

# Resource Not In Archive:

Understanding the behaviour,  
borders and boundaries of the UK  
Web Archive.

Nicola Bingham

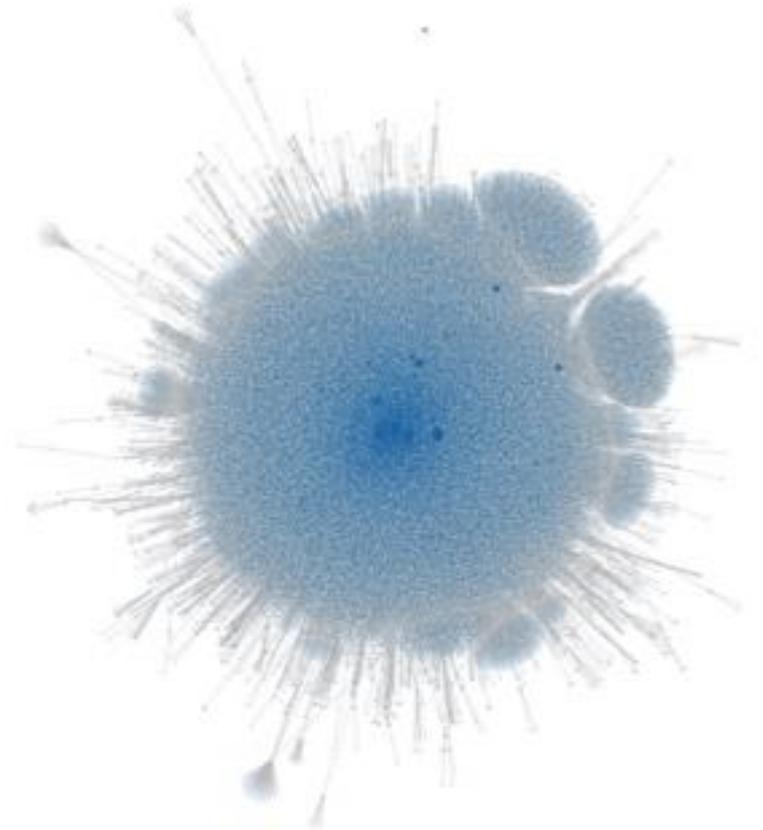
Lead Curator, Web Archiving

14<sup>th</sup> 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



# 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.



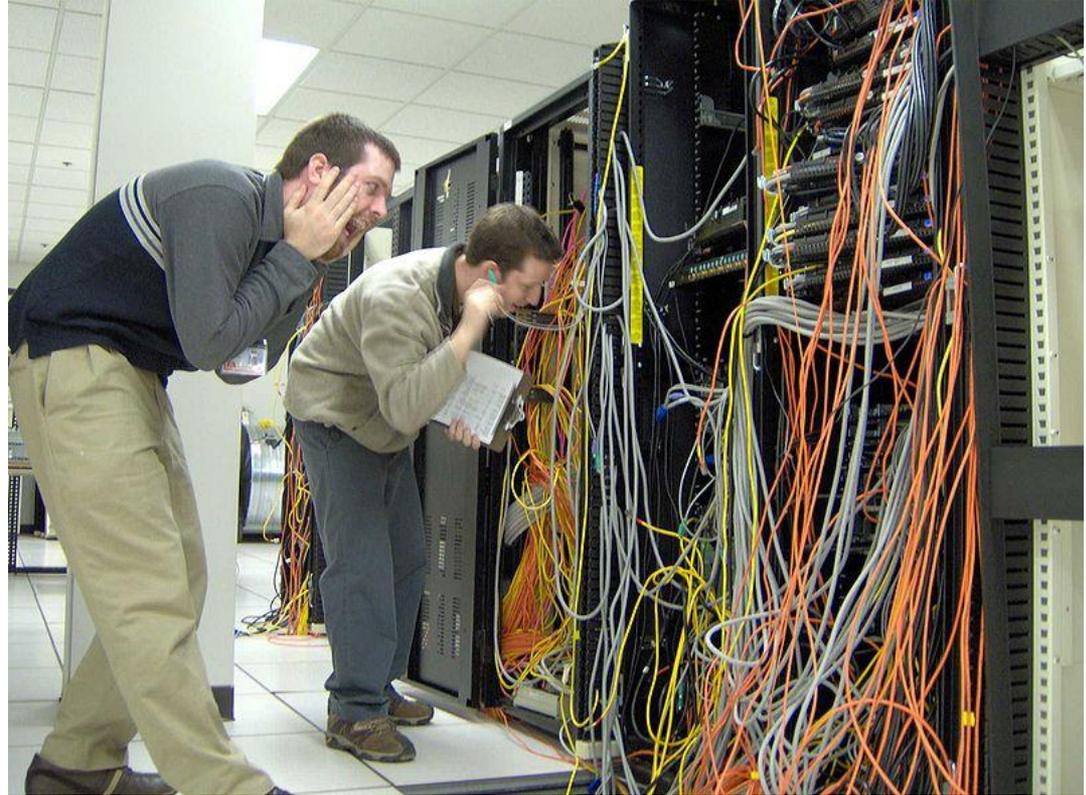
# Web archiving is expensive



- 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

My Sites Reader

ONE WORLD ACTION'S BLOG: REFLECTIONS ON WOMEN'S RIGHTS AROUND THE WORLD

Search... SEARCH

## ONE HUNDRED WOMEN

ABOUT 100 WOMEN: THE UNSEEN POWERFUL WOMEN WHO CHANGE THE WORLD CONTACT

**31 OCT**

### Causes to Celebrate – 100 Unseen Powerful Women Awards and Farewell to One World Action

– Laura Ouseley, Campaigns Coordinator, One World Action

The One Hundred Unseen Powerful Women campaign by [One World Action](#) culminated on 27th October with an awards ceremony at the King's Fund, London. We were joined by over 200 supporters, campaigners, colleagues, and many of the women from the [One Hundred Unseen Powerful Women list](#). The original



ONE HUNDRED WOMEN  
THE UNSEEN POWERFUL WOMEN WHO CHANGE THE WORLD

**RECENT POSTS**

- Causes to Celebrate – 100 Unseen Powerful Women Awards and Farewell to One World Action
- Powerful Women Taking Control
- Standing up for Blind Women's Rights in El Salvador
- The power of older women
- 100 Women List Now Published!

**TWITTER**

Thanks @funnywomen for this really lovely article about One World Action's 100 Unseen Powerful Women event last week bit.ly/uao3m7 4 years ago

#FF orgs making the world better for women, men & children @woman\_kind @UK\_Feminista @OrchidProject @VitalVoices @GenderandDevNet #goodbyeOWA 4 years ago

External links, forms and search boxes may not function within this archived website.

<http://oneworldaction.wordpress.com/> Go

29 Nov 11 - 27 Nov 12

3 captures

OCT NOV MAY 29

2010 2011 2012

Close X

Summary

Help ?

