Building a digital preservation programme: the view from the bottom

ARCHIVES AND RECORDS ASSOCIATION CONFERENCE

AUGUST 2016
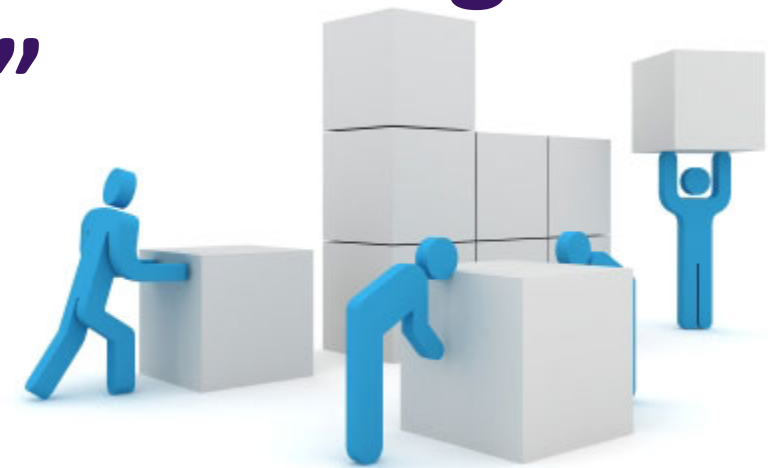
TIP 1: Don't feel you need to be an expert

THE BARRIER TO DIGITAL PRESERVATION: 'the unknown'

Unknown costs…unknown skills…unknown resources…unknown understanding…unknown value

"The only true wisdom is in knowing you know nothing"
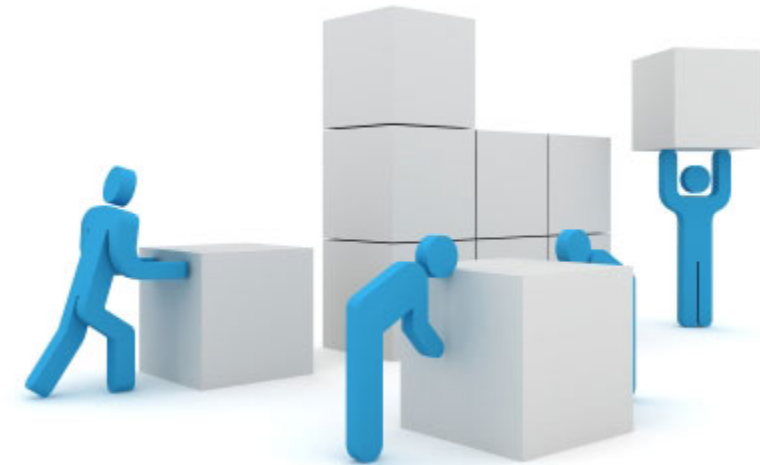
Socrates

TIP 2: Avoid information overload

There is a wealth of information online on digital preservation, digital archiving, digital asset management, open access, digital infrastructure, big data etc, etc!

