



# Building the IIPC Training Program

Maria Praetzellis, Internet Archive  
Abbie Grotke, Library of Congress  
Training Working Group Co-Chairs

Sharon McMeekin, Digital Preservation Coalition

## Training Working Group

Training is a top need and opportunity for IIPC. Exchange of “how to” information is a core purpose of IIPC, and is one of the primary reasons new members join IIPC. Current members have varying approaches to archiving reflecting different institutional mandates, legal contexts, technical infrastructure, etc.), but share a need for various types of training: technical, curatorial, for practitioners and for researchers.

The **Training Working Group**, chartered in late 2017, is working to fulfill the vision of making IIPC the world leader for training on web archiving to its members, web archivists and technologists engaged in web archiving.

- The IIPC training curriculum will be crowdsourced by its members, reflecting real life needs from practitioners, and serving as a form of member engagement and contribution.
- The IIPC training curriculum will be available online under an open (CC) license for asynchronous delivery; it will also be available as a kit for delivery in hands on workshops by “certified” (literally, or figuratively) web archiving instructors for distributed delivery.
- The curriculum will be continuously updated based on feedback and assessment of the effectiveness of the online and in person events and monitoring of new

### > Events

📅 Web Archiving Conference & General Assembly 2018

📅 Web Archiving Conference & General Assembly 2019

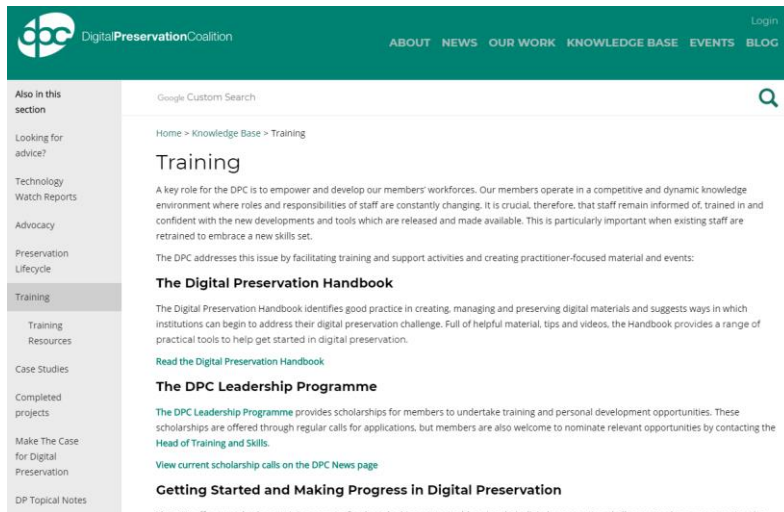
<http://netpreserve.org/about-us/working-groups/training-working-group/>

# TWG CHARTER



- Survey IIPC members for training needs;
- Survey the community at large for existing Web Archiving training curricula;
- Assemble a high quality training curriculum drawing on members' expertise;
- Deliver the training in person and/or via webinar;
- Assess and refine the curriculum based on feedback;
- Deliver a train the trainer session

# PARTNERSHIP WITH DPC



The screenshot shows the Digital Preservation Coalition (DPC) website. The header includes the DPC logo, the text "Digital Preservation Coalition", and navigation links for "ABOUT", "NEWS", "OUR WORK", "KNOWLEDGE BASE", "EVENTS", and "BLOG". A "Login" link is also present. Below the header is a search bar with the text "Google Custom Search" and a magnifying glass icon. The main content area is titled "Home > Knowledge Base > Training" and "Training". It contains a paragraph about the DPC's role in training, a section for "The Digital Preservation Handbook", and "The DPC Leadership Programme". A sidebar on the left lists various categories: "Also in this section", "Looking for advice?", "Technology Watch Reports", "Advocacy", "Preservation Lifecycle", "Training" (highlighted), "Training Resources", "Case Studies", "Completed projects", "Make The Case for Digital Preservation", and "DP Topical Notes".

- Funding from IIPC budget was available to bring on DPC as paid contributors
- Two staff at Digital Preservation Coalition working with IIPC to produce the curriculum.
- Advantages: singular voice, professional focus, experience with training development & curricular delivery
- Three Levels:
  - Beginner (Intro to Web Archiving)
  - Intermediate (Tools, etc.)
  - Advanced (Expert Topics)

# DEVELOPING THE CONTENT OUTLINE

IIPC TWG DPC content outline ☆ 🔄  
 File Edit View Insert Format Data Tools Add-ons Help Last edit was made yesterday at 11:21 AM by anonymous

100% \$ % .0 .00 123 Calibri 11 B I S A 🔍 📄 📑 📏 📐 📊 📈 📉 📉 📉

Why do we archive the web?

