Introduction to Web Archives

Presented by the web archiving team at the Bodleian Libraries
Objectives

- Gain an understanding of web archives, their importance and potential uses
- Become familiar with UK Web Archive and Bodleian Libraries Web Archive interfaces
- Opportunity to play around with both web archives and ask questions
<table>
<thead>
<tr>
<th>Session plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to web archives</td>
</tr>
<tr>
<td>UKWA Overview</td>
</tr>
<tr>
<td>Hands-on: UKWA</td>
</tr>
<tr>
<td>Break (15 minutes)</td>
</tr>
<tr>
<td>BLWA Overview</td>
</tr>
<tr>
<td>Hands-on: BLWA</td>
</tr>
<tr>
<td>Useful tools</td>
</tr>
<tr>
<td>Other web archives and useful resources</td>
</tr>
<tr>
<td>Questions and feedback</td>
</tr>
</tbody>
</table>
Introduction to Web Archives

• What is a 'Web Archive'?
  • A collection of archived websites.
  • Can be accessed through an access platform.
  • Websites can be arranged thematically or as a part of a bulk capture.
  • When the information on a web archive is downloaded they become datasets for researchers.
Introduction to Web Archives

• What is 'Web Archiving'?
  • The process of capturing and preserving websites.
  • Websites are crawled by a web crawler.
  • The crawler saves crawls in the WARC (Web ARChive) format.
  • The same websites are usually archived periodically (e.g. monthly, yearly)
The value of web archives / Why should we web archive?

• Legal / Evidential value:
  • To ensure government, organisational and public accountability
  • Proof that a website existed, websites constantly change and disappear

• Cultural value:
  • To collect valuable websites that form part of our cultural heritage
  • Provide research tools to current / future generations
Page not found (404)

Request Method: GET
Request URL: http://www.jammaplus.co.uk/forum/default.asp/

Using the URLconf defined in isomonitor.urls, Django tried these URL patterns, in this order:

1. ^accounts/login/$
2. ^accounts/logout/$
3. ^accounts/profile/$
4. ^admin/$
5. ^devices$
6. ^device/\(\?P<device_id>[\d-]+\)/setup$
7. ^device/\(\?P<device_id>[\d-]+\)/setup/name$
8. ^device/\(\?P<device_id>[\d-]+\)/setup/input/\(\?P<input_id>[\d-]+\)$
9. ^device/\(\?P<device_id>[\d-]+\)/monitor$
10. ^device/\(\?P<device_id>[\d-]+\)/monitor/cores$
11. ^locations$
12. ^location/add$
13. ^location/\(\?P<location_id>[\d-]+\)/mods$
14. ^location/\(\?P<location_id>[\d-]+\)/setup$
15. ^location/\(\?P<location_id>[\d-]+\)/setup/input/\(\?P<input_id>[\d-]+\)$
16. ^contact-groups$
17. ^contact-group/add$
18. ^contact-group/\(\?P<group_id>[\d-]+\)/setup$
19. ^contact-group/\(\?P<group_id>[\d-]+\)/setup/name$
20. ^contact-group/\(\?P<group_id>[\d-]+\)/del$
21. ^contact-group/\(\?P<group_id>[\d-]+\)/add_contact$
22. ^contact-group/\(\?P<group_id>[\d-]+\)/del_contact/\(\?P<contact_id>[\d-]+\)$
23. ^contact-group/\(\?P<group_id>[\d-]+\)/edit_contact/\(\?P<contact_id>[\d-]+\)$
24. ^tce8$
25. ^tce8Ncdacks$
26. ^doc\description$
27. ^about$
28. ^$ 

The current URL, forum/default.asp/, didn't match any of these.

You're seeing this error because you have DEBUG = True in your Django settings file. Change that to False, and Django will display a standard 404 page.
background and access

Legal Deposit Libraries Act (2003) NPLD, and further enabling regulations to include websites (2013)

Websites in the UK web archive are 'intended to be reflective of contemporary UK life or for their relevance to a particular collection. Judgement is not made on the validity or quality of website content.' UK Web Archive (2018)

UK top level domain without seeking permission – only accessible on dedicated legal deposit library terminals. [Unless the site is Open Access]

2017 data = 500TB

Is the UKWA being safeguarded for future use? - YES!
Curating a special collection
Demo: how to browse, search and navigate the UKWA

Point browser to https://www.webarchive.org.uk
Campaign against Climate Change, The

