SOLOMON R. GUGGENHEIM FOUNDATION

Electronic Records Management Start-Up Project
Electronic Records Processing Manual
Appendix A: SRGF Preservation and Access Formats
Updated May 9, 2014 – Prepared by Anthony Cocciolo, Electronic Records Consultant

Original file is preserved, and files are created for preservation and access. This table is based on the normalization table from Archivematica.[1]

Relevant Pilot	Media type	File formats	Preservation format(s)	Access format(s)	Archivist needs to manually normalize?	Notes
Pilot 5 - Standard Records	MS Office (2007 and greater): Word, PowerPoint	DOCX, PPTX	Original format	PDF	Yes, use script created by consultant to normalize: `pres&access.vbs`[2]	MS Office Open XML files can be unzipped and viewed as XML.
	MS Office (2007 and greater): Excel	XLSX	Original format	Original format	No.	
	MS Office (binary formats): Word & Powerpoint	DOC, PPT	DOCX, PPTX	PDF	Yes, use script created by consultant to normalize: `pres&access.vbs`	MS Office binary files are well documented, however, the open XML versions may be more sustainable because they are text-based.[3]
	MS Office (binary formats):	XLS	XLSX	XLSX	Yes, use script created by consultant to	Not normalized to PDF because

---

[1] https://www.archivematica.org/wiki/Media_type_preservation_plans#Normalization
[2] Script available on: https://github.com/Guggenheim
[3] http://www.digitalpreservation.gov/formats/intro/specifications.shtml

	Excel				normalize: `pres&access.vbs`	spreadsheets can be tedious to view in this format.
	Plain text	TXT	Original format	Original format	No.	
	Portable Document Format	PDF	PDF/A	Original format	No, automatically converted by Archivematica using Ghostscript[4]	PDF submitted should include fonts embedded; avoid "Smallest file size" PDFs which do not embed fonts.[5], InDesign and QuarkXPress files should come in as high-quality or press quality PDFs.
	Raster images – Highly sustainable[6]	JPG, JPG2000, PNG, TIFF	Original format	JPG	No, automatically converted by Archivematica using ImageMagick.	
	Raster images – Questionable sustainability[7]	BMP, PCT, TGA, GIF, PSD[8]	Uncompressed TIFF	JPG	No, automatically converted by Archivematica using	

---

[4] Using the behavior described here: https://www.archivematica.org/wiki/PDF_to_PDF/A_using_Ghostscript
[5] http://help.adobe.com/en_US/acrobat/X/pro/using/WSb2f1a50375cd48d3-1f36d19412ada208ceb-8000.html
[6] Based on Library of Congress' Sustainability of Digital Formats – Still Images: http://www.digitalpreservation.gov/formats/fdd/still_fdd.shtml
[7] Ibid.

					ImageMagick.	
	Adobe Illustrator	AI	PDF/A	PDF	No, most Illustrator files have PDF embedded within it, so it is not necessary to create a separate PDF. However, if PDF is not embedded within the Illustrator file, then create a separate PDF.	The lack of a PDF being embedded within the InDesign file will be caught with FIDO format identification in Archivematica.
	Rich Text format	RTF	Original format	PDF	Yes, use script created by consultant to normalize: `pres&access.vbs`	
	Adobe InDesign	INDD, IND	PDF/A	PDF	Yes, create PDF/A and PDF using InDesign and Acrobat.	InDesign files should generally not be submitted; prefer high-quality PDFs that will be auto-converted into PDF/A.
Pilot 6 - Obscure file formats	AutoCAD [2D model / drawing]	DWG, DXF	Original format	PDF	Yes, export to PDF in DWG viewer.	The DWG file format has been openly specified, and open

---

[8] Adobe Creative Suite files (Photoshop, Illustrator, InDesign, Flash) could face long term preservation challenges because Adobe will only be making those products available via cloud subscription: http://news.cnet.com/8301-1001_3-57582735-92/adobe-kills-creative-suite-goes-subscription-only/ However, Adobe has made great strides in providing open documentation on many of their formats, which helps ameliorate the previously mentioned concern.

