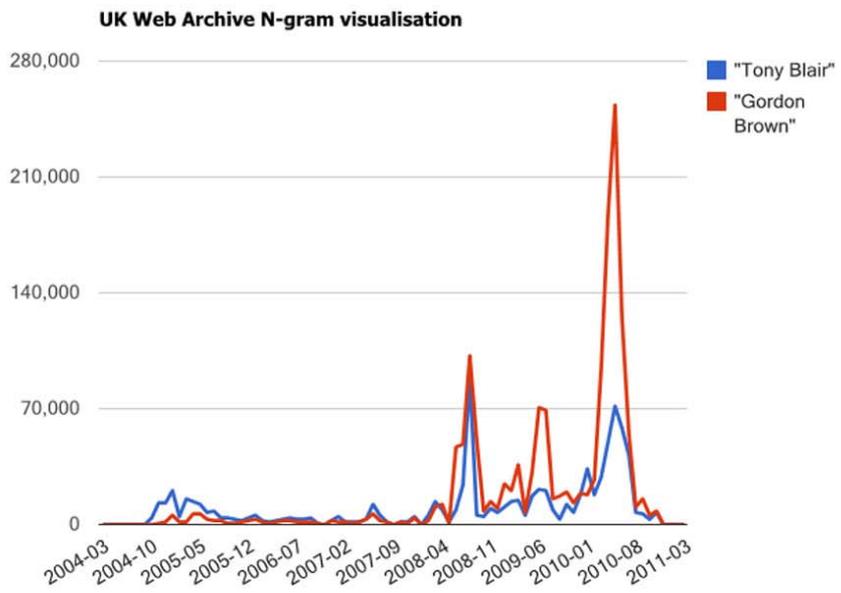
Please enter text

- Title (for a specific archived website)
- Full text (across all the archived websites)

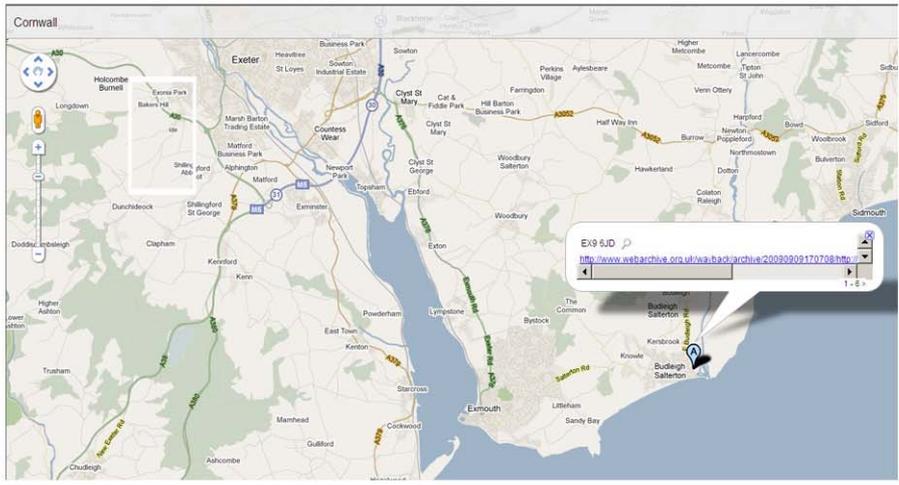
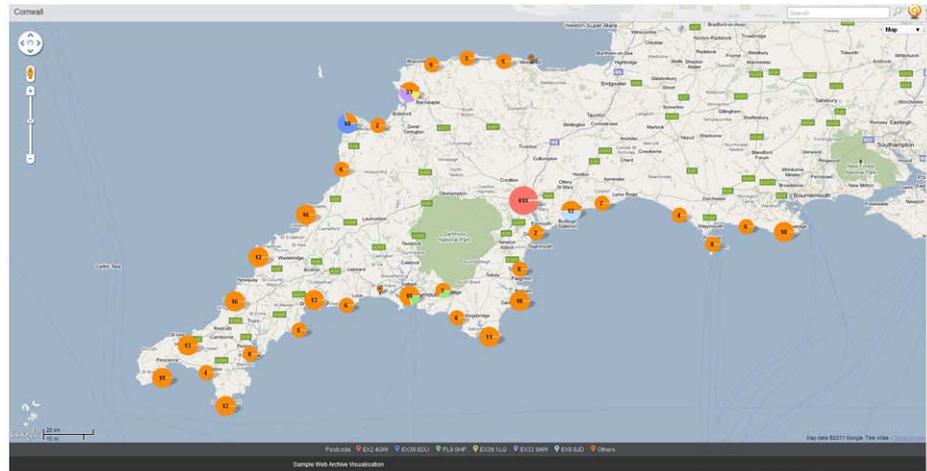
search

