How do you solve a problem like multimedia?

Sharon Grant, Kate Webbink, and Marc Lambruschi
When the World’s Columbian Exposition closed in the autumn of 1893, and its 25 million visitors had gone home, all that remained on the fairgrounds was the Palace of Fine Arts. A year later, the Columbian Museum of Chicago opened. The Chicago Times reported “It was, all like a memory of the fair.”

(http://www.encyclopedia.chicagohistory.org/pages/450.html)
The museum's collections quickly outgrew the space in the old Palace, and in first decades of the twentieth century plans for a new museum building began to take shape. The result, a massive white marble building rising from the mud. The Field Columbian Museum opened to 8,000 visitors in May 1921.

(http://www.encyclopedia.chicagohistory.org/pages/450.html)
The Depression, The Great War, the attack on Pearl Harbour and another name, none of these change The Museum’s goals “Accumulation and dissemination of knowledge, and the preservation and exhibition of objects illustrating art, archaeology, science and history.”
The Field Museum of Natural History.

Into the 21st century, a new focus on scientific research increases The Field Museum of Natural History’s collections to 24 million and 186,000 square feet of storage space is created 2 floors beneath the ground.
What’s all this fuss about digitization?
The Impossible Dream?

Had you noticed that there is an effort underway to digitize collections?
To capture specimen data fast enough to digitise entire collections while maintaining data quality.
   To store and retrieve the assets created.
   To share the information held in these collections.

How do you do this?
   Can we do this?
   Should we do this?

Can we realize the impossible dream?
Intermission
Imaging Frenzy

What to do when the bunnies are multiplying and eating all the grass?

\[ y = 31.843e^{0.0663x} \]

\[ R^2 = 0.5107 \]
Storage Crisis

And we asked for more space and the answer was NO.
Plus our... re-indexing hell

Reindex - Starting Sun Jun 15 01:00:01 2014
registry... # of records updated = 12895
accessionlots... Records updated : 0
ebibliography... Records updated : 0
ecatalogue... Increased size to 2883905 records # of records updated = 2879985
ecollectionevents... Increased size to 1400746 records # of records updated = 1390746
econdition... # of records updated = 7935
econservation... # of records updated = 38424
edocuments... # of records updated = 19748
eevents... Increased size to 2477 records # of records updated = 2252
eexports... Increased size to 50 records # of records updated = 2
efieldhelp... # of records updated = 5
efmnhrepatriation... # of records updated = 5
efmnhtransactions... # of records updated = 48
egazetteer... # of records updated = 7629
egroups... # of records updated = 3894
einurance... Increased size to 50 records # of records updated = 2
einternal... Increased size to 190197 records # of records updated = 180197
eloans... # of records updated = 136
elocations... # of records updated = 0
eparties... # of records updated = 28
erights... Records updated : 118
eright... Records updated : 0
eschedule... Records updated : 0
esites... Records updated : 13
etript... Records updated : 0
estatistics... Records updated : 0
etaxonomy... Records updated : 95
etemplate... Records updated : 0
etheosaurus... Records updated : 0
etransactions... # of records updated = 0
evaluations... Records updated : 0
ewebgroups... Records updated : 0
ewebusers... Records updated : 0
Finished Sat May 17 00:28:04 2014

Reindex audit - Starting Sun Jun 15 01:00:01 2014
Starting system maintenance at Sat Sep 28 22:00:02 CDT 2013
Starting compact maintenance at Sat Sep 28 22:00:02 CDT 2013
Checking for index updates...
Closing database...
Skipped compact...
Reconfiguring database...
Increased size to 111894773 records
Reindexing...
Opening database...
Starting background loads...
Finished compact maintenance at Sat Sep 28 22:00:48 CDT 2013
Finished system maintenance at Sat Sep 28 22:01:11 CDT 2013
Finished Sun Jun 15 14:26:59 2014

Finished Sun Jun 15 14:26:59 2014
The role of multimedia in this clip is played by a fuzzy white bunny.
It wasn’t quite the wild, wild west...
Intermission
HOW DO YOU SOLVE A PROBLEM LIKE MULTIMEDIA
Is a Puzzlement

We started looking in three places:

The files:
Think about content and format.

The back-end:
Review hardware and storage.

The front-end:
Assess workflows and standards.