❑ PLATTER planning tool (DPE)
❑ '*Practical Digital Preservation*' – Adrian Brown
❑ Levels of Digital Preservation (NDSA)
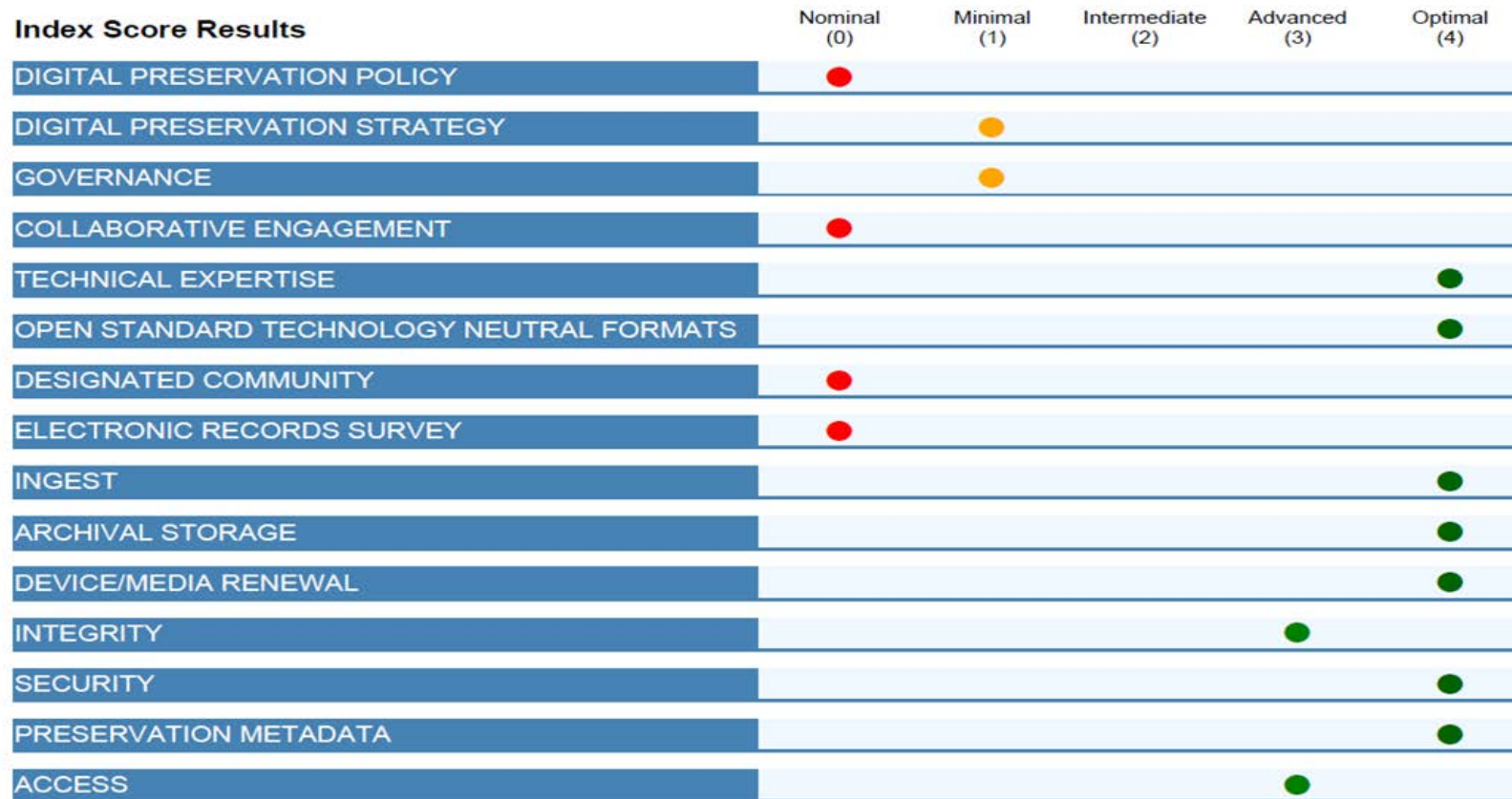❑ Digital Preservation Capability Maturity Model (DPCMM)

SOLUTION:
Be disciplined and stick to a few key resources to kick start your project that suit you and your organisation

TIP 3: Do understand the current situation

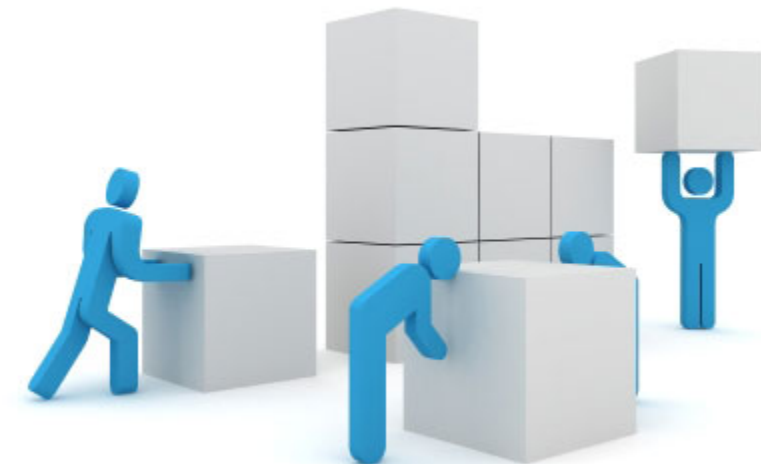| Index Score Results	Nominal (0)	Minimal (1)	Intermediate (2)	Advanced (3)	Optimal (4)
DIGITAL PRESERVATION POLICY	●				
DIGITAL PRESERVATION STRATEGY		●			
GOVERNANCE		●			
COLLABORATIVE ENGAGEMENT	●				
TECHNICAL EXPERTISE					●
OPEN STANDARD TECHNOLOGY NEUTRAL FORMATS					●
DESIGNATED COMMUNITY	●				
ELECTRONIC RECORDS SURVEY	●				
INGEST					●
ARCHIVAL STORAGE					●
DEVICE/MEDIA RENEWAL					●
INTEGRITY				●	
SECURITY					●
PRESERVATION METADATA					●
ACCESS				●	