					software is available for reading it.[9] Also, AutoDesk makes a free DWG viewer – DWG TrueView.[10]
	Bentley Systems CAD format [2D model / drawing]	DGN	DWG	PDF	Yes, convert to DWG and PDF using AutoCAD.	
	VectorWorks and MiniCad [2D model / drawing]	MCD, VWX	DWG [AutoCAD]	PDF	Yes, export to DWG and PDF using VectorWorks.	Major category – 19,130 files in this format. Export to DWG from VectorWorks.
	Rhino [3D Models]	3DM	Original format	Original format, PDF (acts like a preview)	Yes. In Rhino, select Print and Save as PDF. Include the 3DM file and PDF file in the manually normalized access folder.	Major category – 1,111 files in this format. To preserve would need to preserve software and emulate if needed.[11]
	Autodesk 3D Max [3D models]	3DS	Original format	3DM, PDF (acts like a preview)	Yes. Export to Rhino 3dm using Autodesk 3D Max. Create a PDF	Minor format at SRGF; To preserve would need to preserve

[9] http://opendesign.com/files/guestdownloads/OpenDesign_Specification_for_.dwg_files.pdf ; http://www.opendesign.com/guestfiles/teigha_viewer
[10] http://usa.autodesk.com/adsk/servlet/pc/index?id=6703438&siteID=123112
[11] http://www.ijdc.net/index.php/ijdc/article/viewFile/105/80

					preview, and include the 3DM and PDF file in the manually normalized access folder.	software and emulate if needed.[12] Or, convert to Rhino.[13]
	Raw camera files/Digital Negative format**	3FR, ARW, CR2, CRW, DCR, DNG, ERF, KDC, MRW, NEF, ORF, PEF, RAF, RAW, X3F	Uncompressed TIFF	JPEG	No, automatically converted in Archivematica using ImageMagick and UFRaw.	
	General Vector images	EPS, SVG	SVG	PDF	No, automatically converted in Archivematica using Inkscape.	
Pilot 7 – Obsolete file formats	WordPerfect files	WPD	DOCX	PDF	Yes, use script created by consultant to normalize: `pres&access.vbs`	Script leverages MS Word 2007+ for Windows to do conversion.
	QuarkXPress	QXD	PDF/A	PDF	Yes, export to PDF/A and PDF using InDesign and Acrobat. Need fonts used available on system.	QXD files should generally not be submitted; prefer high-quality PDFs that will be auto-converted into PDF/A.

---

[12] http://www.ijdc.net/index.php/ijdc/article/viewFile/105/80
[13] If converting to Rhino, would want to ensure that no information is lost in the process, such as textures.

	Lotus 1-2-3	WK4	XLSX	XLSX	Yes, export using Gnumeric.[14]	
Pilot 9 – Significant Email Correspondence	MS Outlook Personal Folders	PST	MBOX	MBOX (need viewer)	No, uses readpst to create MBOX file.	A viewer for MBOX files is required, and can be installed on an archives computer.
	MS Outlook Message	MSG	EML	PDF	Yes, see Appendix C	EML file is text and MIME based and well-suited for long-term preservation.[15]
	Novell Groupwise Email	MLM	DOCX	PDF	Yes, use script created by consultant to normalize: `pres&access.vbs`	MLM files are WordPerfect 5.X files
	MS Outlook Express Message	EML	Original format	PDF	Yes, see Appendix C	
Pilot 10 – Web archiving SRGM Websites	Web documents stored on Network	HTML. HTM, Misc. files used for web display (JPG, PNG, etc.)	Text based remain original format; other files converted as indicated in this table.	Text based remain original format; other files converted as indicated in this	No.	

[14] http://www.gnumeric.org/
[15] http://www.dpconline.org/component/docman/doc_download/739-dpctw11-01pdf

					table.		
	Web documented retrieved from web archiving	HTML, HTM, Misc. files used for web display (JPG, PNG, etc.)	WARC	WARC	No.	Heretix (retrieve) and WayBackMachine (view). WARC viewers can be used for reading the WARC file.
Pilot 11 – Very large files (or Video/Audio media)	Audio	AC3, AIFF, MP3, WAV, WMA	WAVE (LPCM)	MP3	No, automatically converted in Archivematica using FFmpeg	
	Video	AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, WMV, DV	FFV1/LPCM in MKV	MP4	No, automatically converted in Archivematica using FFmpeg	