TWDG 2014: Sharon Grant, Kate Webbink, Marc Lambruschi, Mike Yoshida – The Field Museum of Natural History. Chicago. IL
File formats
File Formats - Layers

Volumize with LAYERS

Layers allow flexible editing.

Merging of images.

Addition of reference details and scale bars.

Watermarking and security.
File Formats – Layers

Each new layer added to an image increases the size of the file.

This picture consists of a blue background and on top of that a layer of conifers cut using a layer-mask in the shape of a seagull.

The blue sky used as background layer.

The greenery used as top layer.

The layer mask that was applied to the top layer. The white shape is a mask over the black.
The Benefits of BITS

More bits = More colours

More Colours = More Information

More Information = More Data

More Data = More File Space
<table>
<thead>
<tr>
<th></th>
<th>8-bit</th>
<th>16-bit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Approximately 16 million tone variations</td>
<td>Approximately 281 trillion tone variations</td>
</tr>
</tbody>
</table>
Adelotopus parumpunctatus
Archiving S&E Imagery

- Layers
  - 8-bit: x MB
  - 16-bit: 4x MB

+ Layers
  - 8-bit: 2x MB
  - 16-bit: 8x MB

Color (bit-depth)

x 2
File Formats – Colour Depth

8-bit approx. 16 million tone variations; 16-bit approx. 281 trillion tone variations

TWDG 2014: Sharon Grant, Kate Webbink, Marc Lambruschi, Mike Yoshida – The Field Museum of Natural History. Chicago. IL
File Formats – TIFFs Vs DNGs

We archive TIFFS but they are **HUGE** is there anything else we could do?
File Formats – TIFFs Vs DNGs

Four Things About DNGs

Firstly the acronym. DNG = Digital Negative (A two word TLA).

Secondly DNG is “open source” and was created by Adobe as a way to standardize RAW file formats.

Thirdly DNGs are smaller than the TIFF that was used to create them.

Fourthly DNGs retain all the information in the TIFF AND the image is identical.
File Formats – The upshot

- 16-bit TIF working file + layers
- Archival DNG file size from 16-bit TIF
- 8-bit TIF + flattened layers
- Archival DNG file size from 8-bit TIF
Establishing Standards

TIFF working files; DNG archive files

DPI - 300dpi

ICC Color Profiles – AdobeRGB (& sRGB for Web)

Dimensions – < 10,000 x 10,000

Compression – None

8-bit (Approx. 16 million tone variations)

Layers – None

TWDG 2014: Sharon Grant, Kate Webbink, Marc Lambruschi, Mike Yoshida – The Field Museum of Natural History. Chicago. IL
Storage - remember to include Backups

Ingredients:

3 copies of any important file (a primary and two backups)
on 2 different media types (such as hard drive and optical media), to protect against different types of hazards.

1 copy offsite (or at least offline).
## Storage - Costs

<table>
<thead>
<tr>
<th>Solution</th>
<th>Cost per TB for 3-2-1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAN + ExoGrid backup solution</td>
<td>$7,500</td>
<td>$2,490,000</td>
</tr>
<tr>
<td>External Hard-drives</td>
<td>$300</td>
<td>$99,600</td>
</tr>
<tr>
<td>Secret Online Sauce (shhh...)</td>
<td>$3,000</td>
<td>$996,000</td>
</tr>
<tr>
<td>Amazon Cloud Services</td>
<td>To hard to calculate!</td>
<td>Who knows?</td>
</tr>
</tbody>
</table>

Currently storing approx. 82 TB

150 TB is committed for digitization projects over 12-18 months (thanks NSF).

100 TB identified as under desks.
Storage - Considerations

Reimagining Storage

**THE 5Vs**

*Velocity* – Speed of upload and download

*Volume* – Scalability and efficiency

*Value* – Affordability and costs

*Visibility* – Availability to users

*Vigilance* – Security and backup
## Storage - Considerations

<table>
<thead>
<tr>
<th>Solution</th>
<th>Velocity</th>
<th>Volume</th>
<th>Value</th>
<th>Visibility</th>
<th>Vigilance</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAN + ExoGrid backup solution</td>
<td>High</td>
<td>M/H*</td>
<td>Low*</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>External Hard-drives</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Secret Online Sauce (shhh...)</td>
<td>M/L**</td>
<td>High</td>
<td>Medium</td>
<td>High**</td>
<td>High</td>
</tr>
<tr>
<td>Amazon Cloud Services</td>
<td>M/L**</td>
<td>High</td>
<td>Low???</td>
<td>High**</td>
<td>High</td>
</tr>
</tbody>
</table>

