



Format Policy Registry (FPR)

Making sense of migration strategies as a Community

SAA 2014 – Research Forum

Courtney C. Mumma, MAS/MLIS, Archivematica product manager

Why the FPR?

simplify preservation planning

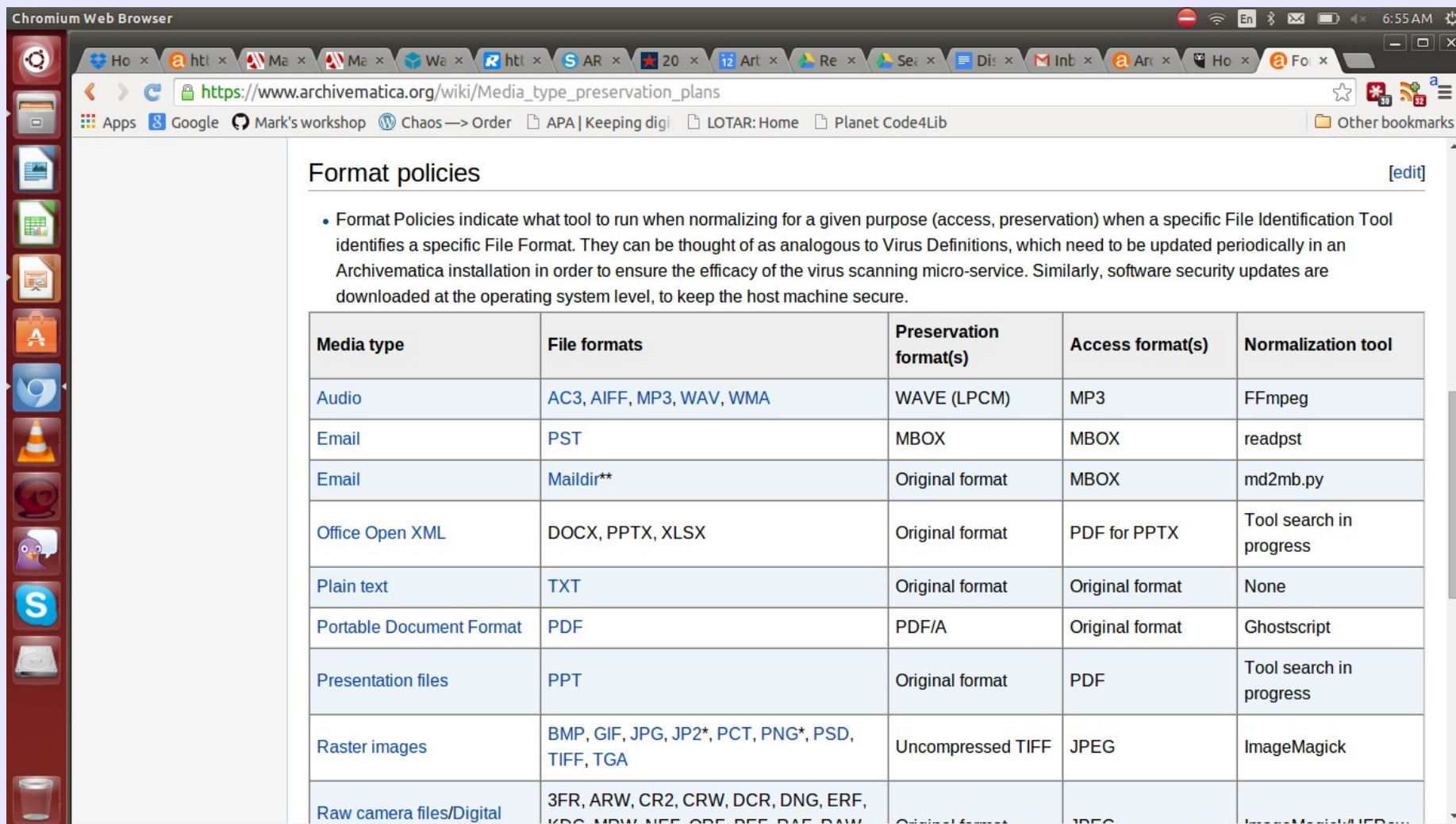
keep current as community standards, practices and tools
improve over time

allow Archivemata users to download and edit the most current
defaults

allow Archivemata users to add their own rules, tools and
commands based on their local preservation plans

aggregate empirical information about institutional format policies
to better identify community best practices

