

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

CID and DPI Systems / Service Desk / Digital Preservation Infrastructure (DPI) overview

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 BFI Knowledge and Collections directorate has developed to preserve and provide access to the digital collections held in the BFI National Archive.

Further Reading

If you would like further information about Digital Preservation theory and practice, the Digital Preservation Coalition has some very useful resources:

- general the Handbook: https://www.dpconline.org/handbook
- general Novice to Know-How online course: https://www.dpconline.org/digipres/train-your-staff/n2kh-online-training
- audiovisual specific Ashley Blewer's 'Pragmatic Audiovisual Preservation' guide: https://www.dpconline.org/news/pragmatic-audiovisual-preservation-publicrelease

A brief description of the systems involved:

Documentation: CID

- No ingest to DPI is possible without a CID record to describe the preservation file in its context
- · Descriptive and some technical metadata about the digital objects (moving image, special collections, library) recorded in CID records
- All technical metadata for each digital file recorded in CID media record (coming soon)

Media Asset Management:

- Originally, Imagen was our Media Asset Management (MAM). Imagen was 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 EON Browser tool on the BK-CI-Access server.
- A permanent replacement for Imagen's core functionality will be developed in Winter 2022 Spring 2023.

Instant access storage: NAS servers - Isilon, QNAP and G-Rack

Adding websites to the Allowed list in the Firewall



Owned by Collections Systems, created with a template ••• Last updated: just a moment ago by Digital Preservation Systems • 4 min read • 🗠 3 people viewed

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.

Instructions

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

CID and DPI Systems / Service Desk DPI Internet destination access reques	st			
Raise this request on behalf of*				
Collections Systems (collectionssystems@bfi.org.uk)				
Reason for request * Normal text ▼ B I ▲ ∨ III IIII IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	() (1) 19 + -			
	Projects / 🙎 User Issues / 🚦 UI-5201			4[£] ⊙ 1 ம ≪ …
Please describe the context for the request in relation to BFI National Archive work. For exam audio processing application, required by the application at start up. Application used for BFI I Show more	DPI Internet destination access request		Waiting for customer 💌	🗸 Actions 🖌
Specific URLs*	Chink issue V Add Checklist ····		Your pinned fields	^
	Hide deta	ails	Team member	3 None
Please specify the full destination (including URLs) for approval. If not known, please provide qnap.com or support qnap.com	View request in portal Description		Details	^
Priority*	File download for restoration		Assignee	Digital Preservation Systems
Medium	Record number		Barrator	
Assign a priority to your request.	[https://silversalt.froms		Reporter	9 ieu fierantes
B Share with BFI Users ✓			Request Type	 DPI Internet destination access request
Send Cancel	Priority = Medium		Knowledge base	15 related articles
			Approvers	

