

# Getting Started with Digital Preservation

Anna Perricci  
Anna.Perricci@dpconline.org



Digital**Preservation**Coalition

# Aims for the Session

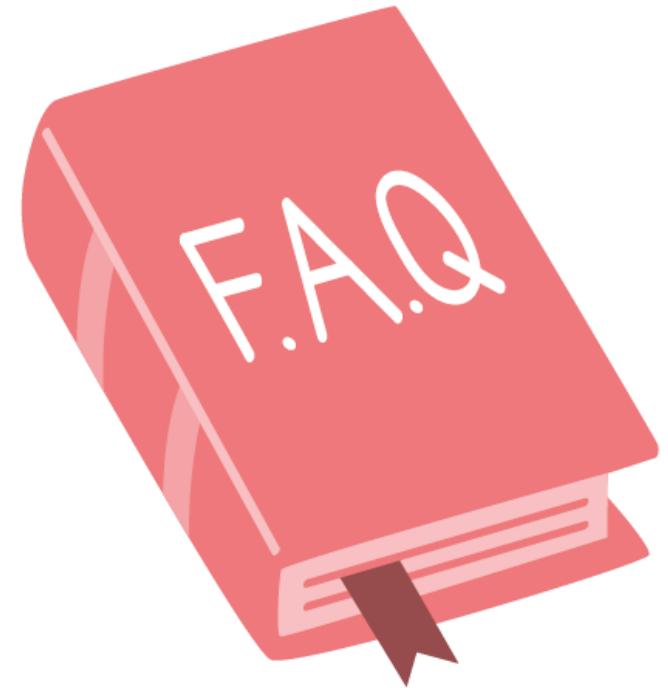
- Establish what we mean by “Digital Preservation”
- Understand the challenges faced
- Identify some key first steps to take
- Know where to look for help and resources



# What is Digital Preservation?

“the series of **managed activities** necessary to ensure **continued access** to digital materials for **as long as necessary.**”

- *The Digital Preservation Handbook*



# Why Preserve Digital Content?



- Legal and Regulatory Compliance
- Increased Efficiency
- New Revenue Streams
- Enabling Research
- Documenting Cultural Heritage
- Ensuring Transparency and Accountability

# Digital Preservation Includes:

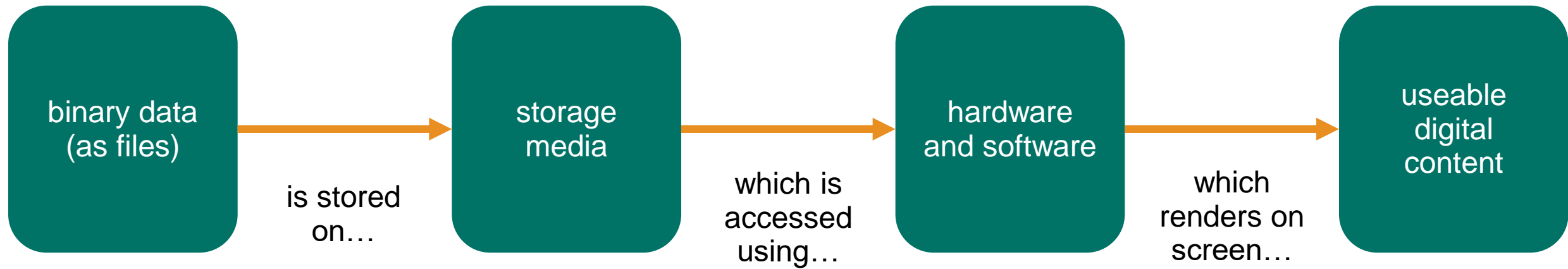
- Resourcing
- Sustainability
- Policy and strategy
- Managing legal issues
- IT capabilities
- Continuous improvement
- Developing workflows
- Preservation planning
- Metadata and documentation
- Understanding user needs
- Discovery and access

# And Digital Preservation Isn't...

- *Only* a technical problem
- Just storage or digitization
- An easy fix with a single solution
- A passing trend
- Something you can 'set and forget'



