

Building a digital preservation programme: the view from the bottom

ARCHIVES AND RECORDS ASSOCIATION CONFERENCE

AUGUST 2016



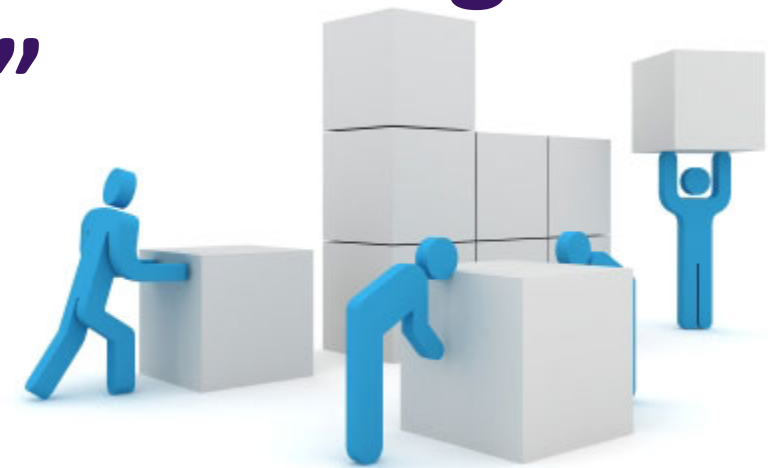
TIP 1: Don't feel you need to be an expert

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

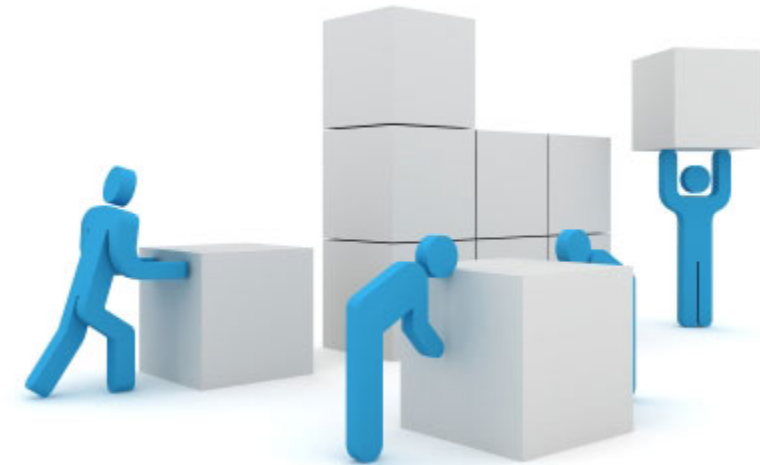
There is a wealth of information online on digital preservation, digital archiving, digital asset management, open access, digital infrastructure, big data etc, etc!



