

Web Archiving for Beginners



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



Day One - Tue 5th BST/UTC+1 (EDT)



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



13:00 (08:00)	Welcome and Introductions
13:15 (08:15)	Why Do We Archive the Web?
13:40 (08:40)	Approaches to Selection of Web Archives
14:15 (09:15)	Group Discussion
14:45 (09:45)	Break
15:15 (10:15)	Web Archiving Workflows: <ul style="list-style-type: none">● Overview● Capture● Preserve● Playback
16:15 (11:15)	Group Discussion
17:00 (12:00)	Wrap-Up Day One

Why Archive the Web?



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



What is Web Archiving?



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



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/>

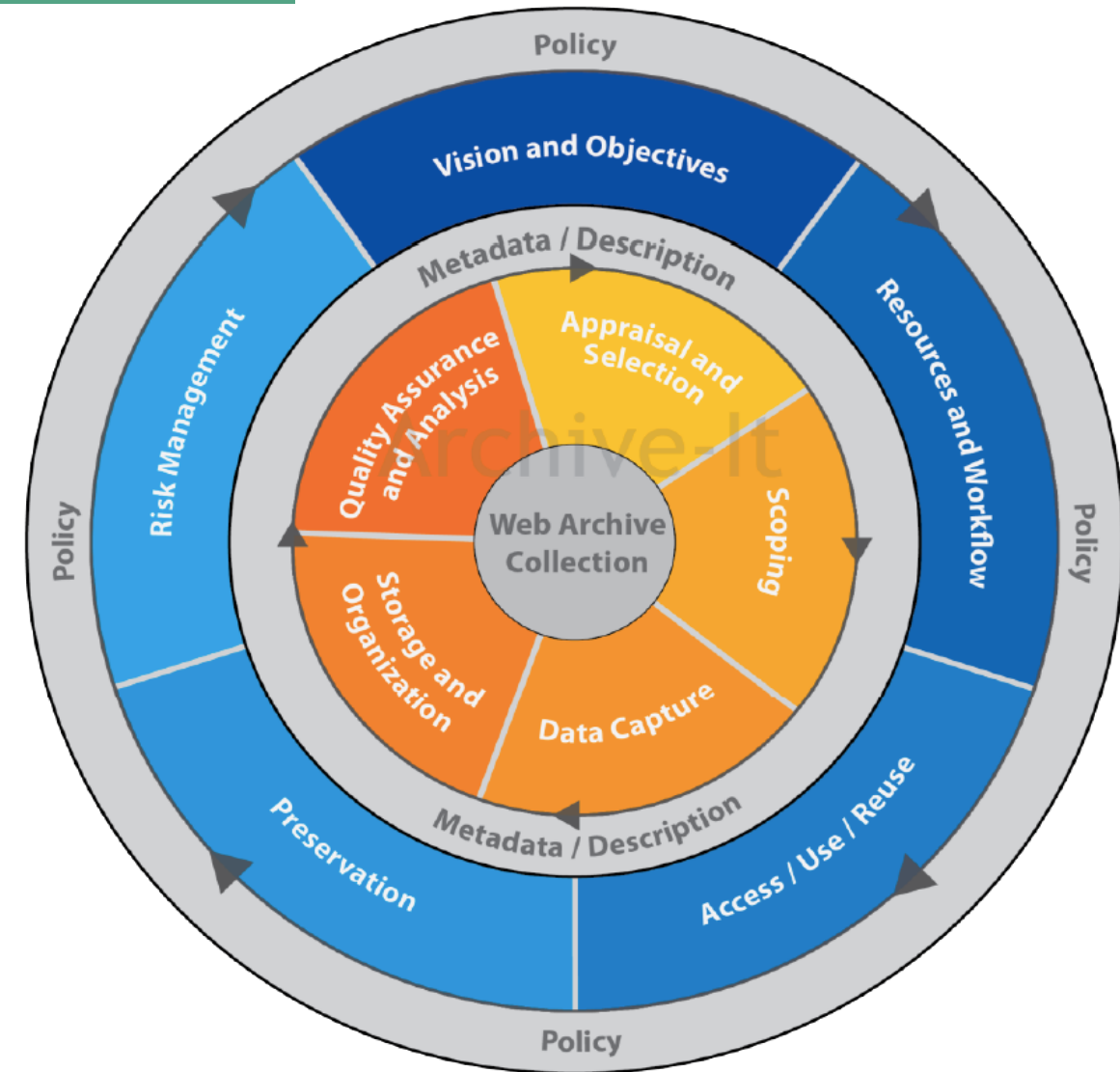
Internet Archive's Web Archiving Life-Cycle Model