# What Are We Preserving?



# File/Data – Media – Access – Rendering – Content

```
0000000 0000 0001 0001 1010 0010 0001 0004 0128
0000010 0000 0016 0000 0028 0000 0010 0000 0020
0000020 0000 0001 0004 0000 0000 0000 0000 0000
0000030 0000 0000 0000 0010 0000 0000 0000 0204
0000040 0004 8384 0084 c7c8 00c8 4748 0048 e8e9
0000050 00e9 6a69 0069 a8a9 00a9 2828 0028 fdfe
0000060 00fc 1819 0019 9898 0098 d9d8 00d8 5857
0000070 0057 7b7a 007a bab9 00b9 3a3c 003c 8888
0000080 8888 8888 8888 8888 288e be88 8888 8888
0000090 3b83 5788 8888 8888 7667 778e 8828 8888
00000a0 d61f 7abd 8818 8888 467c 585f 8814 8188
00000b0 8b06 e8f7 88aa 8388 8b3b 88f3 88bd e988
00000c0 8a18 880c e841 c988 b328 6871 688e 958b
00000d0 a948 5862 5884 7e81 3788 1ab4 5a84 3eec
00000e0 3d86 dcb8 5cbb 8888 8888 8888 8888 8888
00000f0 8888 8888 8888 8888 8888 8888 8888 0000
0000100 0000 0000 0000 0000 0000 0000 0000
*
0000130 0000 0000 0000 0000 0000 0000 0000
000013e
```



## Getting Started with Digital Preservation

Anna Perricci  
Anna.Perricci@dpconline.org



Digital Preservation Coalition

# What Are We Preserving?

- Accessing digital content involves:
  - Being able to find the data
  - Having any necessary security permissions
  - Using the right hardware and software set-up
  - Knowing how to use that hardware and software
- A lot of dependencies and things that can go wrong!
- Digital preservation seeks to actively manage those potential risks

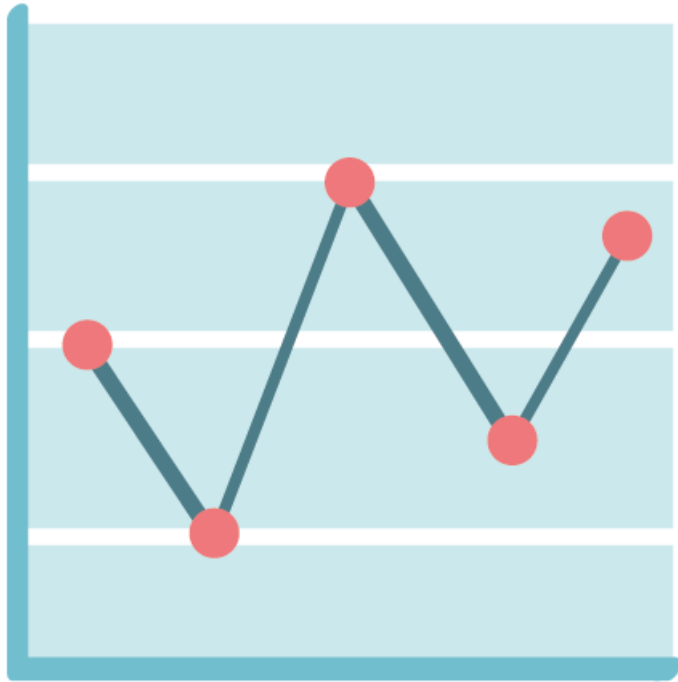
# How Can I Get Started?

info@dpconline.org  
www.dpconline.org



Digital Preservation Coalition

# Understand Where You Are



- Benchmark where you are now
- Helps set goals for the future
- Helps with advocacy for DP
- Continuous improvement approach recommended
  - Facilitates incremental planning and development
  - Setting of achievable targets
- Aim for “good” not “best”

# NDSA Levels of Digital Preservation

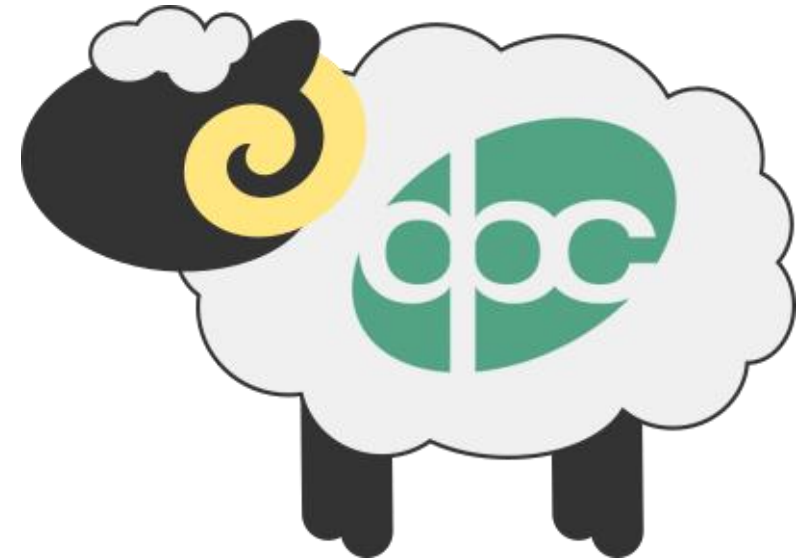
- A maturity model with a focus on:
  - Processes
  - Technological infrastructure
- Simple to understand and use
- Quick to deploy