We were guilty, too



Format policies [\[edit\]](#)

- Format Policies indicate what tool to run when normalizing for a given purpose (access, preservation) when a specific File Identification Tool identifies a specific File Format. They can be thought of as analogous to Virus Definitions, which need to be updated periodically in an Archivemata installation in order to ensure the efficacy of the virus scanning micro-service. Similarly, software security updates are downloaded at the operating system level, to keep the host machine secure.

Media type	File formats	Preservation format(s)	Access format(s)	Normalization tool
Audio	AC3, AIFF, MP3, WAV, WMA	WAVE (LPCM)	MP3	FFmpeg
Email	PST	MBOX	MBOX	readpst
Email	Maildir**	Original format	MBOX	md2mb.py
Office Open XML	DOCX, PPTX, XLSX	Original format	PDF for PPTX	Tool search in progress
Plain text	TXT	Original format	Original format	None
Portable Document Format	PDF	PDF/A	Original format	Ghostscript
Presentation files	PPT	Original format	PDF	Tool search in progress
Raster images	BMP, GIF, JPG, JP2*, PCT, PNG*, PSD, TIFF, TGA	Uncompressed TIFF	JPEG	ImageMagick
Raw camera files/Digital	3FR, ARW, CR2, CRW, DCR, DNG, ERF, KDC, MDW, NEF, ORF, PEF, RAF, RAW	Original format	JPEG	ImageMagick/FFmpeg

format policy = rules + tools + commands

format policies exist for:

format identification
characterization

normalization to access and preservation
extraction
transcription
verification

mumma.archivematica.org/fpr/format/

Apps Google Mark's workshop Chaos → Order APA | Keeping digi LOTAR: Home Planet Code4Lib Other book

archivematica Transfer Ingest Archival storage Preservation planning Access Administration courtney@artefactual.com

Formats

Format Information

Formats [Create New Format](#) [Update](#)

Groups [Show](#) Search:

10 entries

	Description	Group	Actions
Identification	3D Studio Shapes	Unknown	View Edit
Tools	3DM	Unknown	View Edit
Rules	3GPP Audio/Video File	Unknown	View Edit
Commands	7 Zip	Package	View Edit
Format policy registry	7-bit ANSI Text	Unknown	View Edit
Tools			
Normalization			

interface with PRONOM (660)

Format Groups

Format Group Information

Formats

[Create New Format Group](#)

Groups

Show

50

entries

Search:

Identification

Tools

Rules

Commands

Format policy registry

Tools

Normalization

Rules

Commands

Characterization

Rules

Commands

Extraction

Rules

Description

Actions

Audio

Edit | Delete

Disk Image

Edit | Delete

Email

Edit | Delete

Image (Raster)

Edit | Delete

Image (Vector)

Edit | Delete

Package

Edit | Delete

Portable Document Format

Edit | Delete

Presentation

Edit | Delete

Spreadsheet

Edit | Delete

Text (Markup)

Edit | Delete

Text (Plain)

Edit | Delete

Unknown

Edit | Delete

Video

Edit | Delete

Word Processing

Edit | Delete

Showing 1 to 14 of 14 entries

◀ Previous Next ▶

Identification Rule Information

Formats [Create New Rule](#)

Groups Show

10 entries

Format	Command	Output	Tools	Actions
Audio: Audio Interchange File Format (compressed)	Identify by File Extension	.aifc		View Replace Disable
Audio: Audio Interchange File Format: Generic AIFF	Identify by File Extension	.aif		View Replace Disable
Audio: Audio Interchange File Format: Generic AIFF	Identify by File Extension	.aiff		View Replace Disable
Audio: Dolby Digital Audio: AC-3 Compressed	Identify by File Extension	.ac3		View Replace Disable
Audio: MPEG Audio: MPEG 1/2 Audio Layer 3	Identify by File Extension	.mp3		View Replace Disable
Audio: MPEG Audio: MPEG 1/2 Audio Layer I	Identify by File Extension	.mp1		View Replace Disable
Audio: MPEG Audio: MPEG Audio Stream, Layer II	Identify by File Extension	.mp2		View Replace Disable
Audio: WAV: Generic WAV	Identify by File Extension	.wav		View Replace Disable
Audio: Windows Media Audio: WMA file	Identify by File Extension	.wma		View Replace Disable
Disk Image: Advanced Forensic Format: Advanced Forensic Format	Identify by File Extension	.aff		View Replace Disable