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

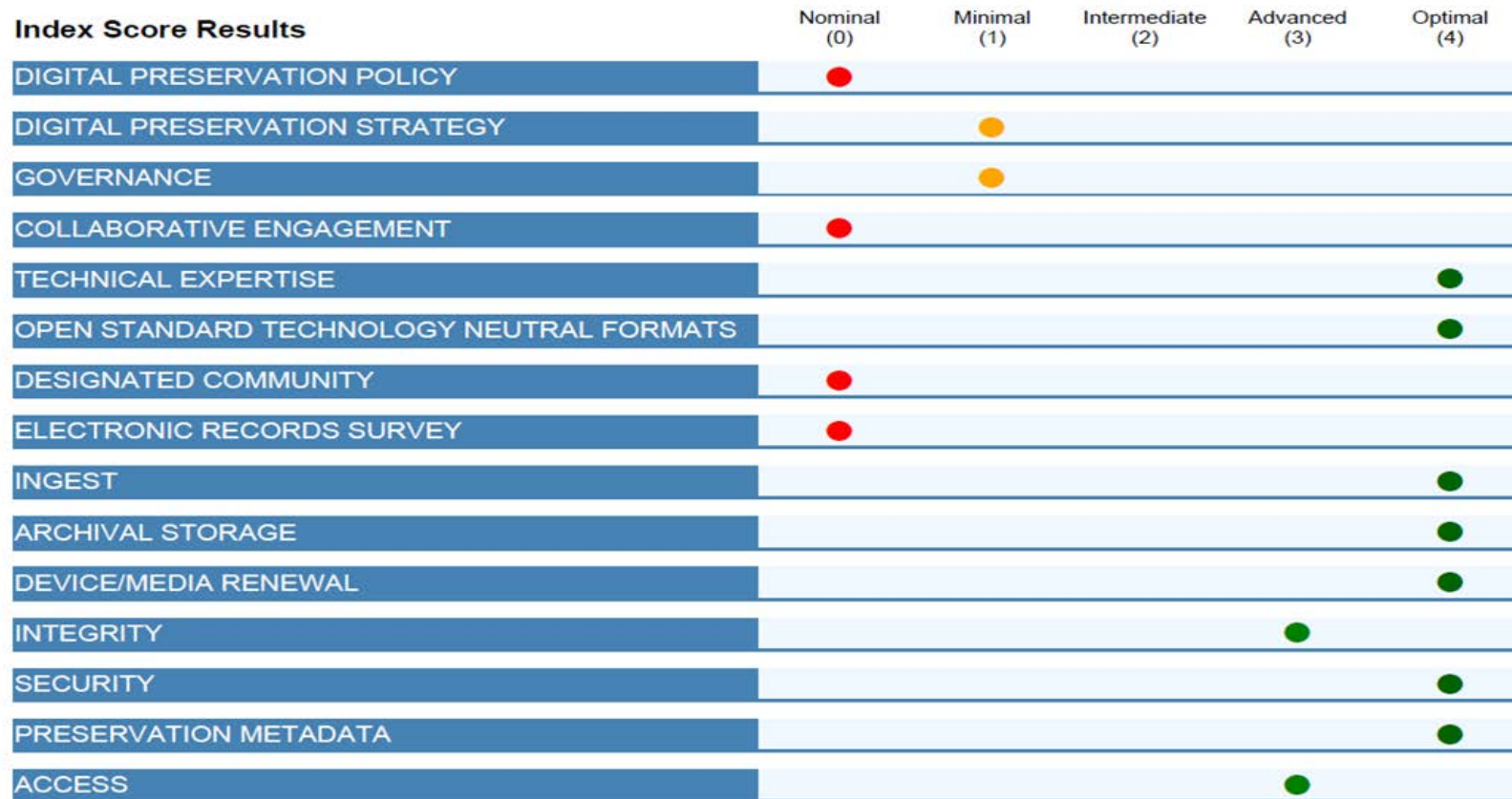
SOLUTION:

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



ted
ity

ers

15.
Access



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

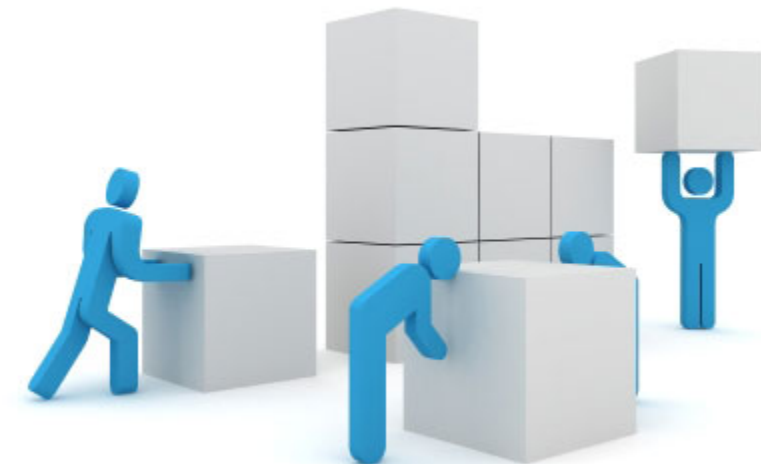
TIP 4: Understand the objectives/deliverables

- ❖ 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!