Assessment across  
5 functional areas:

- Storage
- Integrity
- Control
- Metadata
- Content

# The Rapid Assessment Model

- A maturity model that takes a holistic look at digital preservation
- Assessment across 11 sections:
  - 6 Organizational capabilities
  - 5 Service capabilities
- RAM aims to be:
  - Applicable to any organization and content
  - Strategy and solution agnostic
  - Based on good practice
  - Simple to understand and quick to apply



# Get to Know Your Digital Content

- Important to know what you've got (or will have)
  - Essential for planning
  - Helps with advocacy/building a business case
- Capture information in:
  - Digital Asset Register
  - File Manifest



# Creating a Digital Asset Register



- Useful tool for understanding digital content
  - Know what is held
  - Facilitates management
  - Info for advocacy
- DPC Toolkit offers guidance on:
  - What makes a good DAR
  - Step by step guide to development
  - DAR Template (with stats dashboard)

# Gathering Info on Content for Preservation

Toolkit also provides guidance methods for gathering information for your DAR

- Extracting from existing sources
- Surveying physical storage media
- Using characterization tools
  - e.g. DROID, fido, Seigfried, or JHOVE
- Interviews with or surveys of:
  - Content creators
  - Depositors



# Nothing Works Without Skilled Staff

- Staff: THE fundamental resource need for any successful preservation program
- Digital preservation is a multi-disciplinary undertaking
  - Knowing the skills needed is essential
  - What skills might need to be added to/further developed on your team?
- Skills and roles needed depend on:
  - Organizational context and structure
  - Current and planned DP capabilities



# DP Competency Framework & CAT

- Framework provides info on the skills, knowledge and competencies required for digital preservation
- Core principles:
  - Flexibility – able to be deployed for a variety of purposes
  - Closely linked to DPC RAM
- Competency Audit Toolkit (CAT)
  - Practical implementation of framework
  - Audit individual and organizational skills

| Comp. Area                                    | No. | Skill Element                           |
|---|-----|---|
| <b>Governance, Resourcing, and Management</b> | 1   | Policy Development                      |
|   | 2   | Risk Management                         |
|   | 3   | Resource Management                     |
|   | 4   | Staff Management                        |
|   | 5   | Strategy and Planning                   |
|   | 6   | Analysis and Decision-Making            |
| <b>Communications and Advocacy</b>            | 7   | Effective Communication                 |
|   | 8   | Collaboration and Teamwork              |
|   | 9   | Stakeholder Analysis and Engagement     |
|   | 10  | User Analysis and Engagement            |
|   | 11  | Advocacy                                |
|   | 12  | Training                                |
|   | 13  | Producing Documentation                 |
| <b>Information Technology</b>                 | 14  | General IT Literacy                     |
|   | 15  | Computer Programming                    |
|   | 16  | System Procurement                      |
|   | 17  | Storage Infrastructures                 |
|   | 18  | Information Security                    |
|   | 19  | Workflow Development and Implementation |
| <b>Legal and Social Responsibilities</b>      | 20  | Legal and Regulatory Compliance         |
|   | 21  | Environmental Impact                    |
|   | 22  | Inclusion and Diversity                 |
|   | 23  | Ethics                                  |
| <b>Digital Preservation Domain Specific</b>   | 24  | Metadata Standards and Implementation   |
|   | 25  | Information Management Principles       |
|   | 26  | Approaches to Preservation              |
|   | 27  | DP Standards and Models                 |
|   | 28  | Managing Access                         |