[Advanced search](#)

UK Web Archive N-gram



Postcode-based access



THE OFFICIAL GUIDE TO THE **South West Coast Path** NATIONAL TRAIL

HOME SEARCH GO THINGS TO DISCOVER WALKS FOR EVERYONE USEFUL INFORMATION

SHORT WALKS (up to one day)

WALK NAME: Salcombe to Bolt Head
START LOCATION: South Sands, Salcombe
FINISH LOCATION: South Sands, Salcombe
DISTANCE: 3 miles (4km)
GRADE: Moderate
TERRAIN: Coastal footpath uneven and rocky in places; surfaced road; field footpath.
CIRCULAR WALK: Yes
FREE FROM OBSTACLES & STEEP GRADIENTS: No
RECOMMENDED BY: [South Devon AONB Service](#)
WALK DESCRIPTION:

Passing through an enchanted landscape of rocky spires and jumbled pinnacles, with inspiring views in all directions, this is a coastal walk guaranteed to lift the spirits.

The South Devon AONB Service have produced a leaflet for this walk giving directions, along with information on the wildlife, geology and history of the area. It can be viewed and printed by clicking on the 'Printer friendly page' button on the right hand side, near the bottom of the page.

To find out more about the South Devon Area of Outstanding Natural Beauty and download other walks visit their website - <http://www.southdevonaonb.org.uk>

PUBLIC TRANSPORT INFORMATION:
 There are bus services to Salcombe. 606 from Kingsbridge; 92 from Plymouth and Kingsbridge. From bus stop follow the pedestrian signs to town centre and Whitestrand Pontoon.
 For details visit [Traveline](#) or phone 0870 6082608

NEAREST TOILETS:
 Public toilets at South Sands and at Whitestrand, Salcombe.

NEAREST CAR PARKS:
 Shadycombe Car Park, Salcombe (Postcode for Sat Nav: **TQ8 9ND**) for ferry to South Sands from Whitestrand Pontoon. Pay and display. Tides Reach car park, South Sands.

NEAREST REFRESHMENTS:
 Several pubs, cafes and restaurants in Salcombe. Tides Reach Hotel, South Sands serves refreshments and cream teas.

FURTHER INFORMATION:
[Salcombe Tourist Information Centre](#) or phone 01548 843927870

Salcombe Estuary

CLICK MAP TO ENLARGE

This map is reproduced from Ordnance Survey material with the permission of Ordnance Survey on behalf of the Controller of Her Majesty's Stationery Office © Crown copyright. Unauthorised reproduction infringes Crown copyright and may lead to prosecution or civil proceedings. Natural England Licence Number: 100046223. The Ordnance Survey mapping is included purely to provide a contextual backdrop for the walk and cannot be used for any other purpose.

Vision – UK Web Archiving in 2015

- The British Library will be the primary collector and provider of a web archive which is representative of the UK domain
- The Web Archive will be used for scholarly research in a range of disciplines
- Researchers will be able to search and use the web archive as a part of the Library's overall digital collections
- The UK Web Archive will be known as the place where researchers and general public look for inactive and or historical UK websites
- The Web Archive will be stored in the Digital Library System and preserved into the future

List of URLs

- The UK Web Archive:
<http://www.webarchive.org.uk/ukwa/>
- UK Web Archive data-based visual tools data
- Content Visualisation
<http://www.webarchive.org.uk/ukwa2/seadragon.html>
- N-gram search
<http://www.webarchive.org.uk/ukwa/ngram/>
- Demo Post-code based access (Google Maps replaced with OpenStreet Map)
<http://www.webarchive.org.uk/ukwa2/map.html>