Digital Preservation in Disruptive Times

19th International Conference ■ Champaign-Urbana, Illinois ■ September 19–22, 2023

Creating an End-to-End Process for Implementing a Digital Archiving Workflow: How we are putting theory into practice

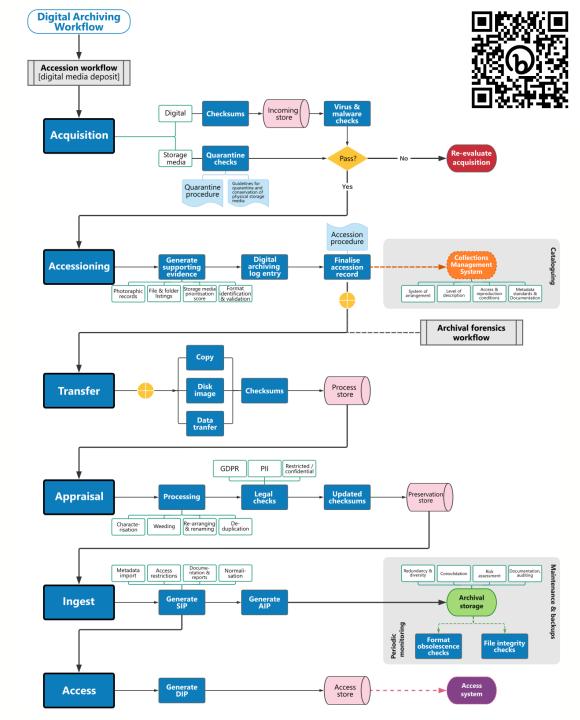
Leo Konstantelos & Emma Yan

Thu 21st Sep 2023, 13:30-15:00



Pilot project ran from Dec 2022 – Sep 2023 to produce an end-to-end process for implementing our digital archiving workflow.

Develop, test and evaluate the necessary framework for delivering robust digital preservation and digital archiving services.





Proposal submission gives more detail on what we intended to achieve



http://bit.ly/3LmEbYZ



Key areas



Develop policies, workflows, procedures and tools for digital archiving



Review existing ASC policies and donor agreements

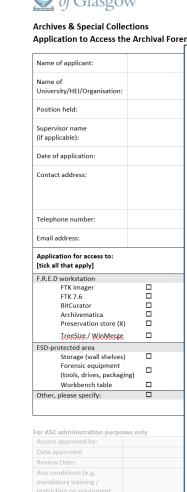


Update digital asset register, create risk register



Outcomes -**Archival Forensics Lab access policy**

- Policy provides conditions of access for individuals seeking access to the Archival Forensics Lab (AFL).
- Follows and/or complies with the specifications defined in:
 - Interpol Global Guidelines For **Digital Forensics Laboratories**
 - UK Gov Forensic science providers: codes of practice and conduct, 2021
- AFL access agreement
- AFL access application form







Archives & Special Collections

plication to Access the A	rchival For	ensics Lab (AFL)		Archival Forensics Lab acces
me of applicant:		6	University of Glasgow	
iversity/HEI/Organisation:			of Glasgow	
sition held:		Archiv	ves & Special Collections	
pervisor name applicable):		Archiv	val Forensics Lab (AFL) Agreement	
te of application:		Audier This agr		ons staff; and for any other party (staff member
ntact address:		or visito		ival Forensics Lab to process collections using the
		Purpos This agr		onsibilities of those who have access to the AFL
ephone number:		-	ures that all parties concerned understand t	
ail address:				
plication for access to:		AFL ag	reement	
k all that apply]		For all A	ASC Staff, Facilities Management staff and \	isitors with access to the AFL:
.E.D workstation FTK Imager FTK 7.6 BitCurator Archivematica Preservation store (X)	0	:	I have read and understood the Archival Fo I will follow the conditions of access to the I understand that any breach of the policy v Staff and Visitors who have been granted	AFL as per the AFL Access Policy. vill have appropriate action taken.
TreeSize / WinMerge		•	I will follow the guidelines and procedures f	or the use of equipment, machines, and not that it is my responsibility to ask for further
O-protected area Storage (wall shelves)			• • • • • • • • • • • • • • • • • • • •	re of any aspect of the use of the AFL and its
Forensic equipment (tools, drives, packaging)			raciities.	
Workbench table				
ner, please specify:		Signed:		
ASC administration purposes o	nly	Name:	(IN BLOCK CAPITALS)	
cess approved by:				
te approved:				
view Date:				
y conditions (e.g.		Date:		
ndatory training /		1		

