

Why do we do it? Understanding the Research Value of Web Archives

Pamela Graham

Director, Humanities & Global Studies

Director, Center for Human Rights Documentation & Research

Columbia University Libraries

New York Technical Services Librarians

Spring 2018 Program

May 2, 2018 The New York Public Library

-
- ▶ **Challenges of user access to web archives**
 - ▶ **Diversity of Web Archiving purposes and relevance to use**
 - ▶ **Researcher initiatives**
-
- 

Challenges for Users of Web Archives

- ▶ Discovery Curve
- ▶ Cultural Understanding of the Web
- ▶ Functional limitations of Web Archives
- ▶ Lack of scholarly standards and practices



Web as	Content	Use	Discovery
Repository	Discrete objects; stand alone value; analogous to pre- web forms;	close reading/ viewing and analysis; used in tandem with non-web	Item / series-level cataloging and description; connection to related non- web items
Platform / field site	Websites, “web spheres”; Context is important	Analysis of language use, networking and linking patterns, “utilitarian” functions	Collection level / website level description, web search as research, data- mining, analytical and visualization tools
Technology	Code, CSS, scripts, design elements	History of the web; programming; experience of the web	Help!



Big UK Domain Data for the Arts and Humanities

Search



[About](#) [Blog](#) [Contact](#) [Documentation](#) [News and events](#) [Bursaries](#)

Web archives are an increasingly important resource for arts and humanities researchers, yet we have neither the expertise nor the tools to use them effectively. Both the data itself and the process of collection are poorly understood, and it is possible only to draw the broadest of conclusions from current methods of analysis. The project will work with the dataset derived from the UK domain crawl from 1996 to 2013 (that is, when legal deposit legislation was extended to cover digital materials), totalling approximately 65 terabytes and constituting many billions of words. For the arts and humanities, this is very big data indeed.

A key objective of the project will be to develop a theoretical and methodological framework within which to study this data, which will be applicable to the much larger on-going UK domain crawl, as well as in other national contexts. Researchers will work with developers at the British Library to co-produce tools which will support their requirements, testing different methods and approaches.

A major study of the history of UK web space from 1996 to 2013, including language, file formats, the development of multimedia content, shifts in power and access, and so on,

RECENT POSTS

- [More project case studies available](#)
- [IHR workshop on web archiving](#)
- [Project case studies now available](#)
- [Web archives as big data](#)
- [Big Data in the Humanities: lessons from papyrus and Instagram](#)

ARCHIVES

- [April 2016](#)
- [November 2015](#)
- [July 2015](#)
- [January 2015](#)
- [October 2014](#)
- [July 2014](#)
- [June 2014](#)
- [May 2014](#)

BUDDAH

- Better methods, frameworks for web archive research
- Build skills and awareness of potential value
- Inform collection development
- Demonstrate overall research value of web archives

<https://buddah.projects.history.ac.uk/>

- *Online reactions to institutional crises: BBC Online and the aftermath of Jimmy Savile* by Rowan Aust. Rowan examines the way the BBC has responded to the scandal, given that Savile is present in so many archived programmes and articles.
- *'all writing is in fact cut ups': the UK Web Archive and Beat literature* by Rona Cran. Rona's study is on the reception of Beat literature in the UK, in both academia and among the general public, at the end of the 20th century and the beginning of the 21st.
- *Revealing British Euroscepticism in the UK Web Domain and Archive Case Study* by Richard Deswarte. Richard reflects on studying his topic via web archives, with the advantages and pitfalls of this methodology.
- *Digital barriers and the accessible web: disabled people, information and the internet*, by Gareth Millward. Gareth explains how his plans for using the web archive to study what was available to the disabled had to be changed because of the particular challenges of using web archives for research.
- *Do online networks exist for the poetry community?* by Helen Taylor. Helen used the UK web archive to investigate poetry communities and how they interact with each other, with a particular focus on [The Poetry Forum](#) and the online presence of the Oxford University Poetry Society.

BUDDAH

PROJECT REPORTS



RESAW

A Research infrastructure for the Study of Archived Web materials



[Home](#)

[About RESAW](#)

[Events](#)

[Projects](#)

[Participants](#)

[Web Archives](#)

[Resources](#)

[Forum](#)

[Contact](#)



Home

Tweets from https://twitter.com/resaw_eu/lists/resaw

RESAW, a Research Infrastructure for the Study of Archived Web Materials, is a network established in 2012, aiming at promoting a collaborative European research infrastructure for the study of archived web materials.

Stories and News

Peter Webster, Web Historian, UK

[ng religion](#)

[Bernstein i](#)

Ian Milligan, Web Historian, Canada

[le: "Ten Simple Rules for Collaborative Web Arch](#)
[son Development"](#) [shop](#)

Web Archives for Historians

[iplinary Event Documentation: "TheFrom Da](#)
[IRC Adventure"](#) [on our D](#)

RESAW

With a view to promoting the establishing of a collaborative European research infrastructure for the study of archived web materials the RESAW network — A Research Infrastructure for the Study of Archived Web Materials — was established in late 2012.

Administration

[Log in](#)

[Entries](#) [RSS](#)

[Comments](#) [RSS](#)

[WordPress.org](#)



AUT PROJECT

Our goal is to make petabytes of historical internet content accessible to scholars and others interested in researching the recent past.





AU 4.0 The British Invasion, June 2014



Home

Welcome

Contact Us

About the Project

Archives Unleashed Toolkit

Archives Unleashed Cloud

Warlight

Get Involved

Events

Toronto Datathon

News (on Medium.com)

Get in touch

@unleasharchives on Twitter

@archivesunleashed on GitHub

Contact via email

Welcome to the Archives Unleashed Project



Welcome

Archives Unleashed aims to make petabytes of historical internet content accessible to scholars and others interested in researching the recent past. Supported by a grant from the [Andrew W. Mellon Foundation](#), we will be developing web archive search and data analysis tools to enable scholars and librarians to access, share, and investigate recent history since the early days of the World Wide Web.

Interested in the project? [Subscribe to our mailing list!](#) Or you can follow the links at left for information [about the project](#), the [Archives Unleashed Cloud](#), [Archives Unleashed Toolkit](#), or [our events](#).

We're always looking for [ways to engage](#) archivists, librarians, researchers, developers, or any others interested in born-digital heritage!

Contact Us

Questions? Comments? Please contact us, either by leaving an issue on one of our [GitHub projects](#) or by [sending us an e-mail](#). Are you a [Slack user](#)? [Join our Slack team!](#)

CC-BY 2018 | Site built with Hugo using the [Material](#) theme

<http://archivesunleashed.org/>

