Resource Not In Archive: Understanding the behaviour, borders and boundaries of the UK Web Archive.

Nicola Bingham
Lead Curator, Web Archiving

14th June 2017
UK Web Archive Overview

Three Collections:

- Open Archive (since 2004)
- Legal Deposit Archive (since 2013)
- JISC Historical Archive (1996-2013)
Size of the UK Web Archive

400 TB
11-12 billion records

wow!
We collect UK websites only

- UK TLD (.uk, .scot, .cymru etc)
- UK hosting: geo-IP look up to locate server
- Manual checks
  - Postal address
  - Correspondence
  - Professional judgement
We only collect published content

Out of Scope:

- Film and recorded sound where AV content predominates, e.g. YouTube.
- Private intranets and emails.
- Social networking sites only available to restricted groups.
Web archiving is expensive

- Annual domain crawls
- 500MB cap per domain
- Selected frequent crawls
There are (many) technical limitations

• Database driven sites
• Programming Scripts
• Plug-Ins
• Proprietary file formats
• Blockers
  – Robots.txt
  – Access denied
There are misrepresentations
Curatorial input

• c. 200 curators
• Across subject areas
• Ensures curation of “important stuff”
• Aids discovery
**Collection Scoping Document**

- **Captures scope, motivation, parameters & timeframe of collection**

---

**UK Web Archive:**
**Request for a Special Collection**

**[Title of collection]**

<table>
<thead>
<tr>
<th>Date of request: dd/mm/yyyy</th>
<th>Requested by: ........................</th>
</tr>
</thead>
</table>

**Special collections are an opportunity for nominated owners to establish collections of archived websites around a given theme of subject.**

Collections must be maintained and quality nominations made on a monthly basis.

Collections should not exceed 400 websites.

Selectors are required to use the online Selection & Permission Tool

Selectors must read the Selection guidelines

<table>
<thead>
<tr>
<th>Type of collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FORMCHECKBOX</strong> Curators own collection</td>
<td></td>
</tr>
<tr>
<td><strong>FORMCHECKBOX</strong> Current event collection</td>
<td></td>
</tr>
<tr>
<td><strong>FORMCHECKBOX</strong> Themed collection</td>
<td></td>
</tr>
<tr>
<td><strong>FORMCHECKBOX</strong> Exhibition collection</td>
<td></td>
</tr>
<tr>
<td><strong>FORMCHECKBOX</strong> Guest collection</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Curator and organisation responsible for selection</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Collection Overview: A brief description of the collection, its relevance to the UKWA selection guidelines, and why it should be considered for a special collection.</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Who will use the collection?</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>How will the collection be promoted?</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>How will permission to archive be secured?</th>
<th></th>
</tr>
</thead>
</table>

- **FORMCHECKBOX** License form signed by website owner
- **FORMCHECKBOX** Crown Copyright content
- **FORMCHECKBOX** Creative Commons content
- **FORMCHECKBOX** Exceptional Archive & Notice permission
- **FORMCHECKBOX** Orphaned works permission
We will expose collection metadata

Explore the special collections

Special Collections are groups of websites brought together on a particular theme by librarians, curators and other specialists, often working in collaboration with key organisations in the field.

British Stand-up Comedy Archive
Collection owned and administered by Elspeth Millar, Briti...

French in London
This collection of websites has been selected by Saskia H...

Scottish Independence Referendum
Provisional title for collection on Scottish Devolution, ...

News Sites
500 titles are included in this collection (30/01/2014), ...
Description (A description of this website)

Each week, a profile of a Londoner is posted on the home page. The profile contains a 3 minute film that gives an insight into the life of the Londoner, as well as their own personal photos of London and some answers to crucial questions about their views on London life.

Browse Subject

Arts & Humanities, Digital Society, Society & Culture, Communities, News and Contemporary Events, Local History

Collections (Special collections that you think this website should appear in)

News Sites, Oral History in the UK, Web only news sites

Nominating Organisation (The organisation that nominated this website)

The British Library

Originating Organisation (the organisation(s) responsible for creating the content)

Nominated by 1000 Londoners

Selector *(The user that nominated this website)

Helena Byrne

Language (the predominant language in which the page is written)

English
Information gathered during the crawling process
What do we know about what researchers want to know?
Questions?

nicola.bingham@bl.uk
@NicolaJBingham

www.webarchive.org.uk
@UKWebArchive