DPCMM - http://www.securelyrooted.com/dpcmm/

TIP 4: Understand the objectives/deliverables

❖ What are the organisational goals/values/strategy

❖ What resources are available? Funding/staff

❖ Standards and best practice driven?

❖ Trusted accreditation?

❖ Advocacy and outreach (conference, social media)?

❖ Communication and training (workshops, 1-2-1, online)?

❖ Development of processes and workflows for appraisal, ingest, preservation and storage

❖ Sector awareness

TIP 5: PLAN, PLAN, PLAN!

Planning Tool for Trusted Electronic Repositories -
http://digital.library.unt.edu/ark:/67531/metadc799759/m2/1/high_res_d/platter.pdf

TIP 6: Assess your needs/requirements

Digital Strong Room - Requirements

Requirement	Essential	Desirable	Type	OAIS functional entity
Ability to restrict access to the original digital objects (only by archivists/administrators)	Yes		Operational	Access
Ability to integrate with archive catalogue system, ArchivesSpace	Yes		Technical	Access
Ability to send the access copy of a digital object to a storage location separate from the preservation copy	Yes		Technical	Access
Ability to apply different levels of access permissions to AIPs (open, closed, partially closed at object level, partially open at object level). Can apply PREMIS rights data at item level but this might only be feasible for items we want to restrict rather than adding rights metadata for all items.	Yes		Functional	Data manage...
Ability to make some DIPs only accessible by permitted individuals (application of EASE user authentication)	Yes		Functional	Data manage...
Ability to generate logs of access requests and failures	Yes		Functional	Administratio...
Ability to delete AIPs and/or digital objects within an AIP	Yes		Functional	Archival stor...
Ability to generate logs of any disposals from the system	Yes		Functional	Administratio...
Ability to generate error logs and flag those errors to Administrators? i.e if issues with users accessing objects, issues with ingesting objects	Yes		Functional	Administratio...

Recommended PREMIS metadata fields

ENTITY	SEMANTIC UNIT	SEMANTIC COMPONENT	SEMANTIC COMPONENT	SEMANTIC COMPONENT	CAPTURE?	RATIONALE	PRIORITY	ARCHIVEMATICA
OBJECT								
1.1	Object identifier							
1.1.1		objectIdentifierType			Yes	Mandatory. Controlled vocabulary would help	Essential	Yes
1.1.2		objectIdentifierValue			Yes	Mandatory. Each object should have a DOI or PUID	Essential	Yes
1.2	objectCategory							
1.3	preservationLevel							
1.3.1		preservationLevelValue			Yes	Mandatory. Bit level/Full. Controlled vocabulary would help.	Desirable	No
1.3.2		preservationLevelRole			Yes	To clarify the level of preservation above and whether it is a requirement to preserve at that level or whether the repository is capable or intends to preserve at that level.	Desirable	No
1.3.3		preservationLevelRationale			No			
1.3.4		preservationLevelDateAssigned			No			
1.4	significantProperties							
1.4.1		significantPropertiesType			No	Not mandatory. Insufficient resource to dedicate.		
1.4.2		significantPropertiesValue			No	Not mandatory. Insufficient resource to dedicate.		
1.4.3		significantPropertiesExtension			No	Not mandatory. Insufficient resource to dedicate.		
1.5	objectCharacteristics							
1.5.1		compositionLevel			Yes	Mandatory. To identify if encoded derivatives (compression/encryption) are present. A controlled vocabulary would help.	Desirable	No
1.5.2		fixity						
1.5.2.1			messageDigestAlgorithm		Yes	Mandatory. To identify the type of hash algorithm used. Controlled vocabulary would help.	Essential	Yes
1.5.2.2			messageDigest		Yes	Mandatory. The output of the hash algorithm.	Essential	Yes
1.5.2.3			messageDigestOriginator		Yes	Optional but could be populated either by the Event record (eventIdentifierValue) or agentIdentifierValue	Desirable	No
1.5.3		size			Yes	Must record the size of the file for storage analysis and file checking	Essential	Yes
1.5.4		format						
1.5.4.1			formatDesignation					
1.5.4.1.1				formatName	Yes	Mandatory. Should record in all instances the file format, even if unknown at point of ingest. Required for preservation planning (migration/tech watch)	Essential	Yes
1.5.4.1.2				formatVersion	Yes	Mandatory. Need to know the version of the format for migration purposes.	Essential	Yes
1.5.4.2			formatRegistry			This container could be repeated with multiple registries		
1.5.4.2.1				formatRegistryName	Yes	It would be useful to identify the registry used to profile the format (PRONOM/UDFR/GDFR)	Essential	Yes
1.5.4.2.2				formatRegistryKey	Yes	This should record the designator within the registry identified above (i.e fmt/412)	Essential	Yes

