DAMS
(a.k.a “DAMNS”, “DAMBS”, “DAMu”, “DAAAMMS”)

https://github.com/fieldmuseum/EMu-Documentation
the silence of the DAMS
A brief history of the FMNH EMu/DAMu-verse
A brief history of the FMNH EMu/DAMu-verse

- **2008**
  - $85 million

- **2010**
  - MM opened to all users

- **2012**
  - Digitization Projects

- **2014**
  - Standardization Project...

- **2016**
  - Catalogue

- **2018**
  - Multimedia

- **2019**
  - DAMS...

- **2020**
  - 20 TB
  - 2,500,000
  - 1,500,000
2008 - 2018

...taking stock of ~10 year’s worth of Multimedia:

- File formats… not so standard
  - e.g., XLSX, 32 flavors of TIFF
- Filenames… also not so standard
  - e.g.: “DÈRP’s b.123(4).jpg”
- Files themselves…
  - sometimes go missing between EMu & storage → [broken icon]
  - dupdupduplicates
- Keywords… how to standard?
  - there was yelling...

[the screaming of the DAMS]
File formats

...not so standard...
...some proprietary...
...reflect a user’s local registry...
File formats

Recommended file formats:
- PDF - for text documents
- JPG - for images
- DNG - for high-quality images
- MP4 - for moving images/video
- CSV - for spreadsheets

Documented standards…
Not set in stone, but reference-able

Standard File Formats permitted in EMu for each Resource Type
- Dataset = CSV, ODS, ZIP* or TAR*
- Dataset | Sequence = ZIP* or TAR*
- Image = 3D model = PLY, STL, or OBJ
- Image | CT Data = ZIP* or TAR* containing folder for 1 scan dataset (set of TIFF or BMP)
- Image | Image Set = ZIP* or TAR* containing folder for 1 image set (set of JPG, TIFF, or BMP)
- InteractiveResource = ZIP* or TAR*
- MovingImage | Video = mp4 or mov
- MovingImage | Animation = mp4 or mov
- Software = ZIP* or TAR*
- Sound = MP3, WAV
- StillImage = JPG, DNG
- StillImage | X-ray = DCM or 16-bit grayscale TIFF
- Text | Document = TXT, PDF, or ODT
- Text | Permit = PDF
Filenames

...also not so standard

- “DÈ[R]P’s b.123 (#4).jpg”

  = an example of a file that will disappear itself from a Multimedia export.

- ...mostly from the “È” character with the diacritic,
- but if we’re fixing that, may as well pull out other special characters)
Filenames

--------sh*tty-filename-query:

... MulIdentifier contains "\[\-\()~!@#$%^&*`,\-\]*/?\;:\"\" or MulIdentifier contains "\.*\.*\.*" or MulIdentifier contains "\[áÁàÀâÂäÄãÃåÅæÆçÇéÉèÈêÊëËíÍìÌîÎïÏñÑóÓòÒôÔöÖõÕøØœŒßúÚùÙûÛüÜ\]*/?\;:\"\"

or SupIdentifier contains "\[\-\()~!@#$%^&*`,\-\]*/?\;:\"\" or SupIdentifier contains "\.*\.*\.*" or SupIdentifier contains "\[áÁàÀâÂäÄãÃåÅæÆçÇéÉèÈêÊëËíÍìÌîÎïÏñÑóÓòÒôÔöÖõÕøØœŒßúÚùÙûÛüÜ\]*/?\;:\"\"

...
Files themselves - when they’re over 2GB...
Files themselves - when they’re over 2GB... get a little mangled

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Changed</th>
<th>Rights</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chloropipo_322549_RawData.tar</td>
<td>2,097,088 KB</td>
<td>4/15/2019 3:37:27 PM</td>
<td>rw-r------</td>
<td>emu</td>
</tr>
<tr>
<td>Chloropipo_322549_RawData.tar#001</td>
<td>2,097,088 KB</td>
<td>4/15/2019 3:38:44 PM</td>
<td>rw-r------</td>
<td>emu</td>
</tr>
<tr>
<td>Chloropipo_322549_RawData.tar#002</td>
<td>2,003,434 KB</td>
<td>4/15/2019 3:39:52 PM</td>
<td>rw-r------</td>
<td>emu</td>
</tr>
<tr>
<td>Chloropipo_322549_ReadMe.txt</td>
<td>76 KB</td>
<td>4/15/2019 3:36:09 PM</td>
<td>rw-r------</td>
<td>emu</td>
</tr>
<tr>
<td>Chloropipo_322549_ScanSheet.pdf</td>
<td>190 KB</td>
<td>4/15/2019 3:36:09 PM</td>
<td>rw-r------</td>
<td>emu</td>
</tr>
</tbody>
</table>