	A	B	C	D	E	F	G	H
1	Session	Status	Topic	Content to Cover	In-Person Format	Online Format	Learning Objectives	Content/Input Needed from TWG
2		1 Reviewed by AG and MP, and updates made.	Why do we archive the web?	To begin, instructor asks participants why they think web archiving is important and jots summaries on flipchart or digital equivalent Overview of institutions who archive the web and why; a few examples, e.g. a research project using web archives, a collection of web and social media content that documents a major social or political event, web archives used as evidence in court; high level overview of main users	Participant feedback with flipchart Presentation with visual examples	Video – recorded slide presentation Additional Links	Attendees will be able to: List the reasons for undertaking web archiving Describe how web archiving is used in specific contexts	Review content, use case, sta how long content lives on web business case for WA
3		2 Awaiting other sessions to be completed before final updates. Feedback from TWG members would be helpful.	What are web archives?	Websites or web content as institutional or business records; web archives as art objects; web archives as research data Some recipes for getting started	Presentations with visual examples	Written content with images/graphics	Attendees will: Be able to understand the main purposes of web archiving Be familiar with web archiving 'first steps'	Review content
4		3 First draft completed, awaiting review by TWG	Main Concepts and Technologies: 3: Overview, 3a: Capture, 3b: Preserve, 3c: Playback	how archiving the web works (Harvesting, Standards, Access/Playback); brief overview of main types of tools/tech; Examples of tools used in web archiving linked to their main functional use and user base (from session 3) [emphasis on tools with friendly GUIs]; brief overview of GitHub and/or other web archiving tool repositories	Presentation with visual examples	Video – recorded slide presentation; Videos of tool demos and practice tasks *not complete: awaiting final version of slides	Attendees will be able to: - Describe the basic processes of web archiving - Describe the types of tools and how they relate to the basic processes - Be able to operate choose web archiving tools based on function and capabilities - Know where to find and download tools for web archiving	All content
				Examples of tools used in web archiving linked to their main functional use and user base (from session 3) [emphasis on	Presentation, 2 or 3 short tool		Attendees will: Be able to operate simple	

## BEGINNER TOPICS

1. Why do we archive the web? -- completed
2. What are web archives? - almost completed, under TWG review
3. Main Concepts and Technologies - almost completed
4. Brief Overview of Tools -- *now folded into session #3*
5. Main users groups - on hold
6. Identifying Benefits and Risks - completed
7. Risk Assessment exercise - completed
8. Main Approaches to Web Archiving - in progress
9. What to include in a policy - in progress
10. How to write a business case
11. Business Case Elevator Pitch exercise
12. Case Studies - *Recording interviews this week!*
13. Workflows Practical Workshop

# SAMPLE: WHY DO WE ARCHIVE THE WEB

## Content to cover:

To begin, instructor asks participants why they think web archiving is important and jots summaries on flipchart or digital equivalent. Overview of institutions who archive the web and why; a few examples, e.g. a research project using web archives, a collection of web and social media content that documents a major social or political event, web archives used as evidence in court; high level overview of main users.

## Learning Objectives:

Attendees will be able to: List the reasons for undertaking web archiving; Describe how web archiving is used in specific contexts

The presentation consists of 16 slides, each with a number in the bottom left corner:

- Slide 1:** Title slide: "Why Do We Archive the Web?"
- Slide 2:** "What is Web Archiving?" - Definition: "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 logo and URL: <http://iipc.netpreserve.org/web-archiving/>
- Slide 3:** "How Does Web Archiving Work?" - A flowchart showing: Selection → Harvest → Preservation → Access.
- Slide 4:** "Who Uses Web Archives?" - List: Local and National Governments, National Libraries and Archives, Public Organisations, Corporate Archives, Research Institutions, Museums and Galleries. Includes an illustration of people looking at a globe.
- Slide 5:** "The Web as an Information Resource" - Logos for YouTube, Google, AP, Microsoft, Netflix, Amazon, Facebook, Twitter, LinkedIn, Uber, and others.
- Slide 6:** "National Centre for Accessible Digital Content" - Screenshot of the Trove website.
- Slide 7:** "The Web Content Quays" - Text: "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." Includes an illustration of a person at a computer.
- Slide 8:** "All-Genre Music" - Text: "The Internet Archive will host 490,000 music tracks lost by Myspace." Includes a screenshot of the Internet Archive website.
- Slide 9:** "New Democracies or Causes" - Text: "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." Includes an illustration of people.
- Slide 10:** "UKWA" - Screenshot of the UK Web Archive website.
- Slide 11:** "Accountability Does for Web" - Text: "Important to capture web resources to ensure accountability." "Governments use the web for: Publishing important resources, Sharing information, Engaging with public, Offering services." Includes an illustration of a person at a computer.
- Slide 12:** "National Archive of Scotland" - Screenshot of the National Archive of Scotland website.
- Slide 13:** "Supporting Business" - Text: "Web archiving can help businesses: Protect copyright/IPR, Protect profits, Increase efficiency, Enable transactions, Support branding/IPR, Provide continued access." Includes an illustration of a person at a computer.
- Slide 14:** "Case-Cox" - Screenshot of the Case-Cox website.
- Slide 15:** "What Would You Miss Without It?" - Illustration of a person at a computer with a large question mark.
- Slide 16:** "Motivations Drive Success" - Text: "Important to understand motivations at the start." "Knowing 'why' helps shape decisions on 'how' and 'what' to keep." Includes an illustration of a person at a computer.

- Ambitious nature of the charter combined with volunteer nature of IIPC made development slower than anticipated - bringing on DPC was critical to keeping things moving
- In creation of training modules:
  - Anticipating varying local policies and situations of organizations and individuals who might be interested in the training
  - Creating customizable slide decks and speaker notes
  - Not being too UK/US centric with examples in training materials
  - Ensuring that we keep materials up-to-date



## NEXT STEPS

- Incorporate case study videos
- Finalize slide decks and speaker notes
- Record video versions of slide presentations
- Plan a train-the-trainer for later 2019
- Begin to post materials online for others to use

## CONTACT US!

Maria Praetzellis -- [maria@archive.org](mailto:maria@archive.org)

Abbie Grotke -- [abgr@loc.gov](mailto:abgr@loc.gov)

### For IIPC members:

The Training Working Group communicates via mailing list: [training@iipc.simplelists.com](mailto:training@iipc.simplelists.com). To join the mailing list send email to [communications@iipc.simplelists.com](mailto:communications@iipc.simplelists.com)

You can also join the #training channel on IIPC Slack