SM Dissubance With Communities

Default dashboard 🛨 Refresh 🗘 Edit 🖋 🚥 ,≝ C3 Q 45 C3 🖏 Pie Chart: User Issues - Open issues Filter Results: User Issues - Open issues (Unassigned) T Key Summary 🛧 P Status Components Created Updated UI-5201 DPI Internet destination access request WAITING FOR APPROV... 13/Sep/23 13/Sep/23 1-1 of 1 () 1 minute ago ,≝ C3 Q Filter Results: User Issues - Assigned to me (Active) T Key 🗸 Summary Projects / DPI Systems Roadmap ? UI-5200 Record deletion request Timeline 📣 Give feedback ሰ Export 🚥 < Share 2 UI-5199 Record deletion request Q 8 Status category 🐱 Epic 🛩 Label 🗸 Type 🗸 ♣ View settings UI-5198 Workflow crashes when trying to add fields to view UI-5197 Barcode not correctly recorded in acquisition record SEP 5 DSR-104 ъ ≪ … × UI-5196 Unable to edit compression code field on Axiell ĥ \odot Status Total Issues: 78 2 UI-5193 Record deletion request Confluence pages Sprints 2023 Septembe 🛿 Actions 🗸 To Do 🐱 Waiting for support 37 UI-5192 Collections Search Crashing Causing WF Jobs to Fail on creation > 5 DSR-3 2023 Routine Tasks @ ♣ ⊘ ☑ ◎ … Details Open 24 2 UI-5188 Record deletion request > 5 DSR-16 2024 Routine Tasks Description Pending Assignee UI-5159 Article Type not showing Add a description... O Unassigned DSR-80 IIIF Universal Viewer - small scal... Waiting for customer 5 2 UI-5157 Data change request Assign to me Waiting for approval 1 DSR-85 Transition to Archive Cybersecu... 1-10 of 25 Child issues Order by × ··· + Labels 50% Done > 🖸 DSR-104 Confluence pages None 1 minute ago 1 minute ago SR-32 Confluence Page (Admin) Re... DONE V + Create Epic Start date None SR IN TESTING / REVIEW V Due date None Development SR DONE - SR DONE - SR DONE -**1** Create branch Create commit ☑ DSR-36 Confluence Page Prep... 📚 SR IN PROGRESS ▼ Releases + Add deploymen -Reporter

Digital Preservation Systems

8

Today Weeks Months Quarters

Add a comment..

JIRA Dashboard and Projects

Axiell Collections

					Action			Clipboard	Occurrence	e .	Fea	tures	
Record details		Ľ	C)	6	6 🛛	+	2	(°) 🔻	=	53	15	~	0 ₀
-Identification-													
Number		C-1575	337		It	em typ)e		Digita	d			
Can ID	:	525075	5A		S	tatus			Viewi	ng			
Usage	!	<u>Digital f</u>	ile can	be acc	cessed on	reques	<u>st</u>						
Collection													
-Title													
Title article	Title						L	anguage		Туре			
	PRESTON	'S EMP	IRE D	ay ma'	Y 22ND 19	09							.
													-
н н т н										1 OCCU	RRENCES	O	7
-Conditions go	verning a	ccess	and	repro	duction								
Access condition		digital o require holders	d clear , contri olders l	ons ple ances ibutors have b	lational Aro ase ensuro from copyr or other een obtain	e the ight	Date						

Access conditions

ß

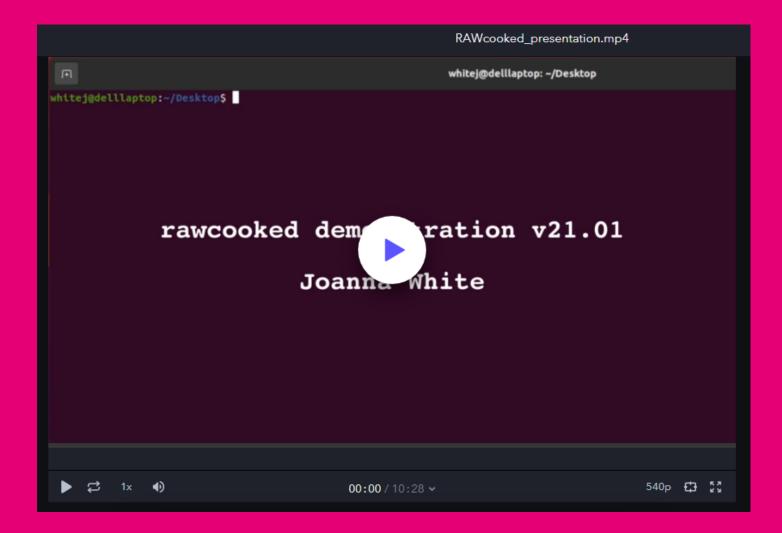
Help

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.

