British Library
Access Policies: Challenges and Approaches

Nicola Bingham, Lead Curator Web Archives
IIPC Conference, Zagreb, June 2019

nicola.bingham@bl.uk
https://www.webarchive.org.uk/
@NicolaJBingham
UK Web Archiving Overview

- Web archiving in the UK since 2004: small scale, selective, permission-cleared.
- UK Legal Deposit Libraries & partners, e.g. The National Archives, JISC.
- 2004 – 2013: 15,000 websites archived.
- 2013 Non-Print Legal Deposit Regulations. LDLs archive the UK Web at scale.

UK Web Archiving Consortium 2004

UK Web Archive 2008

UK Web Archive 2018
UKWA Collection Strategies

• **Annual domain crawl of the UK Web Space**
  – 5-10 million hosts (websites) per annum
  – Over 2 billion items per annum
  – 70 - 100 TB of compressed data per annum
  – Total size: 517 TB compressed data

• **Curated Collections**
  – LDL staff and external partners
  – c. 150 curated collections
  – c. 87,000 curated targets
  – Events e.g. General Elections, ‘Brexit’, Olympic Games
  – Researcher-led e.g. Diaspora Communities in the UK, Muslims, Trust & Cultural Dialogue Project
  – External partner collections e.g. Jersey Archives, British Stand-up Comedy Archive, Wimbledon Lawn Tennis Museum

• **Other content strands & formats** e.g. News, Official publications, Digital Sheet Music, Interactive Fiction

All managed through the Annotation, Curation Tool
UK Non-Print Legal Deposit Regulations 2013

• Enable the six Legal Deposit Libraries to archive the “UK Web” at scale
• UK LDLs: British Library, National Library of Scotland, National Library of Wales, Cambridge University Libraries, Bodleian Libraries, Trinity College Dublin

LD Access Challenges

• Access restricted to reading rooms of the UK LDLs
• Single concurrent access
• Archive accessed inside a Virtual Machine
• Digital copying not permitted
• Cannot cite archived URLs
• Cannot view embedded links or source code
  = Poor user experience
UKWA User Interface Search

- Search c. 87,000 curated records though BL Explore, the main Library catalogue [https://www.explore.bl.uk](https://www.explore.bl.uk)

- Search the UK Web Archive User Interface

- Search across Legal Deposit (reading room only) and openly available websites

- **Full-text search index**

- Search by **URL or keyword**

- Narrow search by **facets** or **collection**
Internet Archive/JISC UK Web Domain Dataset

- Internet Archive crawl of .uk websites 1996 – 2013
- Research access only
- Shine trends analysis
- Secondary datasets available openly e.g.,
  - Format Profile
  - Geoindex
  - Host Link Graph
  - Crawled URL Index

https://www.webarchive.org.uk/shine
https://data.webarchive.org.uk/opendata/
https://github.com/ukwa/opendata
Thank-you!

nicola.bingham@bl.uk
https://www.webarchive.org.uk/
@NicolaJBingham
@UKWebArchive