Version: 1.0. 25/05/2023

of Archives & Special Collections (ASC) and

licies and procedures cies\SAFETY.DOC

icies\Access Policy.doo ations defined in

of practice and conduct, 2021

ng and temporary storage

and material processed in the lab necessitates

luals seeking access to the Archival Forensics Lab

Relations & Digital Preservation team authorised

yed by Facilities Management team.

ho visits AFL premises for any reason.

e the AFL space and/or facilities

s a Visitor during their time on AFL premises.



Outcomes – Digital holdings

 Updated a previous digital scoping survey to build a digital asset register

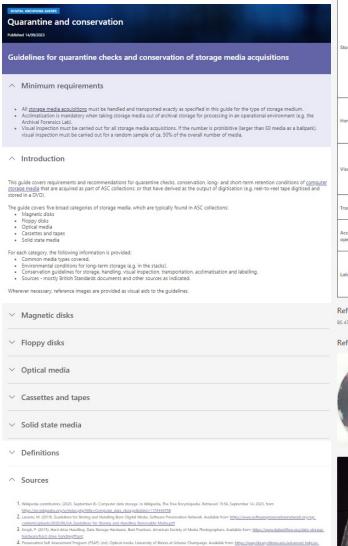
 Need to think about integration with our collection management system, particularly around location logging





Outcomes – Quarantine and conservation

- Guidelines for quarantine checks and conservation of computer storage media
- Covers magnetic disks, floppy disks, optical media, cassettes and tapes, solid state media
- Based on British
 Standards, community
 resources and own
 expertise
- Will be released under CC Licence Q1 2024



5, Brown, A. (2008). Care, Handling and Storage of Removable media. Digital Preservation Guidance Note: 3. The National Archives. Available 1

Optical media

Media

CD-ROM, CD-R, CD-RW, Audio CD, DVD-ROM, DVD-RAM, DVD-RW, Blue-ray Disc, MiniDisc

Long-term environmental conditions

Annual average temperature (°C)	20		
Minimum average temperature (°C)	18		
Maximum average temperature (°C)	22		
Relative Humidity (%)	35-45		

Conservation

Storage	No rapid changes in temperature, humidity or both. Dust- and moisture-free storage and operational areas, away from dust-producing peripherals and moisture-emitting equipment. Store in dark-environment. No exposure to direct sunlight, high heat and sources of ultraviolet light. Avoid environments that produce acidic and oxidizing gases. Remove dist. Foreign material, fleoperprints, stundeges, and flequids by wiping with a clean cotton fabric in a straight line from the centre of the disc toward the edge. Les CD/DVD-cleaning detergent, isopropyl alcohol, or methanol to carefully remove stubborn dirt or material. Avoid original packaging for optical media, especially DVDs, which can be less than desirable as the hubs may be too large or require excessive pressure to be applied in order to remove the disc. Always store vertically, in an upright position. Store in inerty plastic or seted containers with a non-damaging centre hub, if the disc is to be kept long-term. Store in archival quality sleeves in all other cases. Store on shelves within metal, fire resistant cabinets.				
Handling	Handle discs only by the outer edge or the centre hole. Avoid dirt or other foreign matter from touching or sticking to disc, including fingerprints from touching the disc's surfaces. If caddies are used, repeated loading and extention of disks should be avoided to minimize disc damage. Do not allow contact with liquids, dust or snoke. Remove from drives that have been switched off and should be removed from drives that are inactive for long periods. Follow Storage guidelines where suitable while handling.				
Visual inspection	Inspect for surface scratches, gouges, smudges, dirt, and dust that can inhibit playback of the disc. If gouges or scratches are deep enough, they may cause permanent damage to the disc and should be reported. Inspect the container and the disc for mould. Mould on the container is a good indicator that the mould has travelled to or from the disc. Inspect for "disc rot" - oxidisation of the aluminum recording surface, observable indications of disc rot are pin-sized holes in the reflective layer, most visible when viewed against a light source, a bronze discoloration of the disc, and crazing distortion in the reflective appearance. Crazing on the seem is smorthless or fractal-like milky-white patterns on the disc.				
Transportation	Keep disks in protective containers as prepared for Storage. Maintain in cool, day conditions and do not allow condemation to occur on the disc surface. Follow acclimatation and handling guidelines when uppacking at destination.				
Acclimatisation in operational environment	Ideally acclimatise for 24h before use. Minimum acclimatisation of 6h within protective envelope or case. Mandatory 24h acclimatisation for discs that have been stored in environmental conditions outside recommendation above, to avoid condensation forming on disk surfaces.				
Labelling	Adhesive labelling kits must never be used, since these can disturb the mass balance of the disk and damage the data layer. If adhesive labels are present, do not try to peel off or reposition. Discs may be marked on the upper surface using a soft tipped pen with water-soluble, permanent ink. Do not use pencil, pen, fine-tip marker or markers that contain solvents to write on the disc. Never write on the laser-reading used or optical discs. Labels can be safely applied to caddies, sleeves or containers used to store optical discs.				