# Novice to Know-How

- **Access is free for all**
- Self-paced online beginner's DP training
  - Funded by The National Archives (UK)
- Introduces DP & has practical guidance on:
  - Selecting and transferring content
  - Preparing content for preservation
  - Managing long-term preservation
  - Providing access
- Completed by 4000 learners in 62 countries
- Also courses on email preservation & DARs

# Novice to Know-How (N2KH)



Organizational Alignment

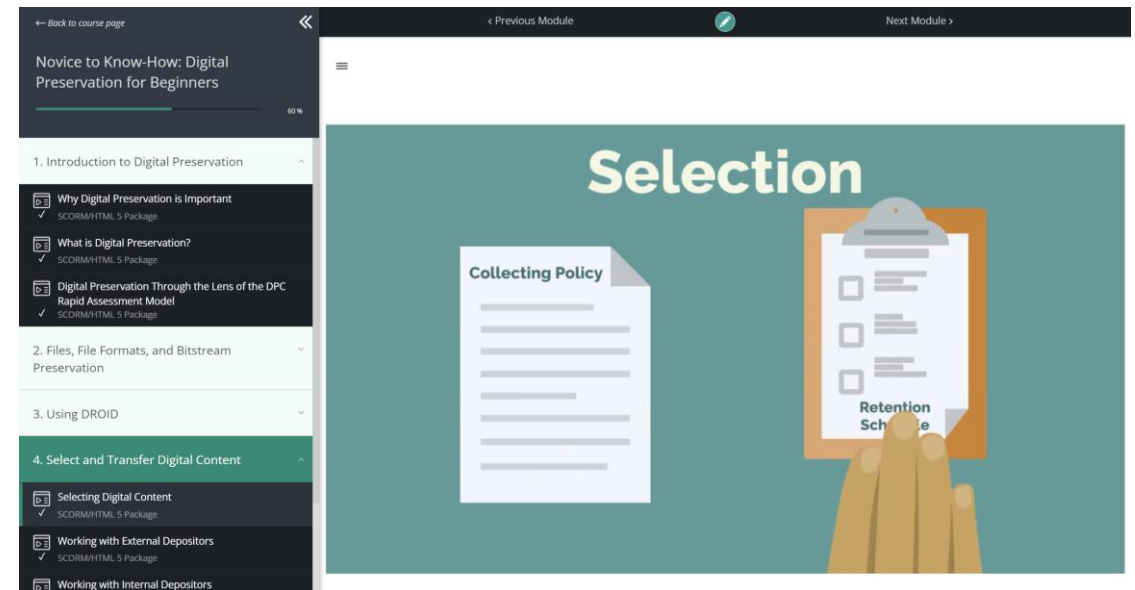
No One-Size-Fits-All

Good vs Best Practice

Start Small

Starting out in digital preservation can seem intimidating! Click on the boxes to the left to find some advice to help you get started and plan for your work.

Previous Next



← Back to course page

Novice to Know-How: Digital Preservation for Beginners

1. Introduction to Digital Preservation

- Why Digital Preservation is Important
- What is Digital Preservation?
- Digital Preservation Through the Lens of the DPC Rapid Assessment Model

2. Files, File Formats, and Bitstream Preservation

3. Using DROID

4. Select and Transfer Digital Content

- Selecting Digital Content
- Working with External Depositors
- Working with Internal Depositors

← Previous Module

Next Module >

## Selection

Collecting Policy

Retention Schedule

[https://dpc.  
getlearnworlds  
.com/](https://dpc.getlearnworlds.com/)

# N2KH Feedback

The course was extremely well structured. Key concepts were explained from the ground up, allowing us to build up a good knowledge base from strong foundations

**Surprisingly,  
not boring!**

The course was excellent in covering larger overviews as well as nitty gritty details

It has made me more confident that digital preservation is something that I could now do rather than just aspire to

I really felt that this was a practical course – I was able to go away from each module and think about it in the context of my organisation and start to experiment with small steps using the knowledge I had learnt through that day's module

# Bonus round WEB ARCHIVING



Digital**P**reservationCoalition

# What is Web Archiving?

- Functional ('browsable') copy of a web page or website
- More than a screenshot – you can browse or interact with pages as they were when they the copy was captured (harvested)
- You cannot use web archives to restore a website to a server (web archives are not a 'back up' of a website)
- Web archives are usually saved as .WARC files