1	STRATEGIC OBJECTIVE PLANS (SOPs) - Access Plan									
2	Responsibilities: Management and relationship with end users									
3										
4	GOAL No.	Category	Title	Example	Deadline	Responsibility	Priority	Unit of measurement	Status	Resource
5	1.1	Designated community	Develop and maintain definitions of designated communities relevant to the repository	<i>Define designated communities and their needs</i>	Phase 3		Long term	Documentation	Uncompleted	
6	1.2	Designated community	Develop and maintain definitions of designated communities relevant to the repository	<i>Review needs of designated communities</i>	Phase 3		Long term	User survey	Uncompleted	
7	2.1	Legislative	Ensure appropriate restrictions are applied to digital records	<i>Define categories of access</i>	Phase 1		Short term	Documentation	Completed	Digital Preservation Policy V8
8	2.2	Legislative	Ensure appropriate restrictions are applied to digital records	<i>Adopt tools for handling data applicable to the different levels of access</i>	Phase 3		Medium term	Availability of tools	In progress	
9	2.2	Legislative	Ensure appropriate restrictions are applied to digital records	<i>Develop appropriate method and process for handling data applicable to the different levels of access</i>	Phase 3		Medium term	Documentation	Uncompleted	
10	3.1	Access	Ensure appropriate methods of access are in place to satisfy designated community needs	<i>Identify methods of access</i>	Phase 3		Medium term	Documentation	Uncompleted	
11	3.2	Access	Ensure appropriate methods of access are in place to satisfy designated community needs	<i>Identify and implement the technical requirements for access methods</i>	Phase 3		Medium term	Successful rendering of digital records	Uncompleted	
12	3.3	Access	Ensure appropriate methods of access are in place to satisfy designated community needs	<i>Review methods of access in line with designated community needs</i>	Phase 5		Long term	Documentation	Uncompleted	
13	3.4	Access	Ensure access copies can be perpetually accessed	<i>Develop a file format policy/register for storage</i>	Phase 1		Short term	Documentation	Completed	Stage A.4 File Format Policy Register
14	3.5	Access	Ensure access copies can be perpetually accessed	<i>Ensure file format policy is implemented</i>	Phase 1		Short term	100% of ingested records following the FFPR	In progress	Stage A.4 File Format Policy Register
15	3.6	Access	Ensure access copies can be perpetually accessed	<i>Review file format policy to ensure it remains up to date and relevant</i>	Phase 5		Long term	Documentation	Uncompleted	Stage A.4 File Format Policy Register
16	3.7	Access	Ensure access copies can be discovered and are discoverable	<i>Identify the descriptive metadata requirements to ensure appropriate data capture prior to ingest</i>	Phase 2		Short term	Documentation	Completed	Stage A.4 File Format Policy Register
17	4.1	Infrastructure	Ensure IT access infrastructure is appropriate and fit for purpose	<i>Identify requirements for access infrastructure and storage</i>	Phase 3		Medium term	Documentation	In progress	
18	4.2	Infrastructure	Ensure IT access infrastructure is appropriate and fit for purpose	<i>Identify tools for satisfaction of access requirements</i>	Phase 3		Medium term	Availability of tools	In progress	
19	4.3	Infrastructure	Ensure IT access infrastructure is appropriate and fit for purpose	<i>Implement tools for access</i>	Phase 3		Medium term	Availability of tools	Uncompleted	
20	4.4	Infrastructure	Ensure IT access infrastructure is appropriate and fit for purpose	<i>Review tools for access</i>	Phase 5		Long term	Documentation	Uncompleted	
21	4.5	Infrastructure	Ensure digital data storage is appropriate and fit for purpose	<i>Identify initial storage requirements for access copies</i>	Phase 1		Short term	Documentation	In progress	
22	4.6	Infrastructure	Ensure digital data storage is appropriate and fit for purpose	<i>Develop plan for upgrade to storage service (media and capacity)</i>	Phase 3		Medium term	Documentation	Uncompleted	
23	4.7	Infrastructure	Ensure digital data storage is appropriate and fit for purpose	<i>Clarify the level of service from IT for the storage service (maintenance and upgrade)</i>	Phase 3		Medium term	Documentation	Uncompleted	
24	4.8	Infrastructure	Ensure digital data storage is appropriate and fit for purpose	<i>Develop process for transfer to storage</i>	Phase 1		Short term	Documentation	In progress	
25	4.9	Infrastructure	Ensure digital data storage is appropriate and fit for purpose	<i>Review plan for transfer to storage</i>	Phase 3		Medium term	Documentation	Uncompleted	
26										
27										
28										
29										
30										