eference

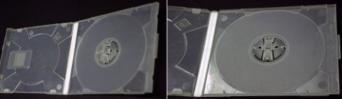
BS 4783: Part 3-1988, [2], [4], [5]

Reference images





Examples of disk rot on the recording and reading surfaces of a disc [4]



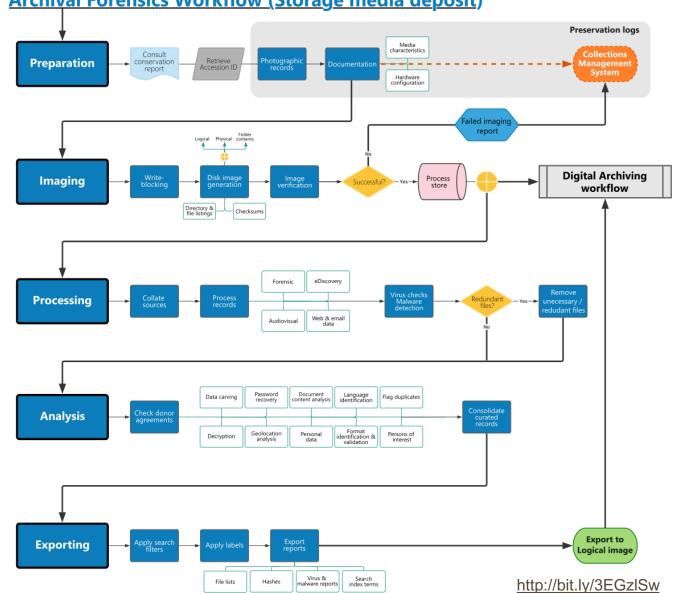
inert plastic container with easy-release hub [4]



