

Why Archive the Web?



This training session was developed in partnership by the International Internet Preservation Consortium (IIPC) and the Digital Preservation Coalition (DPC)



1

What is Web Archiving?

Web archiving is the process of **collecting portions** of the World Wide Web, **preserving** the collections in an archival format, and then serving the archives for **access and use**.

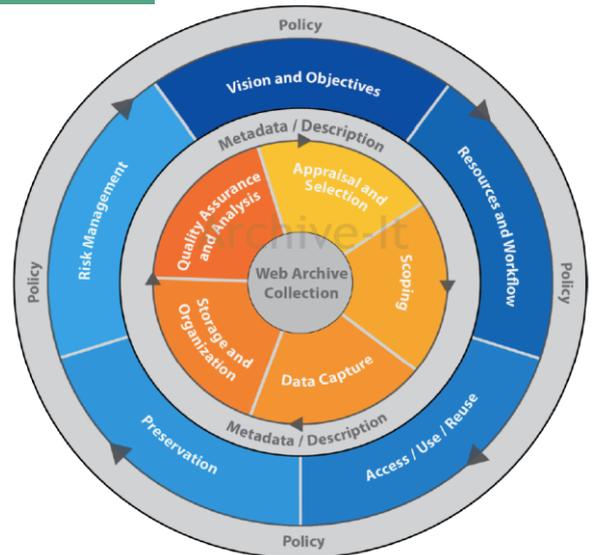
- IIPC

<http://netpreserve.org/web-archiving/>

2

Internet Archive's Web Archiving Life-Cycle Model

- Every step and action taken governed by a **policy** decision
- **Planning and management** keep the cycle moving and support continuous improvement
- **Metadata** and descriptive information about web content captured at multiple stages for different purposes
- Collections generated through on-going process of **selecting, capturing, storing, and quality assuring content**



http://ait.blog.archive.org/files/2014/04/archiveit_life_cycle_model.pdf

3

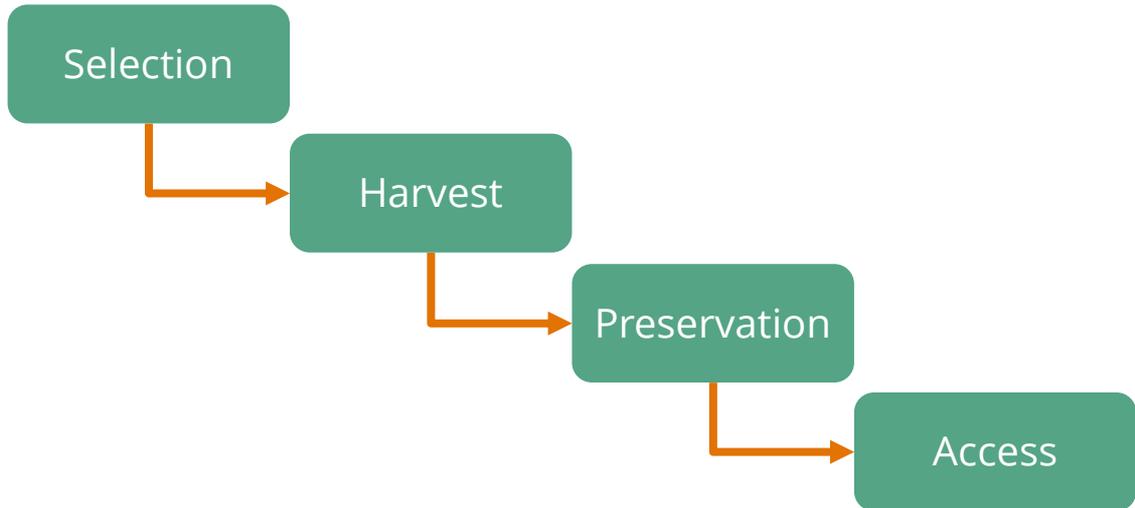
Web Archiving is NOT....

- Search engine **indexing** and summarization (such as by Google or Yahoo)
- Bookmarking
- **Cataloguing** a website
- Moving older web content to a specific part of a website and labelling it '**Archive**'
- **Downloading** or saving a file or page
- "Right-click" an image and "**Save as...**"
- **Recording** the screen while browsing the web
- Using a tool like **Camtasia** or Screencastify



4

Four Stages of Web Archiving



5

Who Does Web Archiving?

- Local and National Governments
- National Libraries and Archives
- Public Organisations
- Corporate Archives
- Research Institutions
- Museums and Galleries



6

Why is archiving the web important?



7

The Web as an Information Resource



8

National Library of Australia Domain Crawl

NATIONAL LIBRARY OF AUSTRALIA

Trove Home About Site news Statistics Tags Help Forum Contact us

It's free and it only takes a minute [Sign up](#) [Login](#)

All Books Pictures, photos, objects Journals, articles and data sets Digitised newspapers and more Government Gazettes Music, sound and video Maps Diaries, letters, archives Archived websites (1996 – now) People and organisations Lists

Keywords or URL

Limit to the gov.au web domain [Advanced search](#) [Search tips](#)

Contents

Archived websites (1996 – now)

The Australian Web Archive captures over twenty years of website snapshots of our cultural and social history.