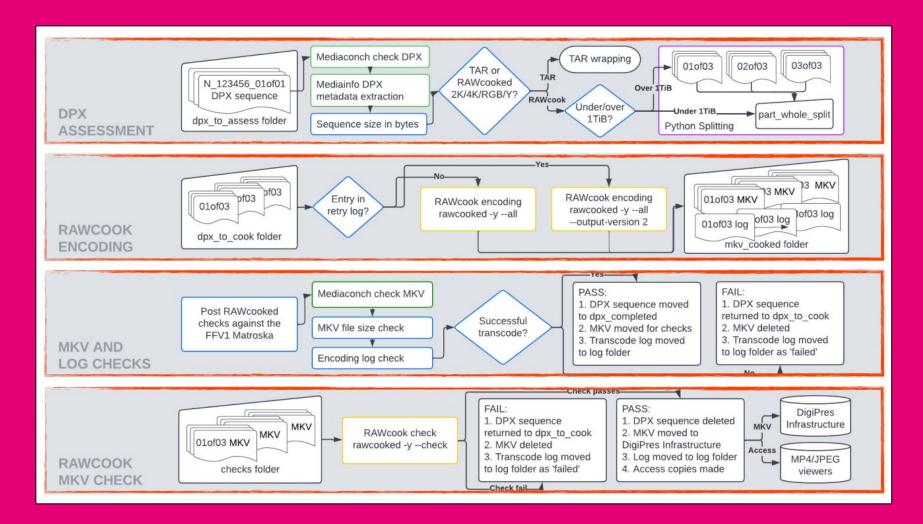
GitHub Repository

E bfidatadigipres/dpx_encoding Note: V Fox 2								
🗘 Code 💿 Issues 🏦 Pull requests 💿 Actions 🖽 Projects 💿 Security 🗠 Insights 🛢 995 KB								
	₽ main - ₽ 1 branch ⊙ 0 t	ags	Go to	file Code -	About			
	digitensions TAR wrapping upd	BFI National Archive automated dpx preservation scripts written in BASH and						
	digiops_isilon	TAR wrapping updates		2 months ago	Python for use with Media Area RAWcooked and other open source			
	film_operations	TAR wrapping updates		2 months ago	programmes.			
	grack_film	TAR wrapping updates		2 months ago	automation digital film matroska			
	qnap_digiops	TAR wrapping updates		2 months ago	preservation ffv1 dpx rawcooked			
	nap_film	TAR wrapping updates		2 months ago	Readme MIT license			
	qnap_filmops	TAR wrapping updates		2 months ago	☆ 10 stars			
	qnap_filmops2	TAR wrapping updates		2 months ago				
	LICENSE	New RAWcooked environment and new TAR script additions	1.06 KB	8 months ago	V 2 forks			
	RAWcooked_Cheat_Sheet.pdf	Updated version.	924.23 KB	9 months ago	Report repository			
	README.md	Update README.md	35.31 KB	2 months ago	Releases			
	flock_rebuild.sh	Flock lock updates for correct script starts	1.62 KB	8 months ago	No releases published			
	rawcooked_dpx_orientation.x	Small XML Mediaconch policy for image orientation	422 B	8 months ago				
	rawcooked_dpx_policy.xml	Updates	4.54 KB	2 months ago	Packages			
	rawcooked_mkv_policy.xml	Updates	3.79 KB	2 months ago	No packages published			
	i≡ README.md				Contributors 2			
	DPX preservation se	cripts			digitensions Joanna White			
	The BFI National Archive recently developed workflows using open source software RAWcooked to convert 2K and 4K DPX film scans into FFV1 Matroska video files for preservation. This has involved working with Media Area's Jérôme Martinez, developer of RAWcooked, to help test and refine features. This repository contains the RAWcooked encoding (and TAR preservation scripts) used for these DPX automation workflows. The aim of these scripts is to turn large DPX image sequences into RAWcooked FFV1 Matroska files for preservation within the BFI's Digital Preservation Infrastructure (DPI). Encoding DPX sequences to FFV1 can reduce the overall file size by half (2K RGB files), and allow the DPX image sequence to be played in VLC or similar software for instant review. These scripts are available under the MIT licence. They have been recently redeveloped and may contain some untested features within the toest. How works to test these yoursel playes create a safe environment to use this code							
	separate from preservation critical files. All comments and feedback welcome.							
	scheduling. As a result there is information about script funct These scripts handle the comp suitability for RAWcokked en dees not meet the basic DPX and passed to a TAR wrap pre files back to a DPX sequence. RAWcokked encoding function and the second assesses the r the error type moving failed fi script to try gain with a differ	ython scripts are not designed to be run from the comm is no built in help command, so please refer to this READI ionality. Jete encoding process from start to finish. including asse dding, splitting of sequences too large for ingest into ou and clean up of completed processes with deletion of DP Mediaconch policy requirements for RAWcooked encodi servation path. These scripts also manage the demuxing and the unpacking to Python tarfile wrapped TAR DPX st and the unpacking to Python tarfile wrapped TAR DPX st is used to these encoding attempts. If an encoding fails d les to a seprate folder, and create a new list which allows ent encoding formula, using 'output-version 2'. If it fail folder in need of human intervention.	ME and the script of essment of the DPX ir DPI, encoding, fai X sequences. If a D ng then the sequen of RAWcooked FFV equences. he RAWcooked enco px_post_rawcook.sls the RAWcooked fi	omments for sequences ilure PX sequence ice is failed (1 Matroska oding path h will assess rst encoding				

