Documentation at the BFI

Lucy Wales and Jennifer Macmillan
How we got to where we are…

How did we manage user support and documentation before?

- Introductory Training
- Cataloguing Manual
- Help Documentation
- Various Inboxes

- Ongoing user support and training
- One portal to raise questions, report bugs and access user help guides and documentation
- Shared team inboxes
Why is digital preservation documentation important to you and your organization?

1) WHY & HOW
2) SUPPORT
3) STANDARDS
What tools or platforms do you use to create and provide access to your documentation?

- User Knowledge Base and Admin Knowledge Base
- CID and Digital Preservation Systems Service Desk
- JIRA Dashboard and Projects
- Axiell Collections
- GitHub Repository
- Videos
User Knowledge Base and Admin Knowledge Base

Digital Preservation Infrastructure (DPI) overview

What is the Digital Preservation Infrastructure?
The Digital Preservation Infrastructure - also known as DPI - is the suite of integrated systems that the ERI Knowledge and Collections directorate has developed to preserve and provide access to the digital collections held in the ERI National Archive.

Further Reading
- General - the Handbook: https://www.dpconline.org/handbook
- General - How to Know More: https://www.dpconline.org/dlpconline-train-your-staff/2021-online-training

A brief description of the systems involved:

Documentation: CD
- An index of the original content and a CD record to describe the preservation file in its context
- Descriptive and some technical metadata about the digital objects (moving images, special collections, library) recorded in CD records
- All technical metadata for each digital file recorded in CD media record (coming soon)

Media Asset Management:
- Originally, images were our Media Asset Management (MAM). Images were removed from DPI in March 2021.
- The DPI ingest process is now managed by Python scripts. These are created and managed by the Data and Digital Preservation department.
- The functionality to download preservation files from data tape is being provided by the IRIS Browser tool on the BRICs Access server.
- A persistent replacement for image's core functionality will be developed in Winter 2023 - Spring 2024.

Instant access storage: NAS servers - Illunx, QNAP and G-Back

Adding websites to the Allowed list in the Firewall

As part of the Cyber Security being applied to DPI network, access to the Internet is being restricted. Basically all sites are blocked unless they have been added to the list of acceptable sites in the Firewall.

The instructions below detail how to add a site to the list.

Stage 1
1. A request will come in to the help desk asking for a site to be authorised. This request will need to be authorised by the Head of Data & Digital Preservation (Stephen McConnachie) or Digital Preservation Manager (Lucy Wales).
2. It is unlikely that the person requesting the website will know the URL. A google search outside of DPI will usually find the website.
3. To identify the exact URL for the licencing server, go to the support pages and search on Firewall.
Service Desk

DPI Internet destination access request

Raise this request on behalf of:
Collections Systems (collectionsystems@bfi.org.uk)

Reason for request:
Normal text

Specific URLs:
Please specify the full destination (including URLs) for approval. If not known, please provide qipr.com or support@qipr.com

Priority:
Medium

Assign a priority to your request:
Share with:
Share with BFI Users

Sent

DPI Internet destination access request

Link Issue

Add Checklist

Description:
File download for restoration

Record number:
https://silverstar.fromsma.org.uk/1567572796631308

Priority:
Medium

Waiting for customer

Your planned fields

Team member:
None

Details

Assignee:
Digital Preservation Systems

Reporter:

Request Type:
DPI Internet destination access request

Knowledge base:
15 related articles

Approvers:

Digital Preservation Systems
JIRA Dashboard and Projects
Axiell Collections

Access conditions

A statement on the level of access permitted to BFI staff and other stakeholders, as agreed by the Head of Collections Management. The access permissions are sometimes informed by the requirements of the lender, copyright holder or donor or are sometimes informed by the copyright conditions or condition of the item.

Before reusing BFI National Archive digital collections please ensure the required clearances from copyright holders, contributors or other stakeholders have been obtained for the specific use.
GitHub Repository

https://github.com/bfidadigipres/dpx_encoding/
Videos
Lucid Chart
Autoingest - Service Overview

Autoingest automates the verification, ingest, full and deletion of media from approved ingest locations.

- Validation
  - Filename must be correctly formatted
  - Media must be fully described in CID
  - Media must be valid
- Ingest
- Persistence
- Visualization
- Related articles

To learn how to use Autoingest, please read the Autoingest - User Guide.

About

Autoingest is a digital preservation workflow service. It automates the validation, ingest, persistence and generation access protocols. It can handle files individually or in bulk.

Autoingest connects to digital preservation infrastructure systems, processes and hardware together.

As the Autoingest service is always running, media is continuously queued for ingest and deletion. Media - is left in place, and will be re-used when needed. Successfully ingested and persisted media is stored in the system.

Autoingest records all media validation, ingest status and persistence status messages in a log file. The log file is documented in CID before it can be processed by Autoingest.

Every file proceeds through the following three stages before it is deleted: validation, ingest and persistence.

- Validation
- Media must pass these three sets of validation tests before it is ingested.

- Ingest

- Persistence

- Visualization

- Related articles

To learn how to use Autoingest, please read the Autoingest - User Guide.

Advanced details

Page information

View Source

Hide related pages...