Archived date: 2015-07-04
https://twitter.com/CambsClimate

Campaign Brief (@CarbonBrief) on Twitter

Archived date: 2015-06-18
https://twitter.com/CarbonBrief/

Carbon Brief, The

Archived date: 2015-06-18
http://www.carbonbrief.org/

Carbon Trust, The

Archived date: 2015-07-04
http://www.carbontrust.com/

Catholic Agency For Overseas Development (CAFOD)
The Carbon Brief

Latest analysis

IEA: China should shoulder greatest burden in raising climate ambition
17 Jun 2015 16:30 | Simon Evans

The world is not on track to avoid dangerous climate change and China should take on more of the effort required to close the gap in climate ambition, according to the International Energy Agency in a special report. Read more

Tags: Energy policy, Climate policy

Yazikoba Arch Bridge | Wikipedia

Daily Briefing | Pope’s views on climate change add pressure to Catholic candidates
17 Jun 2015 09:00 | Carbon Brief Staff

World media focus on what impact the leaked papal encyclical will have on US politics. EU on track to achieve renewable energy goal but UK and Netherlands are lagging behind, and flying green energy leaves 'risk public backlash’. Read more

The Carbon Brief Interview: Dr Fatih Birol
16 Jun 2015 13:15 | Simon Evans

The Carbon Brief Interview with Dr Fatih Birol, chief economist of the International Energy Agency, and responsible for its World Energy Outlook publication. Read more

Tags: Climate science

Daily Briefing | Pope Francis warns of destruction of world’s ecosystem in leaked encyclical

The Carbon Brief

The Carbon Brief Interview: Dr Fatih Birol
16 Jun 2015 13:15 | Simon Evans

The Carbon Brief Interview with Dr Fatih Birol, chief economist of the International Energy Agency, and responsible for its World Energy Outlook publication. Read more

Tags: Climate science

In-depth: Is the 1.5C global warming goal politically possible?

The Carbon Brief

IEA #China should shoulder greatest burden in raising climate ambition
@IAE #DrSimEvens http://t.co/MMzgy8t1Wt

RT @IEA: Chief Economist @IEA on whether #357 declaration signals end of fossil fuel era http://t.co/DSQGwC7Q
<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>June</td>
<td>June 18th, 2015 at 05:30:35</td>
</tr>
<tr>
<td>2016</td>
<td>July</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>August</td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>September</td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>October</td>
<td></td>
</tr>
<tr>
<td></td>
<td>November</td>
<td></td>
</tr>
<tr>
<td></td>
<td>December</td>
<td></td>
</tr>
</tbody>
</table>
Non-Print (Electronic) Legal Deposit Access Policy

The Bodleian Libraries provide access to this item on certain conditions

We expect everyone to read and accept the following guidelines on appropriate use of Non-Print Legal Deposit material before they begin to view material using this system. All usage will be monitored. Infringement of these guidelines may lead to suspension of your reader pass and legal proceedings by the rights owner.

Permitted use according to the Regulations

This material has been supplied to the Bodleian Libraries under the Non-Print Legal Deposit Regulations and as such its use is limited under these regulations.

a) Digital copies of these items are not permitted, but printing is permitted in accordance with Copyright Regulations. Details of the Copyright Regulations are on posters around the Reading Room.

b) Text and data mining is not permitted on this content.

c) For non-commercial or private study of born-digital sheet music, fair dealing rules apply. The Music Publishers Association (MPA) guidelines advise that anything copied should be less than a performable piece to be used for study, and should not be intended or used for performance. A licence from the MPA will be needed for copying for performance.

d) Non-Print Legal Deposit items can only be accessed by a single user at any one time therefore as soon as you have finished with an item please close the browser to allow another reader to access the content.

Full details can be found on the Legal Deposit web pages.

If you intend to print, agreeing to the terms and conditions counts as an electronic signature to the following.

I declare that:

1) I will not use the copy except for the purposes of research for non-commercial purpose, private study, criticism or review or reporting current events, parliamentary or judicial proceedings, or a Royal Commission or statutory inquiry.

2) In relation to a copy of relevant material required for the purposes of non-commercial research or private study:

   a. I have not previously been supplied with a copy of the same material by you or another deposit library: and
   b. To the best of my knowledge, no person with whom I work or study has made or intends to make, at or about the same time as this request, a request for substantially the same material for substantially the same purpose.

I understand that, if the declaration above is false in a material particular, the printed copy supplied will be an infringing copy and that I shall be liable for infringement of the Copyright.