Videos



Lucid Chart



DPC – Documentation Digital Preservation: A Workshop

CID and DPI Systems / Service Deak / Autoingest - Service Overview		☆ ⊙ ใ∎ Share						
Autoingest - Service Overview		Сору	e actions		≝ 🛛	& ₽ @ © ₩ ₩	+~	
		Move Export	>	nowledge Base	Q search	່h ຝ En	iter	
Autoingest automates the verification, ingest, fixity and deletion of media from approved ingest locations About 		Archive				ode snippet	•••	
Validation Filenames must be correctly formatted		Delete			Dis	splay code with syntax highlighting		l
 Media must be fully described in CID Media must be valid Ingest 		Presenter mode	r	t this how-to	8	fo panel ghlight information in a coloured pane	el	a
Persistence Visualisation Related articles		Analytics		ow to set up				зb
(Change page owner				uote ert a quote or citation	>	l
To learn how to use Autoingest, please read the Autoingest - User Guide.		Page history		ictions		ecision	<>	
About		Attachments	>	p-by-step gı	51-	pture decisions so they're easy to trac		
Autoingest is a digital preservation workflow service. It automates the validation, ingest, persistence and generate access proxies. It can handle files individually or in bulk.		Resolved comments	0	s that are sin trations to in	Div	vider		l
Autoingest connects BFI Digital Preservation Infrastructure systems, processes and hardware together to As the Autoingest service is always running, media is continuously queued for ingest and deletion. Media – is left in place, and will be re-queued and reprocessed repeatedly. Successfully ingested and persisted f		Advanced details	، س	8-5 steps per		parate content with a horizontal line		
Autoingest records all media validation, ingest status and persistence status messages in a log file.	View Source	Slack Notifications	Ŭ		.		ſ	Ŀ
Media has to be documented in CID before it can be processed by Autoingest.	view source	Slack Noullications		ght importar		pand ert an expand	-	hi
Every file proceeds through the following three stages before it is deleted: validation, ingest and persister	Hide related pages			options in th	•		•	
Validation Media must pass these three sets of validation tests before it is ingested:	View Storage Format				••• Vie	ew more		
means more pass tress times sets of variable in tests before it is ingested.				ad article	-			

Examples

CID and DPI Systems / Service Deak / Autoingest - User Guide

Error Log

The error log is created at 8am every morning. The log is called current_errors and it is located in the top level of the autoingest folder tree, in all operational shares in Isilon - 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 Isilon share where you work - ie 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 Autoingest.