OBJECT | EVENT | AGENT | RIGHTS

TIP 7: Identify and review your options

TIP 8: Testing, testing, 1-2-3

		Collection size	Action type	Start time	Finish time	Total time	Compression	Pack size	Result
1	Archivematica testing								
2									
3	**Test no.** **Description**		**Action type**	**Start time**	**Finish time**	**Total time**	**Compression**	**Pack size**	**Result**
4	18 10 tiff files with full processing and normalisation for access, no checksum verification	1.35 gb	Transfer			25 mins			
5			SIP			25 mins	5	687 mb	For 10 tiffs it creates 480 files for AIP (almost all Bulk Extractor files!), however compress reduces 1.4 gb ingest to 680 mb for AIP
6					**TOTAL**	**50 mins**			
7	19 10 tiff files with minimal processing, no normalisation, no transcription, no forensics, format ID on ingest only, no structure report, no checksum verification	1.35 gb	Into transfer	09:10:00	09:15:00	5 mins			
8			Transfer	09:15:00	09:27:00	12 mins			
9			SIP	09:27:00	10:03:00	36 mins	1	734 mb	AIP Compression took 28 mins. 10 tiffs producing 18 files for AIP
10					**TOTAL**	**48 mins**			
11	20 10 tiff files with minimal processing, no normalisation, no transcription, no forensics, format ID on ingest only, no structure report, no checksum verification	1.35 gb	Into transfer	11:22:00	11:26:00	4 mins			
12			Transfer	11:26:00	11:30:00	4 mins			
13			SIP	11:30:00	11:57:00	27 mins	9	734 mb	Compression and verification take a while complete
14					**TOTAL**	**35 mins**			
15	21 10 tiff files with minimal processing, no normalisation, no transcription, no forensics, format ID on ingest only, no structure report, no checksum verification	1.35 gb	Into transfer	12:03:00	12:04:00	1 min			
16			Transfer	12:04:00	12:08:00	4 mins			
17			SIP	12:08:00	12:23:00	15 mins	5	686 mb	
18					**TOTAL**	**20 mins**			
19	22 10 tiff files with minimal processing, no normalisation, no transcription, no forensics, format ID on ingest only, no structure report, no checksum verificationRe-test (Test 19 for speed	1.35 gb	Into transfer	12:32:00	12:34:00	2 mins			
20			Transfer	12:34:00	12:37:00	3 mins			
21			SIP	12:37:00	12:49:00	12 mins	1	734 mb	
22					**TOTAL**	**17 mins**			
	23 10 tiff files with minimal processing, no normalisation, no transcription, no forensics, format ID on ingest only, no structure report, no	1.35 gb	Into transfer	12:56:00	12:59:00	3 mins			

