ARCHIVAL PRESERVATION of DIGITAL OBJECTS
INTRODUCTION:

JORDAN BASS

MEDICAL ARCHIVIST (ACADEMIC) | UNIVERSITY OF MANITOBA
VICE-PRESIDENT | ASSOCIATION OF CANADIAN ARCHIVISTS
OBJECTIVES

1. Migration: Understanding digital preservation strategy
2. Format Policy Registries: Selecting preservation formats for digital objects
3. Digital Storage: Choosing storage platforms
4. More Backups, More Problems: How many copies are enough and where do I keep them?
5. Archival Integrity: Checking and recovering data
AND THEN WE TOLD THEM

"WE'RE BUILDING A TRUSTED DIGITAL REPOSITORY"
3 DIGITAL PRESERVATION MYTHS

1. Digital preservation is only done by “digital archivists”

2. Digital objects and systems rapidly become obsolete

3. Digital preservation is for big institutions, not small archives
MIGRATION
NORMALIZATION

“is a form of migration where an archives
minimizes the number of file formats it will
preserve by migrating content to a small number
of preferred formats”
FPR (format policy registries)
“A format policy indicates the actions, tools and settings to apply to a digital object of a particular format (e.g. conversion to preservation format, conversion to access format, extraction of package formats).”
<table>
<thead>
<tr>
<th>Media type</th>
<th>File formats</th>
<th>Preservation format(s)</th>
<th>Access format(s)</th>
<th>Normalization tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td>AC3, AIFF, MP3, WAV, WMA</td>
<td>WAVE (LPCM)</td>
<td>MP3</td>
<td>FFmpeg</td>
</tr>
<tr>
<td>Email</td>
<td>PST</td>
<td>MBOX</td>
<td>MBOX</td>
<td>readpost</td>
</tr>
<tr>
<td>Email</td>
<td>Maildir**</td>
<td>Original format</td>
<td>MBOX</td>
<td>md2mb.py</td>
</tr>
<tr>
<td>Office Open XML</td>
<td>DOCX, PPTX, XLSX</td>
<td>Original format</td>
<td>Original format</td>
<td>Tool search in progress</td>
</tr>
<tr>
<td>Plain text</td>
<td>TXT</td>
<td>Original format</td>
<td>Original format</td>
<td>None</td>
</tr>
<tr>
<td>Portable Document Format</td>
<td>PDF</td>
<td>PDF/A</td>
<td>Original format</td>
<td>Ghostscript</td>
</tr>
<tr>
<td>Presentation files</td>
<td>PPT</td>
<td>Original format</td>
<td>PDF</td>
<td>Tool search in progress</td>
</tr>
<tr>
<td>Raster images</td>
<td>BMP, GIF, JPG, JP2*, PCT, PNG*, PSD, TIFF, TGA</td>
<td>Uncompressed TIFF, JPEG</td>
<td></td>
<td>ImageMagick</td>
</tr>
<tr>
<td>Raw camera files/Digital Negative format**</td>
<td>3FR, ARW, CR2, CRW, DCR, DNG, ERF, KDC, MRW, NEF, ORF, PEF, RAF, RAW, X3F</td>
<td>Original format</td>
<td>JPEG</td>
<td>ImageMagick/UF RAW</td>
</tr>
<tr>
<td>Spreadsheets</td>
<td>XLS</td>
<td>Original format</td>
<td>Original format</td>
<td>None</td>
</tr>
<tr>
<td>Vector images</td>
<td>AI, EPS, SVG</td>
<td>SVG</td>
<td>PDF</td>
<td>Inkscape</td>
</tr>
<tr>
<td>Video</td>
<td>AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, WMV</td>
<td>FFV1/LPCM in MKV, MP4</td>
<td>MP4</td>
<td>FFmpeg</td>
</tr>
<tr>
<td>Word processing files</td>
<td>DOC, WPD, RTF</td>
<td>Original format</td>
<td>Original format</td>
<td>Tool search in progress***</td>
</tr>
</tbody>
</table>
DIGITAL STORAGE
NEVER

NOT.EVEN.ONCE.
KISS PRINCIPLE
KEEP. IT. SIMPLE. STANLEY.
Grant Math

- 100 TIFFs at approx. 300 – 700 MB per image = 70 GB
- 100 JPG derivatives at approx. 300 KB (that’s generous) = 0.03 GB
- Approx. 70.03 GB of storage used for 100 images (preservation and access)
- 3 Terabyte drive will net approx. 8420 archival images
Corporate IT & CLOUD Storage
Backup: How many copies are enough?
Minimum Requirements:

3 External Drives
or
1NAS + 1 External Drive
Initial Transfer of digital objects: External Drives
TEAMWORK MAKES THE DREAM WORK
Initial Transfer of digital objects: NAS
ARCHIVAL INTEGRITY

Checksums

Recovery
Checksums & Data Integrity
1. Digital preservation is only done by “digital archivists”

2. Digital objects and systems rapidly become obsolete

3. Digital preservation is for big institutions, not small archives

TEAMWORK MAKES THE DREAM WORK
THANKS!

Any questions?

You can find me at jordan.les.bass@gmail.com
CREDITS

Special thanks to all the people who made and released these awesome resources for free:

Presentation template by SlidesCarnival
Photographs by Unsplash
Icons designed by Freepik from www.flaticon.com