I agree to the terms and conditions of this policy

Please close the browser tab when you have finished viewing this item to allow others access.
Save a UK website

Fields marked with * are mandatory

Personal details you provide on this form are protected by UK Data Protection Law. Please view our Privacy Statement.

Name
Please enter your name

* Email address [*]
Please enter your email address

* Title of website
Please enter the title of website

* URL of the website
Please enter URL of website

Any other information
Use this space to tell us for example, if you own this website, or if it is about to go offline

We need your email address if you require a response from us. It would also be helpful if you are the owner of the nominated website and would like us to make archival copies of the website available through the UK Web Archive.

I'm not a robot

Submit
UK Web Archive: Worksheet

• Work your way through the UKWA worksheet, or if you prefer just browse and play around on the site.
• If you need any help please raise your hand.
Bodleian Libraries' Web Archive

- Started in 2011
- Hosted by Archive-It
- Contains over 12.8 TB of data and 167,416,167 files
- Focuses on capturing sites related to Oxford and current Bodleian Special Collections.
BLWA Collecting areas:

- Arts and Humanities
- OU Museums, Libraries and Archives
- Social Sciences
- International
- University of Oxford
- Oxford Colleges
- Science, Medicine and Technology
- UO Student societies
Demo: how to search and navigate the BLWA

Point browser to
https://archive-it.org/home/bodleian
Narrow Your Results

Sites and collections from this organization are listed below. Narrow your results at left, or enter a search query below to find a collection, site, specific URL or to search the text of archived webpages.

Enter search terms here

Collections Sites Search Page Text

Page 1 of 1 (7 Total Results)

Sort By: Collection Name (A-Z) | Collection Name (Z-A)

Arts and humanities
Archived since: Mar, 2011
Description: Web sites relating to arts and humanities.
Subject: Arts & Humanities

International
Archived since: Mar, 2011
Description: Web sites from around the world.
Subject: Arts & Humanities, Politics & Elections, Society & Culture
Title: Felix Dennis
URL: http://www.felixdennis.com/
Collection: Arts and humanities
Description: Works of Felix Dennis: publisher, poet and planter of trees
Captured 47 times between Sep 17, 2011 and Aug 10, 2019
Videos: 1459 Videos Captured
Subject: Charity, Literature
Creator: Felix Dennis Trust
Publisher: Felix Dennis Trust
Language: English
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9 pages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 pages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 pages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 pages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* denotes when page was updated
FELIX DENNIS
1947 - 2014

Poet, Publisher, Planter of Trees.

Felix was a publishing legend, famed for his maverick, entrepreneurial style and, more lately, a successful and much loved poet.

He dedicated his legacy to The Heart of England Forest, which will, one day, be the largest deciduous, broadleaf woodland habitat in England.

READ MORE
Potential issues on BLWA

Content not captured

Dynamic content

Videos: Youtube / Vimeo

The page you requested has not been archived in Archive-It.

This could be for a number of reasons.

Most likely the page you are requesting was outside of the crawler's scope. Try another request or click to see other pages from https.
Bodleian Libraries' Web Archive: Hands-on

• Work your way through the BLWA worksheet or play around with the site
• If you need any help please raise your hand.
Extra tools and resources

Time Travel – check if other web archives have captured a site - http://timetravel.mementoweb.org

**time travel**

Find Mementos in Internet Archive, Archive-It, British Library, archive today, GitHub and many more!
Demo: Voyant Tools

Point browser to
https://voyant-tools.org
Example site: https://wayback.archive-it.org/2501/20190901163311/http://www.opcw.org/
Can also be used to compare two sites like an article from *The Guardian* and *the BBC* on Brexit, both written on 22 October 2019
SHINE interface: Trends and searching

https://www.webarchive.org.uk/shine

[Further information about the dataset, including how to cite: http://data.webarchive.org.uk/opendata/ukwa.ds.2/#issues]
Other Web Archives

Government archive
http://www.nationalarchives.gov.uk/webarchive

Internet archive
https://archive.org/

Parliamentary archive
http://webarchive.parliament.uk

Other institutional archives
https://archive-it.org/explore?show=Organizations
Further information on web archives:

- Web Archive LibGuide
  - More information about BLWA and UKWA
  - Other examples of web archives
  - Web archive use cases
  - Other web archiving tools

URL: https://libguides.bodleian.ox.ac.uk/web-archives
Thank you for your attention

Questions?