Outcomes – archival forensics workflow

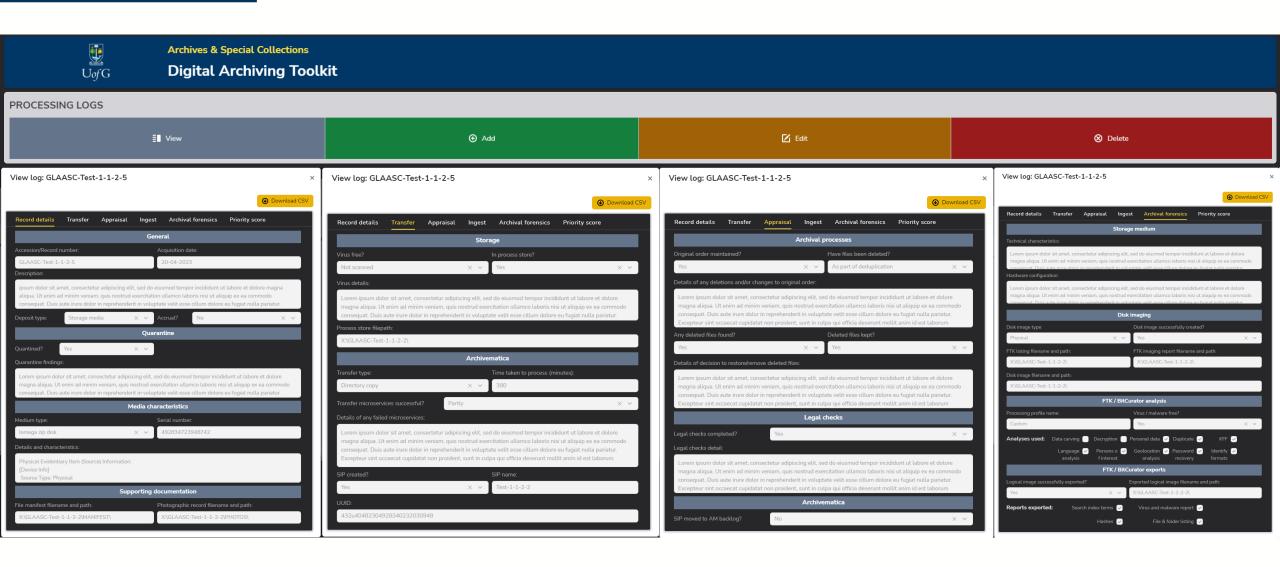


<u>Archival Forensics Workflow (Storage media deposit)</u>





Outcomes – Digital archiving logs





Outcomes – Risk register

47 risks identified across 9 root causes:

- Management
- Legal liabilities
- Operations
- Staffing
- Technical infrastructure
- Security
- Acquisition and ingest
- Preservation and archival storage
- Metadata

Based on:

- Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)
- Risk assessment handbook and self-assessment tool (TNA)
- Managing digital continuity (TNA)
- Digital Preservation Handbook, Risk and change management (DPC)

Risk Ref	Title	Date Raised	Risk Owner	Root Cause	Risk Description	Impact	Category	Initial Impact	Initial Probability	Initial Assessment	Mitigation Type	Mitigation Posture	Mitigating Actions
	Conservation of physical storage media	01/04/2022	(DPWG)	Acquisition & Ingest	Deposited physical storage media (e.g., hard drives) puller from concervation issues, such as pest infestiation and mould.	Conservation issues may compromise the accessibility of deposited physical storage media.	Service Delivery and Operations	3	3	9	Control	Reduce	Develop conservation checks for deposited physical storage media, including quantamine physical storage media, including quantamine 2. Develop conservation practice and documentation for optimal stronge conditions of physical storage products. 3. Manyarian procedures for the timely check of submitted medica, and for determining action with damaged media (disposal, resubmission request).
DP33	Transfer and handing issues with physical storage media	01/04/2022	(DPWG)		Deposted physical storage media are damaged or altered during standers or due to mappropriate handling.	Transfer and handling issues may compromise the accessibility of deposited physical storage media.	Service Delivery and Operations	3	3	9	Control	Reduce	Develop guidelines for transfer and handling of stage media for transportidelivent services to follow. Develop procedures and guidance for depositors, a to properly labely packages as fingles. Mannain procedures for the timely check of submitted media, and for determining action with damaged media (disposa), resubmission request).
DP34	Inability to validate effectiveness of ingest process	01/04/2022	(DPWG)	Acquisition & Ingest	hability to validate effectiveness of ingest process for deposited digital records, and demonstrate that their integrity and authenticity were maintained during ingest.	Between the points of receipt and ingest of digital materials, the received package is subjected to inadvertent changes or comption, which is not monitored or checked, thus rendering the materials inaccessible.	Service Delivery and Operations	3	3	9	Control	Reduce	Create checksums throughout the submission and ingest processes; and compare at the time of ingest to ensure integrity and authentitory of ingested records.
DP35	Loss of confidentiality	01/04/2022	(DPWG)	Preservation & Archival storage	Preserved records protected by conflictershifty agreements are made available to communities, in contravention of those agreements.	The University becomes involved in legal proceedings due to legal labilities resulting from digital preservation activities.	Legal & Governance	5	2	10	Control	Reduce	Ensure software and hardware systems and preservation strategies are capable of meeting confidentiality requirements. Ensure policies and procedures cater for confidentiality requirements. Develop action pathways for dealing with loss of confidetiality cases.
DP36	Loss of availability	01/04/2022	(DPWG)	Preservation & Archival storage	Preserved records cannot be accessed or do not render correctly.	Digital preservation services are unable to provide access to preserved digital records, for which access ought to be available.	Service Delivery and Operations	4	2	8	Control	Resolve	Ensure software and hardware systems and preservation strategies are capable of meeting access requirements. Ensure policies and procedures cater for access requirements. Develop access pathways for preserved digital records.
DP37	Loss of authenticity	01/04/2022	(DPWG)		hability to demonstrate that preserved records are what they purport to be.	The credibility of digital preservation services, as well as the trust credibility of digital preserved records, we understand Potential loss of secondarian.	Service Delivery and Operations	4	2	8	Control	Reduce	I Enzure software and hardware systems and preservation strategies are capable of meeting authernition typerunermin. 2. Mannain and review policies and procedures to ensure adequate recording of provinance information to demonstrate that archived material represent authentic represervation of what was initially deposited or received.
DP38	Loss of integrity	01/04/2022	(DPWG)	Preservation & Archival storage		of Preservation services are incapable of demonstrating that the integrity of information has been maintained since its receipt, and that what is stored corresponds exactly with what is as originally received.	Service Delivery and Operations	4	2	8	Control	Reduce	Ensure software and hardware systems and preservation strategies are capable of meeting integrity requirements. Ensure palcies and procedures careful for integrity requirements. S. financial recording conductors and procedures to ensure adequate recording and companion of checksums to demonstrate that architect