# How are web archives created and used?

- The most common way to create web archives is via a program known as a crawler or spider
  - This is an automated tool to make a functional copy of web pages
  - This is how the Internet Archive and most other services create web archives
- Web archives (WARC files) can generally only be opened with specialized software
- There are also tools for analyzing the contents of web archives

# The Internet Archive

- The oldest and most well-known web archiving service
  - Founded in 1996 by Brewster Kahle
  - Current size of collection: 835 billion web pages
- Can browse websites only if those pages were crawled and saved to the collection
- Archive-It – suite of web archiving tools by subscription
- Save Page Now – free tool to save one page at a time
- End of Term Web Archive (scheduled, very important in 2025)

# Other options and web archiving tools

- The Internet Archive does not have a peer or competitor in all aspects of the services it offers, but there are some alternatives
- Conifer (<https://conifer.rhizome.org/>) relies on individuals to browse pages to ‘record’ them (make functional copy)
  - Previously known as Webrecorder.io
- Webrecorder projects
  - Replay.Web (<https://replayweb.page/>)
  - Browsertrix (<https://webrecorder.net/browsertrix/>)

# Social Media? Apps?

- Big, unsolved challenges here
- Is it a technology problem or a regulatory problem?
  - Likely is both
- No incentive for social media companies to think about long-term stewardship and many are simply trying to block unauthorized harvesting for ingest into data corpora used to train AI



**Stay tuned!**

Keynote by Michael Nelson after lunch:

**A Vision for Trustworthy  
Web Archiving**

# What Next?

info@dpconline.org  
www.dpconline.org

# Just Start Doing Stuff!



- Many people paralysed by fear of doing something wrong when starting...
- But there are lots of low-risk things to try:
  - Experiment with tools, e.g. characterization
  - Build simple workflows/procedures
- Set small, achievable goals
- Make sure to think about legal issues
  - Are agreements and licenses digital ready?
- And always remember documentation!

# Next Steps

- Get everyone to complete N2KH
  - Sets a common foundation of knowledge
  - Can also be sneak attack advocacy!
- Develop a DP Policy
- Prepare a Business Case for more resources
- Consider procuring a repository
- Work gradually to increase capabilities

**There are DPC  
Toolkits to help  
with these!**

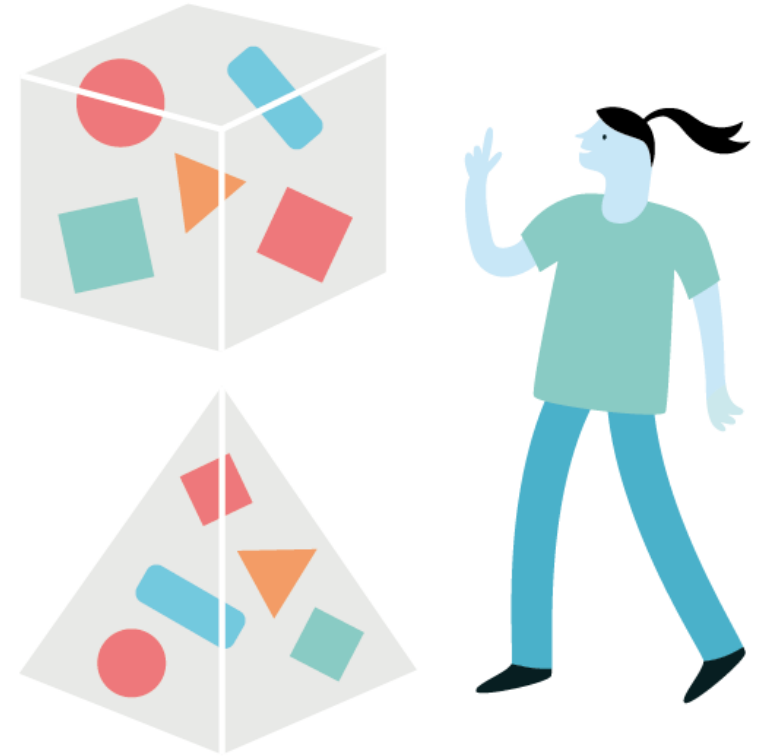
# Some Other DPC Resources

For everyone:

- Tech Watch Reports & Guidance Notes
- Wide range of toolkits
- Digital Preservation Handbook
- #DPClinic and #DPCConnect

For members:

- Topical events
- Special interest groups
- Career development grants



Find everything mentioned  
here and more at:  
**[www.dpconline.org](http://www.dpconline.org)**