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

	A	B	C	D	E
1	Digital Strong Room - Requirements				
2					
3	Requirement	Essential	Desirable	Type	OAIS function
4	Ability to restrict access to the original digital objects (only by archivists/administrators)	Yes		Operational	Access
5	Ability to integrate with archive catalogue system, ArchivesSpace	Yes		Technical	Access
6	Ability to send the access copy of a digital object to a storage location separate from the preservation copy	Yes		Technical	Access
7	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 management
8	Ability to make some DIPs only accessible by permitted individuals (application of EASE user authentication)	Yes		Functional	Data management
9	Ability to generate logs of access requests and failures	Yes		Functional	Administrative
10	Ability to delete AIPs and/or digital objects within an AIP	Yes		Functional	Archival storage
11	Ability to generate logs of any disposals from the system	Yes		Functional	Administrative
12	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	Administrative

1	Recommended PREMIS metadata fields									
2										
3	ENTITY	SEMANTIC UNIT	SEMANTIC COMPONENT	SEMANTIC COMPONENT	SEMANTIC COMPONENT	CAPTURE?	RATIONALE	PRIORITY	ARCHIVEMATIC	
4	OBJECT									
5	1.1	Object identifier								
6	1.1.1		objectIdentifierType			Yes	Mandatory. Controlled vocabulary would help	Essential	Yes	
7	1.1.2		objectIdentifierValue			Yes	Mandatory. Each object should have a DOI or PUID	Essential	Yes	
8	1.2	objectCategory								
9	1.3	preservationLevel								
10	1.3.1		preservationLevelValue			Yes	Mandatory. Bit level/Full. Controlled vocabulary would help.	Desirable	No	
11	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	
12	1.3.3		preservationLevelRationale			No				
13	1.3.4		preservationLevelDateAssigned			No				
14	1.4	significantProperties								
15	1.4.1		significantPropertiesType			No	Not mandatory. Insufficient resource to dedicate.			
16	1.4.2		significantPropertiesValue			No	Not mandatory. Insufficient resource to dedicate.			
17	1.4.3		significantPropertiesExtension			No	Not mandatory. Insufficient resource to dedicate.			
18	1.5	objectCharacteristics								
19	1.5.1		compositionLevel			Yes	Mandatory. To identify if encoded derivatives (compression/encryption) are present. A controlled vocabulary would help.	Desirable	No	
20	1.5.2		fixity							
21	1.5.2.1			messageDigestAlgorithm		Yes	Mandatory. To identify the type of hash algorithm used. Controlled vocabulary would help.	Essential	Yes	
22	1.5.2.2			messageDigest		Yes	Mandatory. The output of the hash algorithm.	Essential	Yes	
23	1.5.2.3			messageDigestOriginator		Yes	Optional but could be populated either by the Event record (eventIdentifierValue) or agentIdentifierValue	Desirable	No	
24	1.5.3		size			Yes	Must record the size of the file for storage analysis and file checking	Essential	Yes	
25	1.5.4		format							
26	1.5.4.1			formatDesignation						
27	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	
28	1.5.4.1.2				formatVersion	Yes	Mandatory. Need to know the version of the format for migration purposes.	Essential	Yes	
29	1.5.4.2			formatRegistry			This container could be repeated with multiple registries			
30	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	
31	1.5.4.2.2				formatRegistryKey	Yes	This should record the designator within the registry identified above (i.e. fmt/412)	Essential	Yes	

TIP 7: Identify and review your options

Preservica
Digital Preservation



@archivematica.