Sheet1   +
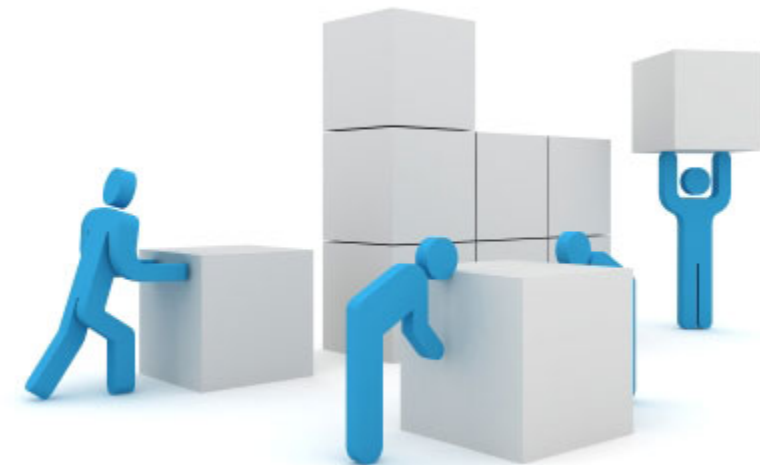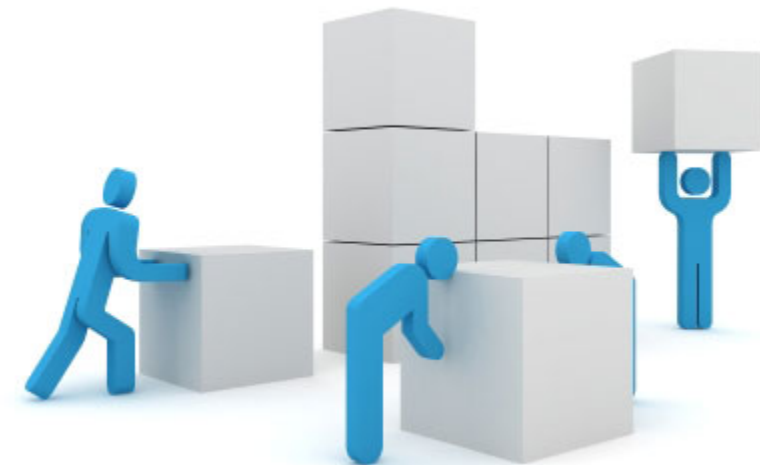
TIP 9: Develop workflows and processes
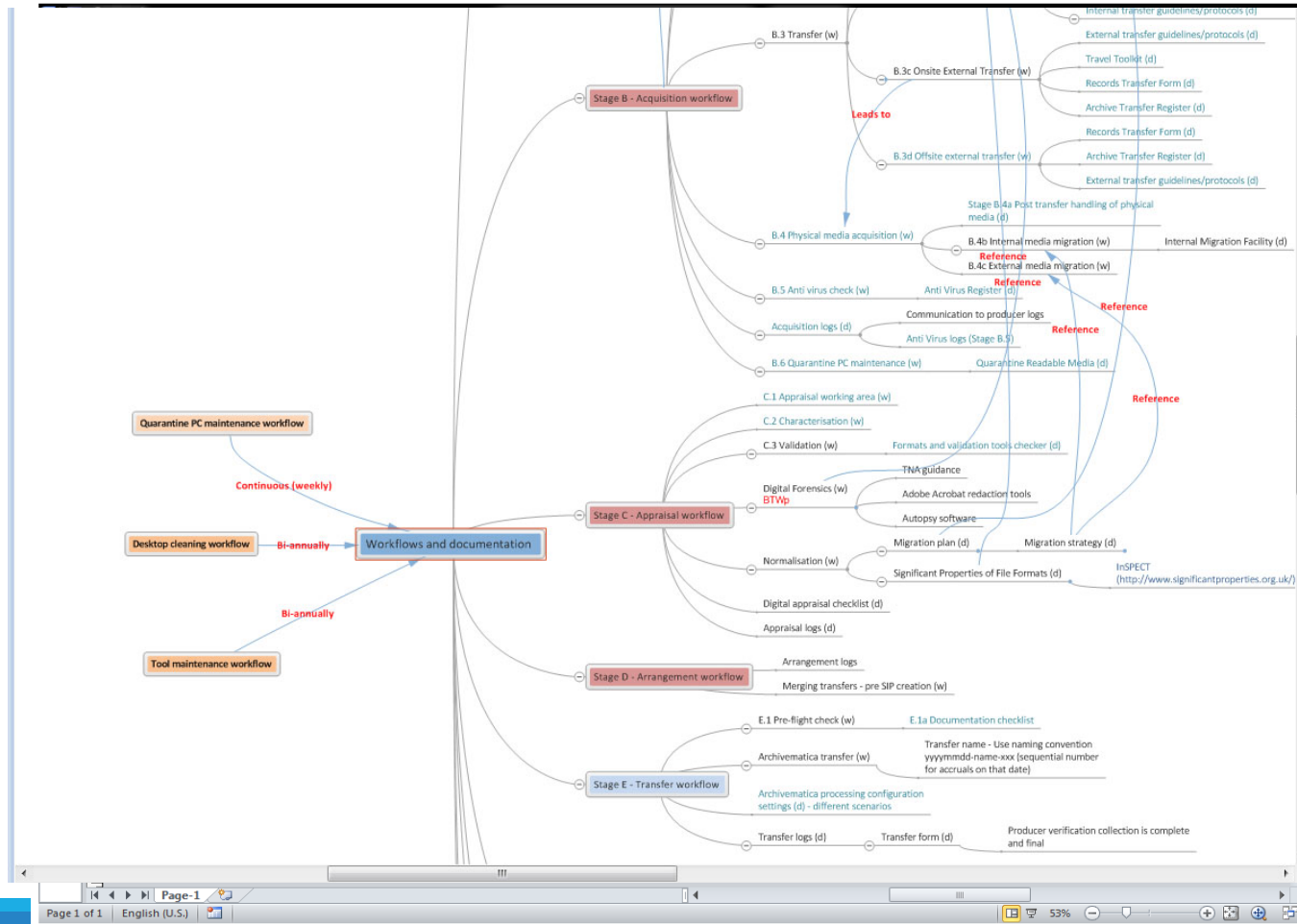
- ❑ Microsoft Visio (Edraw)
- ❑ MindGenius (Coggle, MindMapple)

