MESSES AND MEANINGS

Sheila Morrissey
@sheilaMorr

DPC Re:Format - What is file format obsolescence and does it really exist?
York
22 June 2016
Portico is a preservation service for digital publications, including electronic journals, books, and historical collections.
Text: **pdf 72million** 7.39 postscriptXML, SGML, Javascript, latex, perl, utf-8, css, csv, html(.3), asm, C

<table>
<thead>
<tr>
<th>Format Type</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>application/vnd.corel-draw (cdr)</td>
<td>13</td>
<td>0.00%</td>
</tr>
<tr>
<td>application/x-ptc-paintshoppro (psp)</td>
<td>1</td>
<td>0.00%</td>
</tr>
<tr>
<td>image/gif (gif)</td>
<td>10,771</td>
<td>39.32%</td>
</tr>
<tr>
<td><strong>image/jpeg (jpg)</strong></td>
<td>149,779,348</td>
<td>15.33%</td>
</tr>
<tr>
<td>image/png (png)</td>
<td>9,170,228</td>
<td>0.94%</td>
</tr>
<tr>
<td><strong>image/tiff (tif)</strong></td>
<td>99,870,761</td>
<td>10.22%</td>
</tr>
<tr>
<td>image/vnd.adobe.photoshop (psd)</td>
<td>3,744</td>
<td>0.00%</td>
</tr>
<tr>
<td>image/x-ms-bmp (bmp)</td>
<td>11</td>
<td>0.00%</td>
</tr>
<tr>
<td>image/x-ms-bmp-v3 (bmp)</td>
<td>4,836</td>
<td>0.00%</td>
</tr>
<tr>
<td>image/x-wmf (wmf)</td>
<td>286</td>
<td>0.00%</td>
</tr>
<tr>
<td>Type</td>
<td>Sum of Num Files Not Well Formed</td>
<td>Sum of Num Files Well Formed and Not Valid</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>AUDIO</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>DATA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EXECUTABLE</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MOVING_IMAGE</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OTHER</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PACKAGE</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>PORTICO_TEXT</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>STILL_IMAGE</td>
<td>1,108,628</td>
<td>1,883,062</td>
</tr>
<tr>
<td>TEXT</td>
<td>5,521,112</td>
<td>7,173,121</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td><strong>6,629,905</strong></td>
<td><strong>9,056,183</strong></td>
</tr>
<tr>
<td>Type</td>
<td>Sum of Files Not Well Formed</td>
<td>Sum of Files Well Formed and Not Valid</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>AUDIO</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>DATA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EXECUTABLE</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MOVING_IMAGE</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OTHER</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PACKAGE</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>PORTICO_TEXT</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>STILL_IMAGE</td>
<td>1,108,628</td>
<td>1,883,062</td>
</tr>
<tr>
<td>TEXT</td>
<td>5,521,112</td>
<td>7,173,121</td>
</tr>
<tr>
<td>Grand Total</td>
<td>6,629,905</td>
<td>9,056,183</td>
</tr>
</tbody>
</table>
Undetermined

Audio: mpeg (undetermined)

Data:
fits, maple mathematica, rtf, mathcad, stata data and programs
pdb  chem  cdx/cif/jmol-voxel,mol,(80K)
word 527,142
excel 136,335
ppt 141,006
open office 390K

Executable
Moving Image
realmedia, flash, avi, mp4, mpeg, quicktime, wmv

Still-image not determined: almost entirely PNG 1.4 % of our image files

text not determined dataset.toc, dataset.xml

Package: epub, stuffit

OTHER Bytestream: not identified
“I come to praise PDF, not to bury it”
Well, maybe just a little...

OMG-WTF-PDF
[PDF Ambiguity and Obfuscation]

Julia Wolf
FireEye
2010 March 31
Troopers 11

Calvin is a cataract, a primeval forest, a demonic power, something directly down from Himalaya, absolutely Chinese, strange, mythological; I lack completely the means, the suction cups, even to assimilate this phenomenon, not to speak of presenting it adequately.


Incredible accomplishment 23 years ago 2013

Interoperable (before Java)
“Interchange PostScript” Analogous to simplification from SGML to XML
Portable document format
(for example, files produced on older versions of Apple’s operating system, which include Apple Single and Apple Double encoding information preceding the header)

Other developers of PDF applications clearly have struggled with the gap between PDF “in theory” and PDF instances in actual practice. Comments in the PDFParser source code from Adobe’s PDFBox Java application note some of the various compensatory actions taken silently to repair non-conforming documents.[12] The widely-used iText library includes an “isRebuilt()” method to indicate that it has “repaired” syntactic errors
Tools break – iText, PDFBox – we just chain them split PDF files
Characterize the renderers

VeraPDF better expression of failures of conformance
Duff Johnson, PDF Association – DPC, OPF

Match it up to render characterizer
Interoperable

Never validated - NWF, WFNV, Encoding (easy stuff - syntax)
Trickier: procedural semantics (Display italic vs regular,), “Implied text” - boiler plate – good practice – hard for preservation (encapsulate)
Month 32

Impedance mismatch -- fine/coarse grain  Synonymy

It's a tree

Sell-by date -- sophisticated use
Separation of concerns -
Hello, World!!
Hello, World!
Embroidery Trouble Shooting Guide

Thread Breakage

Causes:

Improper Thread
Try re-threading the machine; make sure the thread goes through all guides.

Burr
There may be burrs in the needle’s eye, on the thread guides, needle plate or the hook. Replace the needle and try buffing the thread guides and needle plate. Buffing may alter the timing, so it’s a good idea to replace a damaged hook.

Dry Hook
Because it dries out faster, the hook needs frequent lubrication. The hook assembly should be lubricated every 4-8 hours of machine running time.

Needle too small
You may have to change to a larger needle.

Too much Adhesive spray
You should be using only Embroidery designed adhesive sprays, never ever use them. Too much adhesive build up on the needle can cause them to drag and break threads. If this is happening you should examine the way you use your adhesive spray
Causes:

Improper Thread
Try re-threading the machine; make sure the thread goes through all guides.

Burr
There may be burrs in the needle’s eye, on the thread guides, needle plate or the hook. Replace the needle and try buffing the thread guides and needle plate. Buffing may alter the timing, so it’s a good idea to replace a damaged hook.

Dry Hook
Because it dries out faster, the hook needs frequent lubrication. The hook assembly should be lubricated every 4-8 hours of machine running time.

Needle too small
You may have to change to a larger needle

Too much Adhesive spray
You should be using only Embroidery designed adhesive sprays, never overt use them. Too much adhesive build up on the needle can cause them to drag and break threads. If this is happening you should examine the way you use
“Rendering Matters”
POSSIBILITIES

- It was an accident
- It was intentional
- It doesn’t matter
Why might an “original experience” matter

https://commons.wikimedia.org/wiki/File:NMABG_Aphrodite_Archer_1.jpg
Why might an “original experience” matter

Data, Software, & Reproducibility
This is a joint Project of the ACM Digital Library and the ACM Technology Committees
Meanings matter

The meaning and value of these data derive not only from the raw content, but from the ways people interact with the technologies that create them.

Sara Day Thompson – preserving transactional data
Questions?