ExLibris
Rosetta



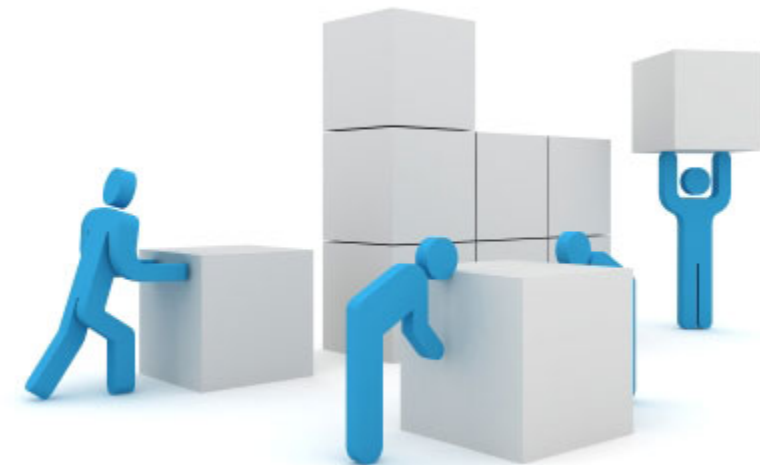
DSPACE



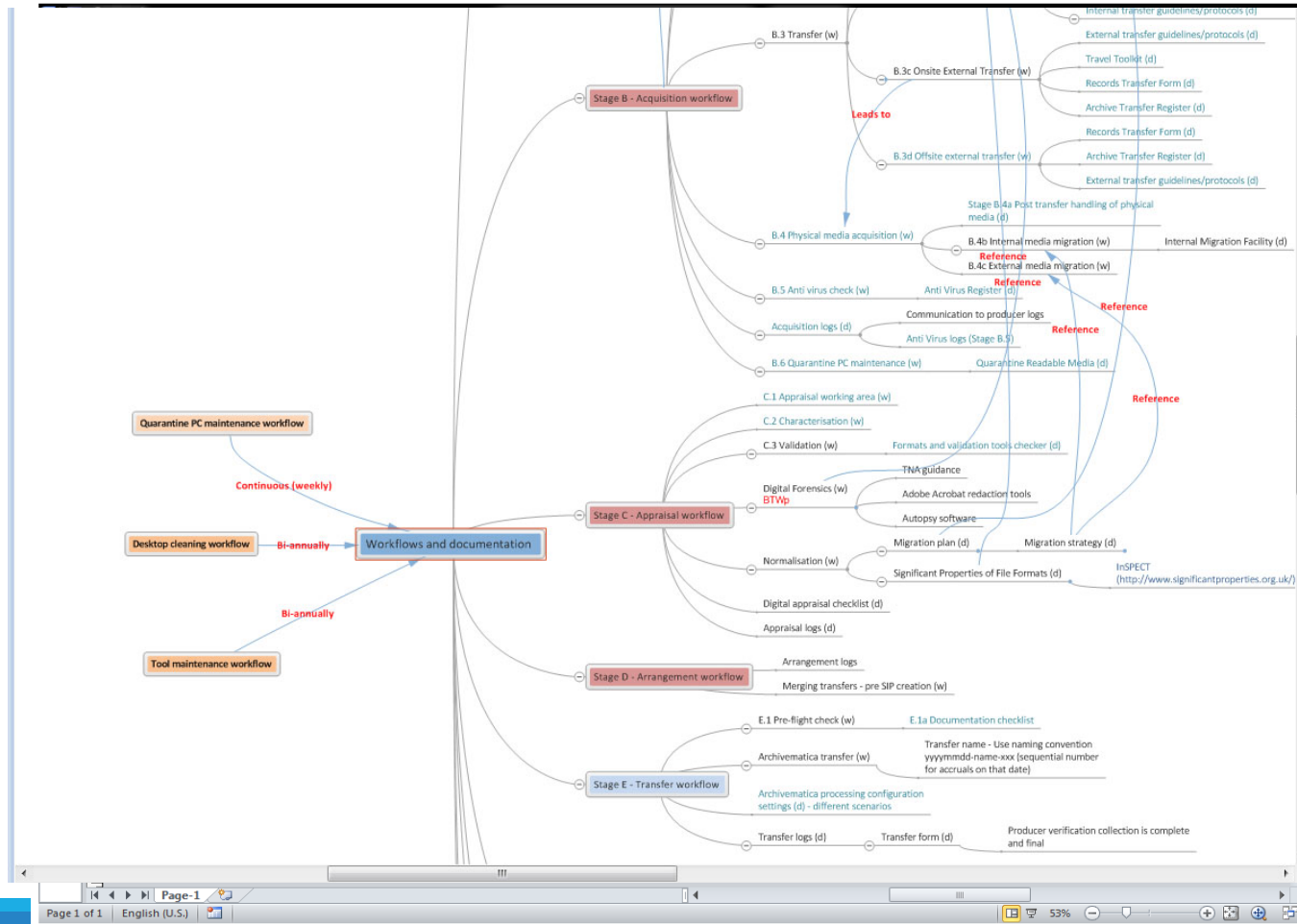
TIP 8: Testing, testing, 1-2-3

1	Archivematica testing									
2										
3	Test no.	Description	Collection size	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 compression 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			AIP Compression took 28 mins. 10 tiffs producing 18 files for AIP
9				SIP	09:27:00	10:03:00	36 mins	1	734 mb	
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			Compression and verification take a while to complete