Showing 1 to 10 of 618 entries ◀ Previous Next ▶

identification

file or extension

Other tools can be added
results are basis of other
format policy actions

Format Policy Registry Tool Information

Formats

[Create New Tool](#)

Groups

Show

Search:

25

entries

Identification

Tools

Rules

Commands

Format policy registry

Tools

Normalization

Rules

Commands

Characterization

Rules

Commands

Extraction

Rules

Commands

Description	Version	Actions
7Zip	9.20	View Edit
Archivematica Script	1.0	View Edit
convert	6.6.7-9	View Edit
cp	(GNU coreutils) 8.20	View Edit
echo	GNU coreutils 8.12.197-032bb	View Edit
ExifTool	9.27	View Edit
ffmpeg	0.10.6-6:0.10.6-0ubuntu0jon1~precise1	View Edit
FFprobe	2.2.0	View Edit
FITS	0.6.1	View Edit
Ghostscript	9.05 (2012-02-08)	View Edit
inkscape	0.48.3.1 r9886 (Jan 29 2013)	View Edit
MediaInfo	0.7.52	View Edit
ps2pdf	9.05 (2012-02-08)	View Edit
Sleuthkit	4.1.3	View Edit
test	5.97	View Edit
unrar-free	0.0.1	View Edit

Showing 1 to 16 of 16 entries

◀ Previous Next ▶

Format Policy Rule Information

Formats

[Create New Rule](#)

Groups

Show

Search:

100

entries

Identification

Tools

Rules

Commands

Format policy registry

Tools

Normalization

Rules

Commands

Characterization

Rules

Commands

Extraction

Rules

Commands

Transcription

Rules

Commands

Event Detail

Rules

Commands

Verification

Rules

Commands

Purpose	Format	Command	Success	Actions
Preservation	Generic PDF	Transcoding to pdfa with Ghostscript	0 out of 0	View Replace
Preservation	Nikon Digital SLR Camera Raw Image File	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Kodak Digital Camera Raw Image File	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Sony RAW Image File	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Canon CR2 Raw	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Raw Bitmap	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Generic TGA	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Truevision TGA Bitmap 2.0	Transcoding to tif with convert	1 out of 1	View Replace
Preservation	Truevision TGA Bitmap 1.0	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Generic PICT	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	PICT 1.0	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Adobe Photoshop	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Exchangeable Image File Format (Uncompressed) 2.2	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Exchangeable Image File Format (Uncompressed) 2.1	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Exchangeable Image File Format (Uncompressed) 2.0	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Generic DNG	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	Geographic Tagged Image File Format (GeoTIFF)	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	PICT 2.0	Transcoding to tif with convert	0 out of 0	View Replace
Preservation	TIFF	Transcoding to tif with convert	1 out of 1	View Replace

normalization rules for access, preservation and thumbnails (for internal search)

extraction

Format Policy Rules

Format Policy Rule Information

Formats

[Create New Rule](#)

Groups

Show

Search:

50 ▾

Identification

entries

Tools

Rules

Commands

Format policy registry

Tools

Normalization

Rules

Commands

Characterization

Rules

Commands

Extraction

Rules

Commands

Transcription

Rules

Commands

Event Detail

Rules

Purpose	Format	Command	Success	Actions
Extract	7Zip	Extracting 7Zip compatible format.	0 out of 0	View Replace
Extract	Advanced Forensic Format	tsk_recover	0 out of 0	View Replace
Extract	BZip	Extracting 7Zip compatible format.	0 out of 0	View Replace
Extract	BZip2	Extracting 7Zip compatible format.	0 out of 0	View Replace
Extract	CAB	Extracting 7Zip compatible format.	0 out of 0	View Replace
Extract	Generic RAR	Extracting RAR format.	0 out of 0	View Replace
Extract	GZip	Extracting 7Zip compatible format.	0 out of 0	View Replace
Extract	ISO Disk Image File	tsk_recover	0 out of 0	View Replace
Extract	RAR 2.0	Extracting RAR format.	0 out of 0	View Replace
Extract	RAR 2.9	Extracting RAR format.	0 out of 0	View Replace
Extract	Raw Disk Image	tsk_recover	0 out of 0	View Replace
Extract	Tape Archive Format	Extracting 7Zip compatible format.	0 out of 0	View Replace
Extract	ZIP file	Extracting 7Zip compatible format.	0 out of 0	View Replace

Showing 1 to 13 of 13 entries

◀ Previous Next ▶

transcription

Format Policy Rules

Format Policy Rule Information

Formats

[Create New Rule](#)

Groups

Show

Search:

10

entries

Identification

Tools

Purpose	Format	Command	Success	Actions
---------	--------	---------	---------	---------

Rules

Transcription	JPEG 1.00	Transcribe using Tesseract	0 out of 0	View Replace
---------------	-----------	----------------------------	------------	--

Commands

Showing 1 to 1 of 1 entries

◀ Previous Next ▶







Format policy

dashboard decision-making

archivematica **Transfer** Ingest ² Archival storage Preservation planning Access Administration panther ▾

Standard ▾ /home ▾ Browse Start transfer

Type Transfer name Accession no.

Transfer	UUID	Transfer start time	
✔ MultiMediaAug11	d977431c-bfa7-4f60-8658-b1c7985f1b03	2014-08-11 08:43	 
▶ Micro-service: Create SIP from Transfer			
▶ Micro-service: Complete transfer			
▶ Micro-service: Characterize and extract metadata			
▶ Micro-service: Examine contents			
▶ Micro-service: Extract packages			
▼ Micro-service: Identify file format			
Job: Identify file format		Completed successfully	
Job: Determine which files to identify		Completed successfully	
Job: Select file format identification command		Completed successfully	
Job: Move to select file ID tool		Completed successfully	
▶ Micro-service: Clean up names			
▶ Micro-service: Generate transfer structure report			
▶ Micro-service: Scan for viruses			
▶ Micro-service: Quarantine			
▶ Micro-service: Generate METS.xml document			
▶ Micro-service: Verify transfer checksums			
▶ Micro-service: Reformat metadata files			

preconfiguring policy decisions

archivematica [Transfer](#) [Ingest](#) [Archival storage](#) [Preservation planning](#) [Access](#) [Administration](#) [panther](#)

Processing configuration

- General
- Failures
- Transfer source locations
- AIP storage locations
- AtoM DIP upload
- CONTENTdm DIP upload
- Archivists Toolkit DIP upload
- PREMIS agent
- REST API
- Users
- Version

<input checked="" type="checkbox"/> Send transfer to quarantine	No
<input type="checkbox"/> Approve normalization	Yes
<input type="checkbox"/> Store AIP	Yes
<input checked="" type="checkbox"/> Transcribe files (OCR)	No
<input checked="" type="checkbox"/> Generate transfer structure report	No
<input type="checkbox"/> Remove from quarantine after	<input type="text"/> days
<input type="checkbox"/> Create SIP(s)	--Actions--
<input checked="" type="checkbox"/> Extract packages	Yes
<input type="checkbox"/> Normalize	--Actions--
<input checked="" type="checkbox"/> Reminder: add metadata if desired	Continue
<input checked="" type="checkbox"/> Examine contents	Skip examine contents
<input type="checkbox"/> Select file format identification command (Transfer)	--Actions--
<input checked="" type="checkbox"/> Select file format identification command (Ingest)	Use existing data

FPR future

front-facing server website – better usability

ability to upload local policies and share reasoning

leverage FPR for shared experimentation with format policies
and preservation planning

interface with other registries, including linked data

FPR as a service, partner with other FOSS projects

and ??????????

archivematica.org

Twitter: @archivematica

Online demo:

sankofa.archivematica.org

User: demo@example.com

Password: demodemo

Me: courtney@artefactual.com

[@arxiverdama](#)

I have free stickers and pins