One World Action's Blog: Reflections on women's rights around the world

Search

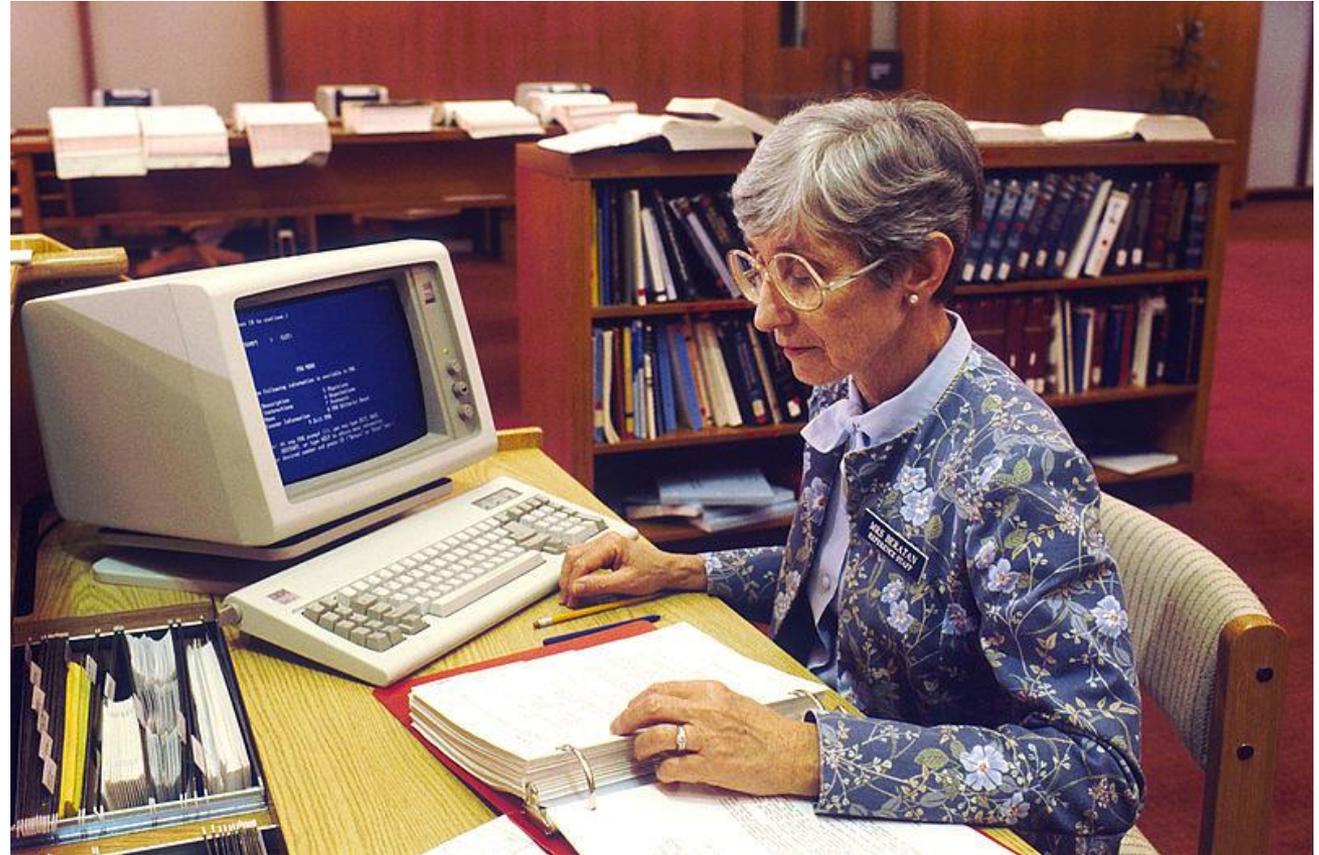
Search... Search

### One Hundred Women

- [About](#)
- [100 Women: The Unseen Powerful Women Who Change The World](#)
  - [Public Service](#)
    - [Nihma Cande](#)
    - [Leila Bashir](#)
    - [Runa Naila](#)
    - [Sister Josephine and the staff of St Anne's](#)
    - [Liz Douglas](#)
    - [Katie Hagley](#)
    - [Firoza](#)
    - [Katie Rizvi](#)
    - [Iman Abou Atta](#)
    - [Serrin M. Foster](#)
    - [Dr Caroline Bressey](#)
    - [Abha Shrestha](#)
    - [Dr Manda Mune](#)
    - [Esther Ngumbi](#)
    - [Marceline Kongolo-Bice](#)
    - [Jennifer A.C. Albano](#)
    - [Hariyatu Bangura](#)
    - [Nan Sloane](#)
    - [Lyn Lusi](#)
    - [Aarti Naik](#)
    - [Rahima Khanun](#)
    - [Anvys Darkwa](#)
    - [Dr Hawa Abdi](#)
    - [Ana Hurtado Abad](#)
    - [Dr. Maslumi Rama Singh](#)
    - [Eshana Faruqi-Stecker](#)
    - [Dr. Beatriz Acevedo](#)
    - [Ena Abrahams](#)

# Curatorial input

- c. 200 curators
- Across subject areas
- Ensures curation of “important stuff”
- Aids discovery



[title of collection]

Date of request: dd/mm/yyyy

Requested by: .....

- Collection Scoping Document
- Captures scope, motivation, parameters & timeframe of collection

*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*

Type of collection	<input type="checkbox"/> Curators own collection <input type="checkbox"/> Current event collection <input type="checkbox"/> Themed collection <input type="checkbox"/> Exhibition collection <input type="checkbox"/> Guest collection
Curator and organisation responsible for selection	
Collection Overview <i>A brief description of the collection, its relevance to the UKWA selection guidelines, and why it should be considered for a special collection.</i>	
Who will use the collection?	
How will the collection be promoted?	
How will permission to archive be secured?	<input type="checkbox"/> License form signed by website owner <input type="checkbox"/> Crown Copyright content <input type="checkbox"/> Creative Commons content <input type="checkbox"/> Exceptional Archive & Notice permission <input type="checkbox"/> 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](#)



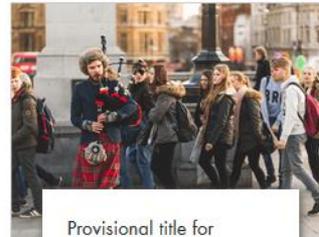
Collection owned and administered by Elspeth Millar, Briti...