Evidence-based approach to delivering digital preservation and digital archiving services.



Implement an end-to-end digital archiving case study using one of our collections to test workflows, explore benefits of archival forensics.



Highlight areas where improvements and further investment are required.



Description	Dance House Glasgow was a creative arts organisation involved in supporting the city's professional dance sector and offering community development programmes for over 20 years. In 2018, it lost its Creative Scotland funding and ceased operating.	DANCE HOUSE
Variability	The whole collection consists of records relating to Dance House Glasgow dating from c.1990 to 2018. It includes governance, financial, staff and project records, along with photographs, audio and video, press cuttings, and promotional material.	GLASGOW
Volume (digital)	6.3TB	
Representative	Hybrid – both born digital and paper records.	
ness	Digital records across three HDDs and 90 optical media discs.	
	At least one of the hard drives and some of the CD-ROMs not functioning, as per a survey of the material in 2021.	
Legal issues	The records came to UofG ASC via the Business Archives Surveying Officer and were gifted to us in 2019. As a creative arts collection, we expected issues around IPR and rights for the music used in the collection. As a business collection, we expected issues around personal information, confidential records etc.	















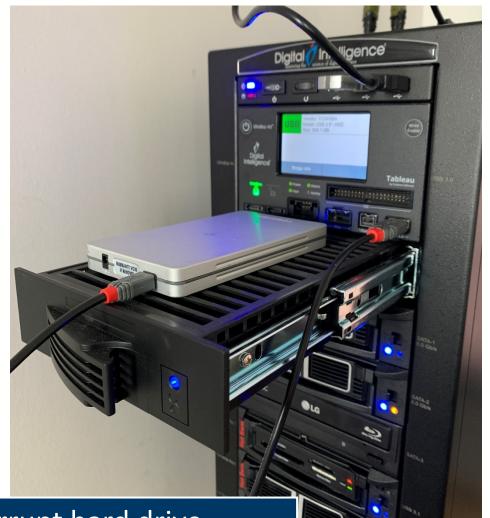


Dance House Glasgow collection: computer storage media





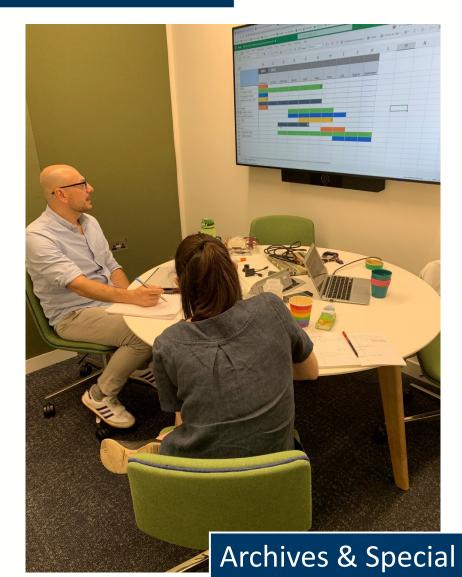


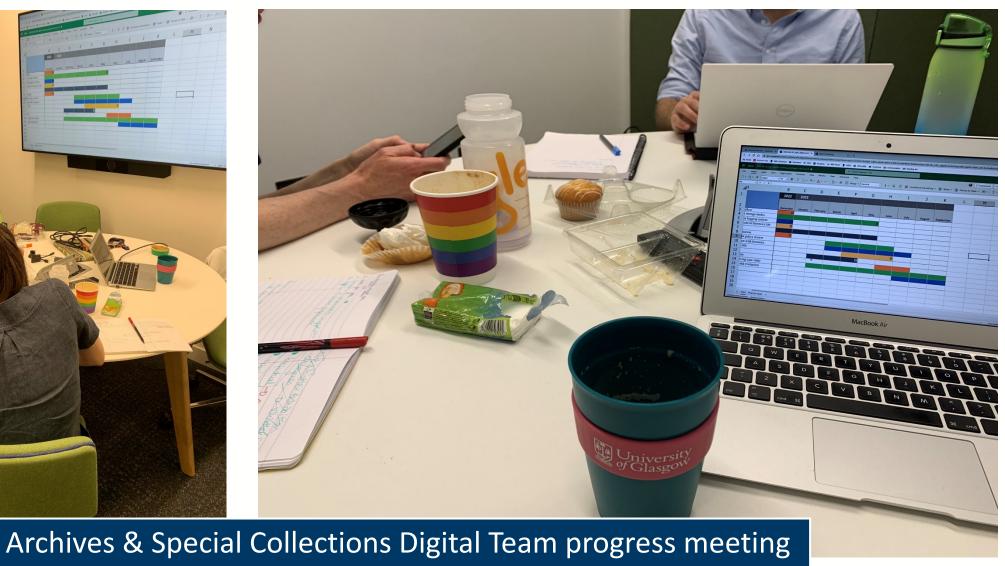


F.R.E.D workstation imaging a corrupt hard drive























Examples of born-digital items in the collection