The material includes:

- more than two decades of the PANDORA Web Archive - a collection created in partnership with cultural institutions around Australia
- government websites formerly accessible through the Australian Government Web Archive; and
- websites from the .au domain collected annually through large scale crawl harvests

Browse curated collections by subject

Arts	Health	People & Culture
Business & Economy	History	Politics
Defence	Humanities	Sciences
Education	Indigenous Australians	Society & Social Issues
Environment	Industry & Technology	Sports & Recreation
Government & Law	Media	Tourism & Travel

Featured items

<https://trove.nla.gov.au/website>

9

The Web Changes Quickly

- Web pages change, move or disappear in c. 90-100 days
- Affects everything from social media to scholarly publications
- Web archiving must be active and consistent to counteract those losses



10

British Library Gender Equality Collection

<https://beta.webarchive.org.uk/en/ukwa/collection/1942?page=1>

Cymraeg

UKWA
UK WEB ARCHIVE

[Home](#)

[Topics and Themes](#)

[Save a UK website](#)

[About Us](#)

[Contact Us](#)

Gender Equality

You are here: [Topics and Themes](#) > Gender Equality

This collection gives an insight into attitudes in the UK relating to gender equality. The collection includes material relating to bodily autonomy; domestic abuse; gender equality in the workplace; gender identity; the gender pay gap; parenting; and women's suffrage, including material relating to the centenary of the 1918 Representation of the Peoples Act....
[Read more](#)

Search within "Gender Equality"

[Bodily Autonomy](#)

A subsection of the Gender Equality collection.

[Domestic Abuse / Gender-Based Violence](#)

A subsection of the Gender Equality collection.

[Gender Equality in the workplace](#)

A subsection of the Gender Equality collection.

[Gender Identity](#)

A subsection of the Gender Equality collection.

[Gender Pay Gap](#)

A subsection of the Gender Equality collection.

[Parenting](#)

A subsection of the Gender Equality collection.

[Women's Suffrage](#)

A subsection of the Gender Equality collection.



13

Accountability Using the Web

- Important to capture web resources to ensure accountability
- Governments use the web for:
 - Publishing important resources
 - Sharing information
 - Engaging with public
 - Offering services



14

National Records of Scotland

About Home



NATIONAL RECORDS OF SCOTLAND WEB ARCHIVE

This collection of archived websites has been selected for preservation by National Records of Scotland as part of our Web Continuity Service. National Records of Scotland collects, preserves and produces information about Scotland's people and history, and makes it available to inform present and future generations.

Content ▾

[Advanced Search...](#)

index / filter: **All** A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

List **Full**

Accessing Scotland's Past

<http://www.accessingscotlandspast.org.uk/>

Accountant in Bankruptcy

<https://www.aib.gov.uk/>

Architecture and Design Scotland

<http://www.ads.org.uk>

<https://webarchive.nrscotland.gov.uk/>

15

Supporting Business

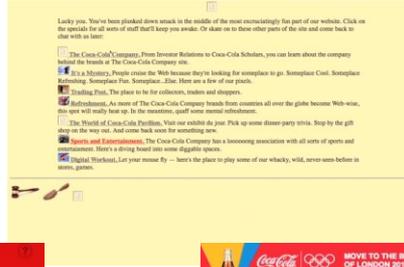
Web archiving can help businesses:

- Protect copyright/IPR
- Legislation/Regulation
- Protect profits
- Increase efficiency
- Enable transactions
- Support branding/PR
- Provide continued access



16

Supporting Business

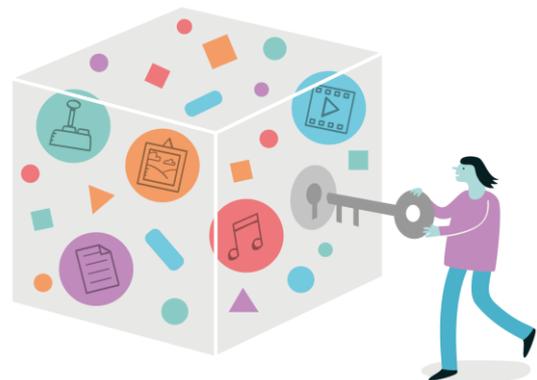


© The Coca-Cola Company - <https://web.archive.org/web/20170909015656/http://www.coca-colacompany.com/history/1s-and-0s-the-history-of-the-coca-cola-companys-website>

17

Motivations Drive Selection

- Important to understand motivations at the start
- Knowing 'why' helps shape decisions on 'how' and 'what' (to keep)



18