TIP 10: Tools!

Acquisition/Transfer:

Exactly, Rich Copy, USB write blocker,ExactFile

Appraisal:

TreeSize Pro, Quick View Plus, Media Info, Inkscape, Ghostscript, VLC media

Digital Forensics:
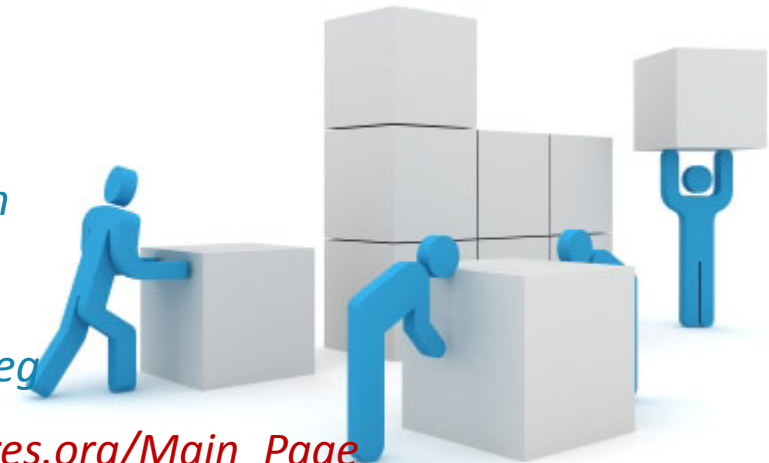
Bulk Extractor, Autopsy

Characterisation/Validation:

DROID, Jhove, VeraPDF, Jpylyzer, MDQC, MediaConch

Migration:

File Merlin, ImageMagick, Inkscape, Ghostscript, ffmpeg

COPTR Community Owned Digital Preservation Tool Registry = http://coptr.digipres.org/Main_Page

Thank you and goodbye!

Kirsty Lee

Digital Preservation Curator

kirsty.lee@ed.ac.uk

#thebitgardener

libraryblogs.is.ed.ac.uk/bitsandpieces