Warning: Back of a cigarette packet assessments!
The Server
The Server - Upgrading

**TOBY ($18,568.41)**
*Live server specifications - Purchased 2009*

**Base Unit:** DELL PowerEdge R710 with Chassis for Up to Eight 2.5-Inch Hard Drives

**Processor:** PowerEdge R710

**Memory:** 12GB Memory (6x2GB), 1333MHz Dual Ranked UDIMMs for 2 Processors, 50GB Solid State Drive SATA Mainstream 2.5in HotPlug Hard Drive, RAID 5 for H700 or PERC 6/i Controllers, SSD Hard Drives

**Hard Drive:** 50GB Solid State Drive SATA Mainstream 2.5in HotPlug Hard Drive. [Hard Drive Controller: PERC 6/i SAS RAID Controller 2x4 Connectors, Internal, PCIe256MB Cache, x8 Chassis]

---

**ROSS ($48,513.01)**
*Live server specifications - Purchased 2014*

**Base Unit:** Cisco C240

**Processor:** Dual Intel Xeon 3.5 GHz E5-2643

**Memory:** 512GB of DDR3-1866-MHZ LR DIMM/PC3 RAM

**Hard Drive:** Dual 300GB 10K RPM SAS drives for OS and a single 785GB MLC FusionIO drive
The Server - Upgrading

This is Toby our old server

This is Ross our new server
The Server – The Results

Batch inc eaudit – Start Fri Oct 3 23:00:01 CDT 2014

eaccessionlots... Records updated : 0
ebibliography... Records updated : 0
eckatolgue... Records updated : 2213
econdition... Records updated : 3
econservation... Records updated : 12
edocuments... Records updated : 0
eevents... Records updated : 1
exports... Records updated : 0
efieldhelp... Records updated : 2
efmnhrapatiation... Records updated : 0
efmntractions... Records updated : 0
egazetteer... Records updated : 0
egroups... Records updated : 0
einsurance... Records updated : 0
einternal... Records updated : 704
eloans... Records updated : 1
elocations... Records updated : 0
eluts... Records updated : 75
emovements... Records updated : 0
eports... Records updated : 0
enarratives... Records updated : 0
enhmstratigraphy... Records updated : 0
eparties... Records updated : 10
erights... Records updated : 0
eschedule... Records updated : 0
esites... Records updated : 6
estatistics... Records updated : 0
etaxonomy... Records updated : 0
ethemesaurus... Records updated : 0
etemplate... Records updated : 0
etriansactions... Records updated : 0
evaluations... Records updated : 0
ewebgroups... Records updated : 0
ewebusers... Records updated : 0

Finished Sat Oct 4 06:15:37 CDT 2014

Reindex – Start Sat Aug 2 22:00:01 CDT 2014

eregistry... Number of records updated = 13321
eaccessionlots... Number of records updated = 4405
ebibliography... Number of records updated = 33116
ectatolgue... Increased size to 2913319 records Number of records updated = 2903319
econdition... Increased size to 1404107 records Number of records updated = 1394107
econservation... Increased size to 279641 records Number of records updated = 269641
edocuments... Increased size to 44400 records Number of records updated = 40364
edocument... Number of records updated = 1973
ereports... Number of records updated = 244
exports... Increased size to 50 records Number of records updated = 49
efieldhelp... Number of records updated = 0
efmnhapatiation... Number of records updated = 11111
efmnhtractions... Number of records updated = 48531
egazetteer... Number of records updated = 1624
egroups... Number of records updated = 76747
eiinsurance... Increased size to 1065828 records Number of records updated = 1057828
eloans... Number of records updated = 3188
elocations... Number of records updated = 7624
eparties... Number of records updated = 10
erights... Increased size to 50 records Number of records updated = 11
eschedule... Increased size to 50 records Number of records updated = 4
esites... Increased size to 452751 records Number of records updated = 442751
estatistics... Number of records updated = 78612
etaxonomy... Increased size to 555903 records Number of records updated = 545903
etemplate... Increased size to 50 records Number of records updated = 6
ethemesaurus... Increased size to 907026 records Number of records updated = 897026
etriansactions... Increased size to 50 records Number of records updated = 0
 evaluation... Number of records updated = 1361
etemplates... Increased size to 100 records Number of records updated = 84
ewebusers... Increased size to 100 records Number of records updated = 82