Forensic analysis – FTK and BitCurator:

- Data carving
- Decryption and password recovery
- Personal data analysis (Bulk extractor)
- Document content analysis
- Format identification and validation (FTK, Droid, JHOVE)
- Duplicate flagging
- Index search in FTK

FTK – Weeding:

- Duplicates
- Deleted files
- Free space / file slack
- Recycle bin items
- OS / system files

		,		
	Pre-appraisal	Post appraisal	Difference	Notes
Total volume (GB)	6363	423	-5941	Total volume reduced by 5.94TB
Total file size	55936	17954	-37982	Total files reduced by 37,982
Selected file categ	ories:			
Archives	1572	48	-1524	Difference includes deleted and temporary files
Databases	2	1	-1	Difference includes deleted and temporary files
Documents	1802	219	-1583	Difference includes deleted and temporary files
Email	2	2	0	
Executables	510	0	-510	
Graphics	24225	15278	-8947	Difference includes deleted and temporary files
Internet	35	35	0	
Multimedia	1456	979	-477	Difference includes deleted and temporary files
OS/File system	1081	0	-1081	
Presentations	2	2	0	
Slack/free space	20512	0	-20512	
Spreadsheets	12	11	-1	Difference includes deleted and temporary files
Selected file status	s:			
Deleted files	557	9	-557	
Duplicate items	14995	3607	-11388	
KFF Alert files	37	0	-37	
KFF Ignore files	1288	0	-1288	
From recycle bin	3060	0	-3060	



Outcomes - final thoughts

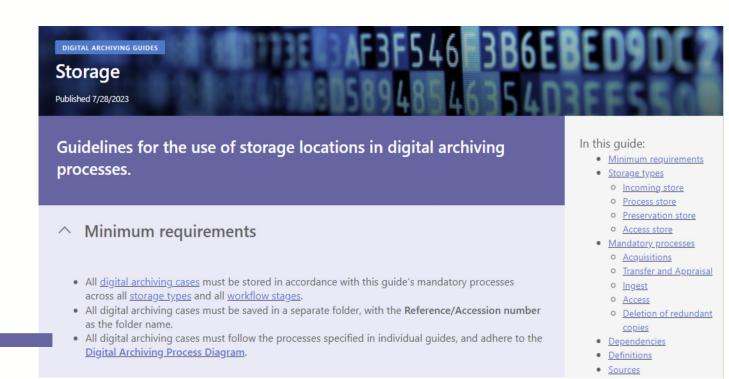
What we didn't complete:

- Donor agreement updates
- Collection Development policy update
- Research data end of life

Moved to planning for 2023/2024 academic year.

By the end of the pilot, we achieved what we needed to, plus more.

Digital archiving guides



Read about our work on the new DPC Digital Preservation Documentation Guide