12				Transfer	11:26:00	11:30:00	4 mins			
13				SIP	11:30:00	11:57:00	27 mins	9	734 mb	
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	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			

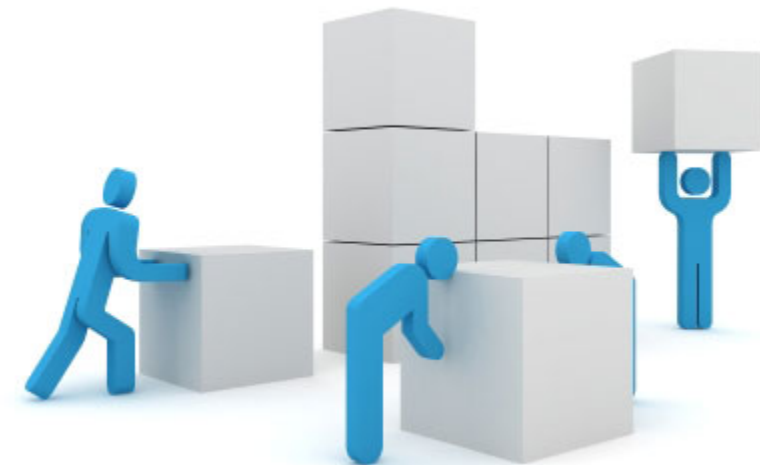
Sheet1



TIP 9: Develop workflows and processes



- ❑ Microsoft Visio (Edraw)
- ❑ MindGenius (Coggle, MindMaple)



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

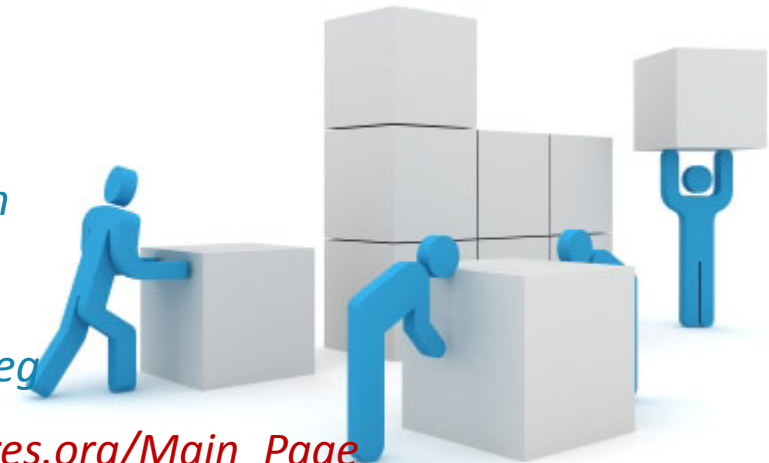
Characterisation/Validation:

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