MD5 Checksum: 24126b4a42781646b514d62f7b8145eb

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Usage</th>
<th>MIME Type</th>
<th>MIME Format</th>
<th>Width (pixels)</th>
<th>Height (pixels)</th>
<th>File Size (bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chloropipo_322549</td>
<td>Research</td>
<td>application</td>
<td>pdf</td>
<td>193851</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chloropipo_322549</td>
<td>Research</td>
<td>text</td>
<td>plain</td>
<td>75571</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chloropipo_322549</td>
<td>Research</td>
<td>application</td>
<td>x-tar</td>
<td></td>
<td></td>
<td>6346352128</td>
</tr>
</tbody>
</table>
Files themselves - sometimes go missing

...into the mystery-void between EMu & storage → [broken icon]
Files themselves - sometimes go **missing**

...into the mystery-void between EMu & storage → [broken icon]

- On each side of the void is an audit trail, (EMu audit + Filer log)
- but without cross-checking between systems…
On each side of the void is an audit trail (EMu audit + Filer log)

- but without cross-checking between systems...

- [Broken icon]

- On each side of the void is an audit trail (EMu audit + Filer log)

- but without cross-checking between systems...

github.com / fieldmuseum / Collections-Scripts
Files themselves...

...get re-re-dup-duplicated

- sometimes folks don’t look before they leap
- Or they do, but their workflows get in the way
  - (e.g., checking for dups by name...after renaming dups)
Keywords

Out-of-sync duplicated data
Keywords: Audubon Core

Out-of-sync duplicated data

Mapping reverse-attached fields to Audubon Core:

- Taxon Coverage
- Vernacular Name
- Related Geography
- Temporal Coverage
- Sex
- Life Stage
- Subject Orientation
- Subject Part
- Capture Device
- Access URI
- Related Resource ID
- Derived From
- Funding Attribution
- Identifier
- Associated Specimen
- Associated Observation
2018 - 2019

...taking stock of 5 year’s worth of Multimedia + 1 year of cleanup:

- File formats… “stabilized”?
  - XLSX → open-docs formats; still need conversion tools for DNGs

- Filenames… renamed
  - “DÈRP’s b.123(4).jpg” → sh*ty-file-renamer.pl → “DRPs_b_123_4_.jpg”

- Files themselves…
  - sometimes go missing → check/compare logs between EMu & filer
  - Duplicates → backlog deleted + uniqueness on ChaMD5sum

- Keywords… mapped Audubon Core reverse attachments
  - there is less yelling...for now...

[the silence of the DAMS?]
2019 - Setting up the DAMS server...

So far, we’ve:

- set up the DAMS server (“Palmer”)
- pointed it at our EMu server (“Shackleton”)
- […testing stuff… (users / eevents / photo archives workflow?) ]
All clear going forward?

Or just the calm before the next storm?

(And who exactly is ‘Hannibal Lecter’ in this scenario?)

(And how many different misspellings of “DAMS” will there be?)
Citations/Sources

- Clarice EMu / JodiEMu Foster - https://animals.sandiegozoo.org/animals/emu
- Nathaniel B Palmer -
- Ernest J Shackleton -
- Shackleton crew with puppies -
- Silence of the Lambs poster - © Orion Pictures
- Screaming Lambs - https://twitter.com/screaminglxmb
- Screaming Charlton Hestons - Soylent Green - © Metro-Goldwyn-Mayer
- Death’s Head Moth - http://twmoth.tesri.gov.tw/peo/MothInfo/V19-20150515-174
  - (Acherontia lachesis - occurrenceID: V19-20150515-174)
Questions, Comments, Thought?

(No real lambs were harmed in the making of this talk.)
the silence of the DAMS

Jodie Foster = EMu?