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



- 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**



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



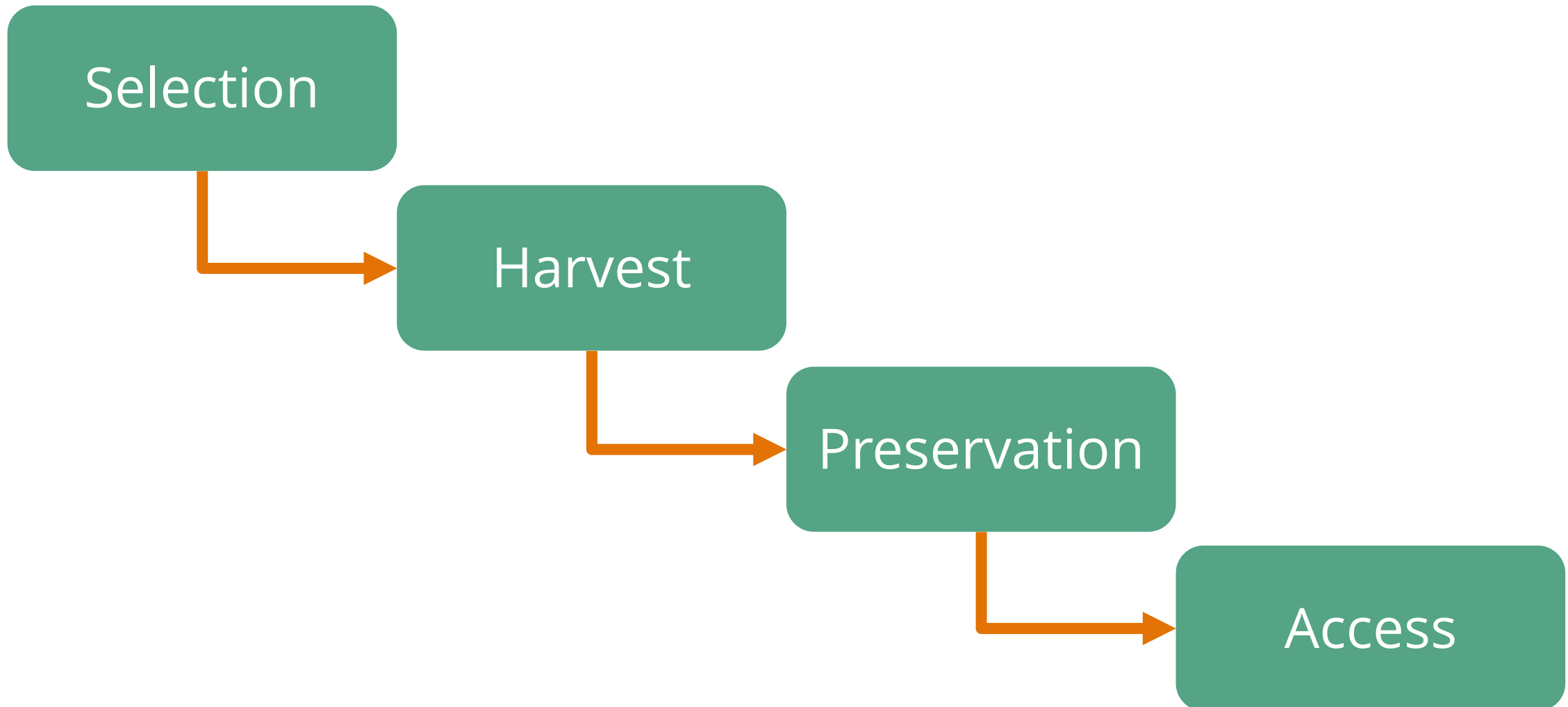
INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



Four Stages of Web Archiving



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



Who Does Web Archiving?



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



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



Why is archiving the web important?



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



www.digitalbevaring.dk

The Web as an Info Resource



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



NETFLIX



WIKIPEDIA
The Free Encyclopedia



National Library of Australia Domain Crawl



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



NATIONAL LIBRARY OF AUSTRALIA



[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
---------------------	-----------------------	---	--	---	-------------------------------------	--	----------------------	--	--	--	-----------------------

Search

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

Contents

[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](#)

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](#)
[Business & Economy](#)
[Defence](#)
[Education](#)
[Environment](#)
[Government & Law](#)

[Health](#)
[History](#)
[Humanities](#)
[Indigenous Australians](#)
[Industry & Technology](#)
[Media](#)

[People & Culture](#)
[Politics](#)
[Sciences](#)
[Society & Social Issues](#)
[Sports & Recreation](#)
[Tourism & Travel](#)

Featured items

The Web Changes Quickly



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



- 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



MySpace Music



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



engadget_{UK}

3 related articles ▾



Login



Latest in Entertainment



Wirecutter's best deals:
Save \$80 on Samsung's 2TB
T5 portable SSD

🕒 51m ago



Snapchat now lets you play
live, multiplayer games
with friends

The Internet Archive will host 490,000 music tracks 'lost' by MySpace

Nearly half a million tracks have been put back onto the internet.



