Development of a UK sports collection for the UK Legal Deposit Web Archive

12th April 2018
The UK Legal Deposit Libraries
What are we collecting?

All of the public UK web

• The domain crawl scopes in websites using the following criteria:
  • URL has a UK top level domain: e.g. .uk, .scot, .cymru, .wales, .london
  • Registered in WHO-IS as a UK site
  • Site is stored on a UK server

We also identify web content that may otherwise be missed from the domain crawl.
Legal Deposit and special collections

Total number of websites is taken from figures for the 2016 whole domain crawl, and is based on number of hosts. The 2017 domain crawl is still being analysed.

9 million websites*
54,461 curated websites
2,274 open access websites
Does that mean we are collecting “everything”? 
Does that mean we are collecting “everything”?
Does that mean we are collecting “everything”?

We don’t collect:

• Anything requiring credentials or login
• Intranets
• Databases
• Material stored in the cloud
• Flash
• Most video or audio content
• Facebook
• Not much Twitter
Where to find the collections

**UK Web Archive website**
- All archived UK websites since 2005
- [webarchive.org.uk/ukwa](http://webarchive.org.uk/ukwa)

**Legal Deposit**
- Domain & frequently crawled UK websites with no OA permission
- UK LD library reading rooms

**Shine Interface**
- All archived .uk websites from 1996-2013
- [webarchive.org.uk/shine](http://webarchive.org.uk/shine)
Special collections: thematic, event based and rapid response

- Thematic (71), 58%
- Event based (42), 34%
- Rapid Response (10), 8%
The UK Web Archive (UKWA) collects millions of websites each year and preserves them for future generations.


About us

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.

View all special collections
Creating a dedicated sport collection

- It is a subject which has not traditionally been well served by libraries and museums

- Selection: the National Library of Scotland nor British Library have Curators completely dedicated to collecting Sport
Creating a dedicated sport collection

Lead Curators:
• Helena Byrne, British Library
• Trevor Thomson, National Library of Scotland
• Additional Selectors:
  • Web archiving staff and other curators from the UK Legal Deposit Libraries
  • Some nominations from the public
Structuring Sport

- Sports Collection (850)
  - Athletics (24)
  - Ball Sports (excluding Football) (109)
    - By Hand (10)
    - On a Table (0)
    - With a Stick (33)
    - With Ball (10)
    - With Club (7)
    - With Racket (Racquet) (4)
  - Combat/Self Defence Sports (10)
  - eSports (1)
  - Fishing, Hunting, Shooting and Target Sports (20)
  - Gymnastics/Acrobatics (39)
  - Horse and Other Animal Sports (10)
  - Motor Sports (13)
  - News/Media (44)
  - Orientation Sports (25)
  - Stadiums (69)
  - Strength Sports (5)
  - Target Games (9)
  - Water Sports/Aerial Sports (105)
  - Wheel Sports (135)
  - Winter Sports (25)

- Sport: Football (1278)
  - Gaelic Football (72)
  - Other (1)
  - Rugby (293)
    - Rugby League (56)
    - England (2)
    - Northern Ireland (0)
    - Scotland (5)
    - Wales (49)
  - Rugby Union (207)
  - Soccer (908)
    - Clubs (606)
      - England (190)
      - Northern Ireland (11)
      - Scotland (250)
      - Wales (152)
      - Commercial (1)
      - Governing Bodies (145)
    - Related Sites (170)
      - Fans (83)
      - Forums (2)
      - History and Research (43)
      - News/Media (38)
Challenges

• Selection

• We need input from the public, but also academic community

• Sport is international…Websites are international… but Legal Deposit is not!

• Collection maintenance
Top level domain suffices attached to 832 “sport and recreation” sites
What we have saved
Adapted from: https://www.flickr.com/photos/barron/34073939122/ under https://creativecommons.org/licenses/by-nc/2.0/#
Where can I find out more?

- UK Web Archive
  
  https://www.webarchive.org.uk/ukwa/

- UK Web Archive Open Data
  
  data.webarchive.org.uk/opendata

- The IIPC Awesome list; a good general guide to open source tools and training material:
  
  • https://github.com/iipc/awesome-web-archiving/
“History books remember the big players, Tweets remember the individuals”

Chido Muchemwa
THANK YOU

Eilidh MacGlone, Web Archivist

e.macglone@nls.uk
https://www.nls.uk/

@dalmailing

Except for images or where otherwise stated this presentation is © National Library of Scotland and is licensed under the Creative Commons Attribution 4.0 International Licence. To view a copy of this license, visit: http://creativecommons.org/licenses/by/4.0/