Error Message	Meaning	Action
Cannot parse partWhole from filename	 This indicates poorly formed partWhole statement. Some common examples: missing underscore after Object Number missing leading zero before a number below 10 - 1of6 instead of 01of06 presence of extra text such as 'Reel1' or 'R1' - R1ofR6 instead of 01of06 presence of trailing version increment - 01of01-2, 01of01-3, etc part greater than whole - 05of03, 10of01, etc first part is 00, and not 01 	Change filename to use correct partWhole 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 0001of1900. Also make sure to use a leading zero for a number below 10, and no extra characters such as Reel, R, Part, Pt. Also make sure the part is not greater than the whole - ie 04of05 and 05of05 are fine, but 05of04 will fail. Please make sure the 'of' is lowercase letters with no gaps between.
Extension does not match <file_type> in record</file_type>	This indicates a mismatch between the file_type in the CID record and the file extension. Some common examples: CID Item states MP4, file extension is .mov CID record states JPG, file extension is .tif	In some cases, this error may indicate incorrect Object Number use in filename, so that should be checked first. Most commonly, the CID record must be changed, to document the file type accurately.
ffprobe failed to read file: [1] status or ffprobe failed to read file: [None] status	 This indicates a serious problem with the file's metadata headers, 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 file's usability 	In most cases this indicates a poorly encoded file, which must be re-encoded or re-acquired, as it cannot be ingested to DPI. Use MediaInfo to examine the file's metadata, and re-encode / re-acquire as appropriate.

ueued for ingest. Files which are successfully ingested are persisted as exact copies; additionally, access proxies

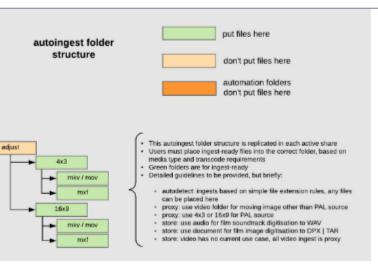
*

ws are supported. Files may simply be stored; or stored with an additional proxy made available; or stored with an

cks contains an autoingest directory structure. The preservation workflow used by Autoingest is no longer sest directory structure.

the video/adjust or proxy/image subfolders, viewing proxies for video or image files will be made successfully However, existing automated workflows still use different ingest folder paths so their structure is being retained for to continue their existing ingest practises, or simplifying ingest by using the autodetect folder.

e files. The directory tree is mapped to autoingest by a configuration file. The tree is organised hierarchically and options for storing, transcoding and adjusting media.



Examples

Admin Knowledge Base

- Heritage 2022 Processes
 - Adding Imagen 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
 - BFI file checking sampling (MN WIP)
 - BFI Supplier Invoices
 - CID manual data input processes for H22 BFI Film Digitisation BFIPlayer titles
 - Delivering Files for BFI Replay (WIP)
 - Exceptions Editing CID Item records
 - File Workflow
 - Film Digitisation Workflow and Processes
 - Framework Supplier allocation spreadsheet
 - · Google Sheets Data gathering for Google Sheets
 - · Items digitised through shared container
 - Logging WIP MN
 - MediaConch FSN File Validation with MediaConch WIP
 - MediaConch Failures 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 filesize for multiple HDDs:
 - CID import data
 - Copying files to HDD with 7-Zip Encryption (WIP)
 - Generating RNA filelist with checksums: (WIP MN)
 - · Rights mp4 requests download, rewrap and supply
 - RNA file requests download and supply
 - Splitting WARNING / INFO Reports
 - Troubleshooting Processing; Segmentation; Autoingest
 - Updating the Google Data Studio H22 Dashboard

When do you create documentation and how often is it reviewed and updated?



Inception (2010), dir. Christoper Nolan

What is the update process and how do you manage versions?



What is next for digital preservation documentation at your organization?



prime video

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