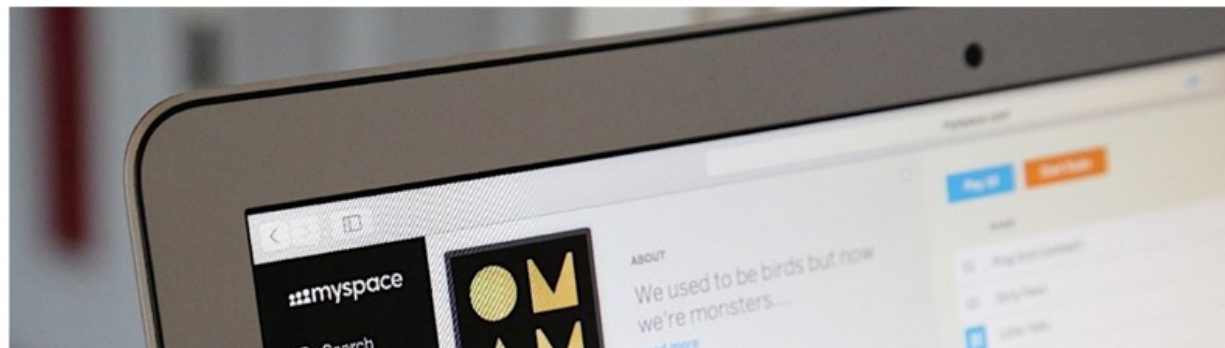
Rachel England, @rachel_england
7h ago in [Internet](#)

12

Comments

198

Shares



Start saving **now**.

Rich Documentation of Culture

- Perhaps the most democratic representation of society
- Offers windows into both dominant and sub-cultures
- Web archiving can enable broader and fairer representation
- Can capture communities outputs without interpretation by a 3rd party



British Library Gender Equality Collection



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



Cymraeg

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



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.



Accountability Using the Web



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



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



National Records of Scotland



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



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.

Search

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/>

Supporting Business



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



Web archiving can help businesses:

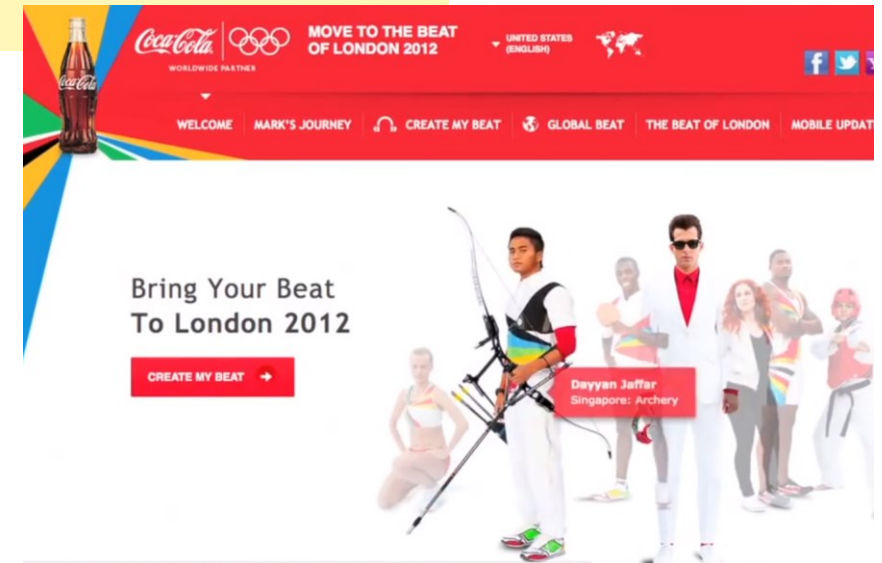
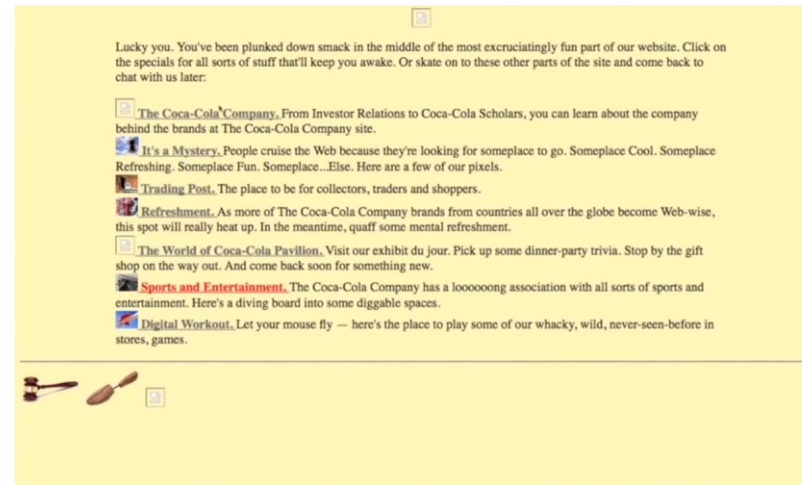
- Protect copyright/IPR
- Legislation/Regulation
- Protect profits
- Increase efficiency
- Enable transactions
- Support branding/PR
- Provide continued access



Supporting Business



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



Motivations Drive Selection



INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM



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