Finished Sun Aug 3 01:42:06 CDT 2014

Reindex audit – Start Sun Oct 5 22:00:01 CDT 2014

eaudit... Increased size to 48095230 records Number of records updated = 48095230
eregistry... Number of records updated = 13321
eaccessionlots... Number of records updated = 4405
ebibliography... Number of records updated = 33116
ectatolgue... Increased size to 2913319 records Number of records updated = 2903319
econdition... Increased size to 1404107 records Number of records updated = 1394107
econservation... Increased size to 279641 records Number of records updated = 269641
edocuments... Increased size to 44400 records Number of records updated = 40364
edocument... Number of records updated = 1973
ereports... Number of records updated = 244
exports... Increased size to 50 records Number of records updated = 49
efieldhelp... Number of records updated = 0
efmnhapatiation... Number of records updated = 11111
efmnhtractions... Number of records updated = 48531
egazetteer... Number of records updated = 1624
egroups... Number of records updated = 76747
eiinsurance... Increased size to 1065828 records Number of records updated = 1057828
eloans... Number of records updated = 3188
elocations... Number of records updated = 7624
eparties... Number of records updated = 10
erights... Increased size to 50 records Number of records updated = 11
eschedule... Increased size to 50 records Number of records updated = 4
esites... Increased size to 452751 records Number of records updated = 442751
estatistics... Number of records updated = 78612
etaxonomy... Increased size to 555903 records Number of records updated = 545903
etemplate... Increased size to 50 records Number of records updated = 6
ethemesaurus... Increased size to 907026 records Number of records updated = 897026
etriansactions... Increased size to 50 records Number of records updated = 0
 evaluation... Number of records updated = 1361
etemplates... Increased size to 100 records Number of records updated = 84
ewebusers... Increased size to 100 records Number of records updated = 82

Finished Mon Oct 6 06:51:40 CDT 2014

13 hours per week

TWDG 2014: Sharon Grant, Kate Webbink, Marc Lambruschi, Mike Yoshida – The Field Museum of Natural History. Chicago. IL
EMu-Magic

In the above clip please substitute EMu for the word OZ
The broom stick plays the role of digitized specimen data with problems.
The role of the faithful collection manager is played by Dorothy
How exactly do you get help some multimedia help around here?
Know thy system.

Empower your database

Match your server to your system.

Change your workflow NOT your database.

Change your database NOT your workflow.

Connect to other peoples data and systems.
ImageMagick® is a software suite to create, edit, compose, or convert bitmap images. It can read and write images in a variety of formats (over 100) including DPX, EXR, GIF, JPEG, JPEG-2000, PDF, PNG, Postscript, SVG, and TIFF. Use ImageMagick to resize, flip, mirror, rotate, distort, shear and transform images, adjust image colors, apply various special effects, or draw text, lines, polygons, ellipses and Bézier curves. - http://www.imagemagick.org/

dcraw- (pronounced "dee-see-raw"), has become a standard tool within and without the Open Source world. It's small (about 9000 lines), portable (standard C libraries only), free (both "gratis" and "libre"), and when used skillfully, produces better quality output than the tools provided by the camera vendor. - http://www.cybercom.net/~dcoffin/dcraw/
EMu – ImageMagick