View Storage Format

Copy

Move

Export

Archive

Delete

Presenter mode

Analytics

Change page owner

Page history

Attachments

Resolved comments

Search

Code snippet

Display code with syntax highlighting

Info panel

Highlight information in a coloured panel

Quote

Insert a quote or citation

Decision

Capture decisions so they're easy to track

Divider

Separate content with a horizontal line

Expand

Insert an expandable section

View more
Examples

CID and DPI systems  /  Service Desk  /  Auto ingest  /  User Guide

Error Log

This error log is created at 6pm every morning. The log is called current_errors and it is located in the top level of the auto ingest folder tree in all operational shares in Film - Audio Operations, Digital Operations, Film Operations, Video Operations, Special Collections - as well as in the ingest share used by the Data team.

There is a CSV version. Check the error log to find out which files have not successfully ingested. You can filter on path to view only those rows that relate to the ingest share where you work - excluding the other operational areas where you have no ingest responsibility.

This table below gives the error message, the meaning and actions to take to address the problem and progress the file in the next cycle of Auto ingest.

<table>
<thead>
<tr>
<th>Error Message</th>
<th>Meaning</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot parse part/whole from filename</td>
<td>This indicates poorly formed part/whole syntax. Some common examples: missing underscore after Object Number, missing leading zero before a number below 10 - 19f instead of 01f, presence of extra text such as Rea1 or K.1 = Rea1K1 instead of 01f06, presence of trailing version increment - 01c0f1-2, 01c0f1-3, etc, part greater then whole - 05c0f3, 05c0f1, etc, first part is 00 and not 01</td>
<td>Change filename to use correct part/whole syntax making sure to use an underscore to separate Object Number, not a hyphen or space. Numbers up to 4 digits long are accepted eg 0000101999. Also make sure to use a leading zero for a number below 10, and no extra characters such as Rea1, R, Part, Pt. Also make sure the part is not greater than the whole - e.g. 01c0f1 and 01c0f2 are fine, but x01c0f2 is not. Please make sure the 'f' is lowercase with no gaps between.</td>
</tr>
<tr>
<td>Extension does not match &lt;file_type&gt; in record</td>
<td>This indicates a mismatch between the file type in the CID record and the file extension. Some common examples: CID name states jpg, file extension is mov. CID record states .jpg, file extension is tif.</td>
<td>In some cases this error may indicate Incorrect Object Number use inFilename, so that should be checked first. Most commonly, the CID record must be changed, to document the file type accurately.</td>
</tr>
<tr>
<td>If header failed to read file [file status]</td>
<td>This indicates a serious problem with the file's metadata header at a level that prevents it from being preserved in DPI. Some common examples: No video stream within the container, fundamental metadata missing from container or video stream, which would compromise the files usability.</td>
<td>In most cases this is a poorly encoded file, which must be re-encoded or re-acquired, as it cannot be ingested to DPI. Use Metadata to examine the files metadata, and re-encode / re-acquire as appropriate.</td>
</tr>
</tbody>
</table>

Note: For ingest, files which are successfully ingested are promoted as exact copies; additionally, access profiles we are supported. Files may simply be stored or stored with an additional proxy made available or stored with an XML file. This file contains an auto ingest directory structure. The preservation workflow used by Auto ingest is no longer part directory structure.

The video, audio, or proxy/image subfolders, viewing proxies for video or image files will be made successfully whenever a missing automated annotation still uses different ingest folder path on their structure being retained or to continue their existing ingest practices, or simplifying ingest by using the autodetect folder.

A file. The directory tree is mapped to auto ingest by a configuration file. The tree is organized hierarchically and options for storing, transcoding and adjusting media.
Examples

Admin Knowledge Data

- Heilige 2022 Processes
  - Adding Images URLs to H22 Google Sheets (redundant)
  - Adding proxy URLs to H22 RNA Google Sheets
  - Aspera - Building Aspera Reports in Aspera Console
  - Aspera Onboarding
  - Aspera System Support
  - 5F file checking sampling (VW WIP)
  - 5F Supplier invoices
  - CID manual data Input processes for H22 5F Film Digitisation BPF Player titles
  - Delivering Flats for BFI Reelplay (WIP)
  - Exceptions - Editing CID item records
  - Film Workflow
    - Film Digitisation Workflow and Processes
    - Framework Supplier allocation spreadsheet
    - Google Sheets - Data gathering for Google Sheets
    - Items digitised through shared container
    - Logging - WIP NW
    - MediaConch - TSN File Validation with MediaConch - WIP
      - MediaConch Policies
      - Mediatheque Restrictions
    - Monthly Stats for H22 Access strand
    - Monthly Stats - NEW
  - Proxy supply to rightsholders (WIP)
  - Removing items from Containers post digitisation
  - Returning files to RNAs
    - Calculating transfer for multiple HODs
    - CID Input data
    - Copying files to HDD with 7-Zip Encryption (WIP)
    - Generating RNA Tiffist with checksums (WIP NIN)
  - Rights nap4 requests - download, rewrap and supply
  - RNA tax requests - download and supply
  - Splitting WARNING / INPD Reports
  - Troubleshooting - Processing Segmentation: Autoingest
  - Uploading the Google Data Studio H22 Dashboard
When do you create documentation and how often is it reviewed and updated?

Inception (2010), dir. Christopher Nolan
What is the update process and how do you manage versions?
What is next for digital preservation documentation at your organization?
What tips do you have for people starting out on documenting their digital preservation activities?

- Engage + research your audience
- Research systems for documentation
- Formatting is key
- Link to other articles
- Make documentation a key part of all of your processes, so it’s regular and sustainable
Thank you

Any Questions?

lucy.wales@bfi.org.uk
jennifer.macmillan@bfi.org.uk