### [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](#)



558 titles are included in this collection (30/01 /2014). ...

[View more collections...](#)

View Edit

- Overview **Metadata** Crawl policy and Schedule NPLD scope (+) UKWA Licencing (+)

Watched Target

## Description (A description of this web site)

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 web site should appear in)

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

## Nominating Organisation (The organisation that nominated this web site)

The British Library

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

## Selector \* (The user that nominated this web site)

Helena Byrne

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

English

## Instances

<http://www.1000londoners.com/>

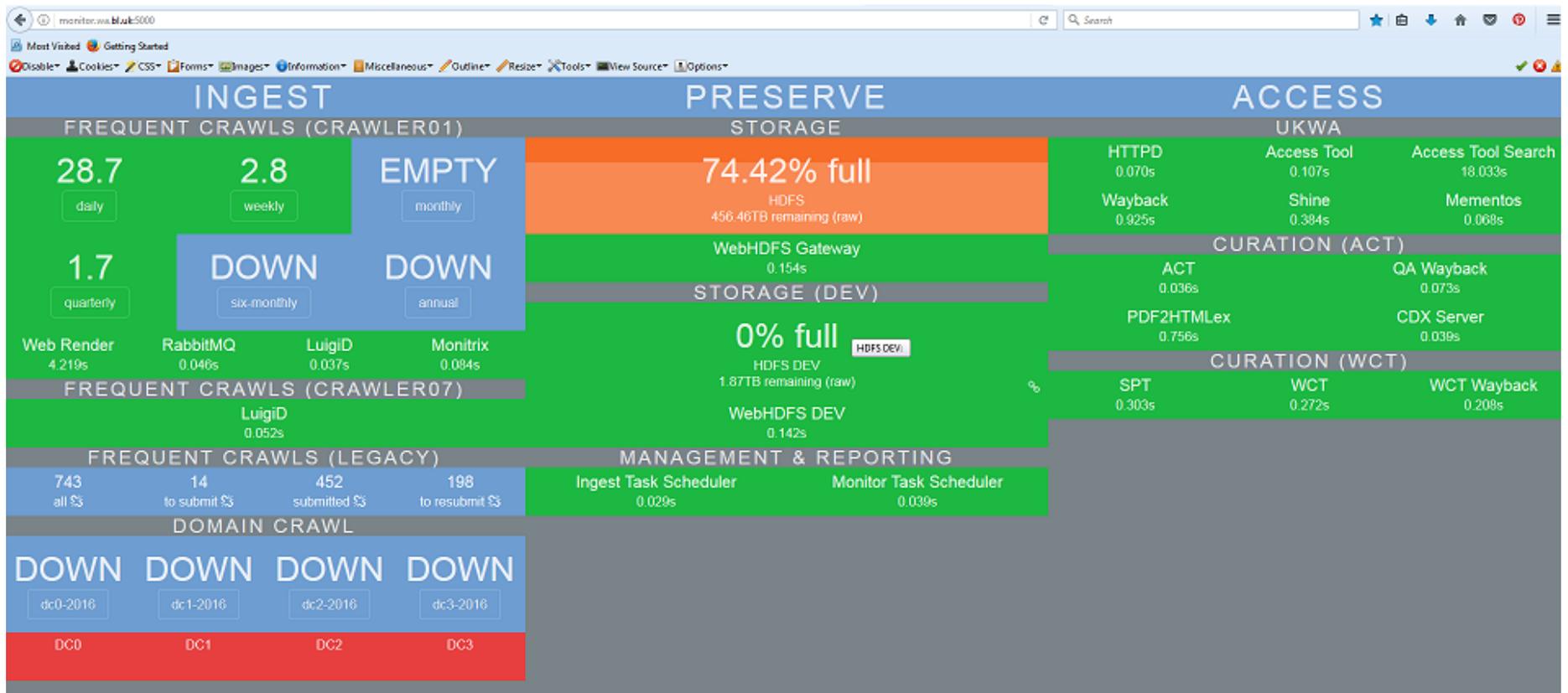
- Search QA Wayback
- Search Open UK Web Archive
- Search Other Web Archives

## Crawl Logs

<http://www.1000londoners.com/>

- Find URL in crawl logs
- Find all URLs with this prefix in the crawl logs

# Information gathered during the crawling process



# What do we know about what researchers want to know?



# Questions?

[nicola.bingham@bl.uk](mailto:nicola.bingham@bl.uk)  
@NicolaJBingham

[www.webarchive.org.uk](http://www.webarchive.org.uk)  
@UKWebArchive