File formats available through ImageMagick

<table>
<thead>
<tr>
<th>File Extension</th>
<th>Codec</th>
<th>Read/Write/Multi page</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>RAW</td>
<td>+</td>
<td>Raw alpha samples</td>
</tr>
<tr>
<td>AVS</td>
<td>AVS</td>
<td>-</td>
<td>AVS x image</td>
</tr>
<tr>
<td>B</td>
<td>RAW</td>
<td>+</td>
<td>Raw blue samples</td>
</tr>
<tr>
<td>BIE</td>
<td>JPEG</td>
<td>-</td>
<td>Joint Bi-level image experts Group interchange format (1.5)</td>
</tr>
<tr>
<td>BMP</td>
<td>BMP</td>
<td>-</td>
<td>Microsoft Windows bitmap image</td>
</tr>
<tr>
<td>C</td>
<td>RAW</td>
<td>+</td>
<td>Raw cyan samples</td>
</tr>
<tr>
<td>CMYK</td>
<td>CMYK</td>
<td>-</td>
<td>Raw cyan, magenta, yellow, and black samples</td>
</tr>
<tr>
<td>CMYK</td>
<td>CMYK</td>
<td>-</td>
<td>Raw cyan, magenta, yellow, black, and opacity samples</td>
</tr>
<tr>
<td>CUR</td>
<td>CUR</td>
<td>-</td>
<td>Microsoft icon</td>
</tr>
<tr>
<td>CUT</td>
<td>CUT</td>
<td>-</td>
<td>DR Halo</td>
</tr>
<tr>
<td>DCM</td>
<td>DCM</td>
<td>-</td>
<td>Digital Imaging and Communications in Medicine image</td>
</tr>
<tr>
<td>PPC</td>
<td>PPC</td>
<td>-</td>
<td>Soft Imaging, multi-image Paintbrush</td>
</tr>
</tbody>
</table>
EMu - ImageMagick

Using EMu to empower your ASSETS

EMu uses ImageMagick libraries to handle multimedia.

ImageMagick is NOTHING to do the KE Software!

ImageMagick sometimes uses other programs called delegates to handle files.

dcraw is the delegate that handles DNGs.
EMu - ImageMagick

TIFF->DNG

TIFF->JPG (photoshop)

DNG->JPG (EMu)
EMu - ImageMagick

<table>
<thead>
<tr>
<th>File Size/Type</th>
<th>Conversion</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>7MB (Nikon raw --&gt; DNG)</td>
<td>yes</td>
<td>10s</td>
</tr>
<tr>
<td>25MB (Canon raw --&gt; DNG)</td>
<td>yes</td>
<td>20s</td>
</tr>
<tr>
<td>70MB (Mamiya raw --&gt; DNG)</td>
<td>yes</td>
<td>40s</td>
</tr>
<tr>
<td>18MB raw --&gt; DNG</td>
<td>yes</td>
<td>20s</td>
</tr>
<tr>
<td>25MB TIF --&gt; DNG</td>
<td>yes</td>
<td>10s</td>
</tr>
</tbody>
</table>

3000 25MB DNGs takes 8.5 hours!
EMu - Work flow

The Supplementary Tab
EMu - Client Changes

The Multimedia Tab
EMu - Client Changes

The Associations Tab
Know thy system - Emu Connections
The whiteboard challenge

Everything is easier on a whiteboard!
Finale
Thank-you

The roles of curators, collection managers, and volunteers working together to digitize specimens are played by Can-can dancers.

Thanks to Janeen Jones, Marc Lambruschi, Kate Webbink, Mike Yoshida, Rob Zschernitz and all the FMNH EMu digitizers.

TWDG 2014: Sharon Grant, Kate Webbink, Marc Lambruschi, Mike Yoshida – The Field Museum of Natural History. Chicago. IL
Things to Read.

DIGITISING COLLECTIONS

- [http://zookeys.pensoft.net/articles.php?id=2916](http://zookeys.pensoft.net/articles.php?id=2916)
- [http://soyouthinkyoucandigitize.wordpress.com/](http://soyouthinkyoucandigitize.wordpress.com/)
- [https://www.idigbio.org/content/gbif-training-manual-1-digitization-natural-history-collections-data](https://www.idigbio.org/content/gbif-training-manual-1-digitization-natural-history-collections-data)
- [http://images.library.amnh.org/digital/](http://images.library.amnh.org/digital/)
- [https://journals.ku.edu/index.php/jbi/article/view/3992](https://journals.ku.edu/index.php/jbi/article/view/3992)
- [http://www.gbif.org/resources/2775](http://www.gbif.org/resources/2775)

FILE FORMATS

- [http://www.mosaicarchive.com/2013/05/01/the-raw-truth-about-dng/](http://www.mosaicarchive.com/2013/05/01/the-raw-truth-about-dng/)
- [http://www.barrypearson.co.uk/articles/dng/respectability.htm](http://www.barrypearson.co.uk/articles/dng/respectability.htm)