From 0 to 400 GB:

Confronting the challenges of born-digital campus photographs

#saa16 - #s201
Panelists

- **Ed Busch**, Michigan State University
- **Chris Prom**, University of Illinois at Urbana-Champaign
- **Molly Tighe**, Chatham University
- **Greg Wiedeman**, University at Albany
- **Kristen Yarmey**, University of Scranton
Why are we here?

- Photograph collections are starting to look a *teensy* bit different.

Historic photograph collection storage, Virginia Tech
Why are we here?

- Born-digital photos present some new challenges...
Why are we here?

- ...but also some new opportunities.
Why are we here?

- We panelists have tried a few things. Some of them have worked.
- We thought we’d share them with you.
Time Keeping

● Panelists address key questions (10 min each):
  ○ How have you worked with the creator/donor of the materials?
  ○ How do you grapple with embedded metadata?
  ○ How have you leveraged automation/tools?
  ○ How do you provide access?

● Discussion/Q&A (15 min)
Ed Busch

Electronic Records Archivist
Michigan State University Archives & Historical Collections

@edbusch
MSU Overview

- MSU Archives established in 1969 by mandate of the Board of Trustees
- 1000s of analog photographs from the university (established in 1855)
  - Primarily University photographers
  - Others (faculty, staff, students, non-MSU)
- Digital photos arriving early 2000s
  - Floppies, Jaz/Zip, CD/DVD, Flash, file transfer, etc
- Created Trusted Digital Repository in 2013
  - Using Archivematica to ingest digital files
  - No access capability implemented
  - Have held off on mass ingesting of photos
- Planned Preservica evaluation during 2016-17
MSU Creators and Donors of Digital Photos

- Communications and Brand Strategy (CABS) office; primary source
  - 663GB received
  - 1.4TB on their site
  - On their servers?
- College of Human Medicine
  - 220GB received in 2008
- Other Colleges (22 total)
- Departments/Offices/Faculty etc.
- Athletics
  - Not transferred to archives
- Non MSU collections
MSU - Metadata

- What metadata!
- CABS (Campus photographers) generally will have caption on their favorites
- Sometimes files may be named with useful information
- I have used campus publications to add some metadata
- Face recognition software
MSU - Tools

- **Google Picasa**
  - Experimented using; face recognition on CHM photos
  - No longer supported by Google

- **Adobe Lightroom**
  - Added facial recognition function in 2015
  - Transferred Picasa facial metadata as keywords

- **Zenfolio**
  - Annual fee with unlimited storage
  - File organization
  - Search function
  - Storefront
MSU - Access

● Public Access
  ○ KORA (similar to ContentDM, developed by MSU’s MATRIX group)
  ○ FLICKR

● Internal
  ○ Zenfolio

● Future
  ○ Preservica
  ○ Islandora or Hydra
Kristen Yarmey

Associate Professor and Digital Services Librarian
University of Scranton

@kristenyt
University of Scranton Born Digital Photos

● **Terry and Paula Connors Photograph Collection** (1976-Present)
  ○ Archives/Special Collections hybrid (University and local events)
  ○ Physical/digitized/born digital hybrid
    ■ ≈100,000 negative strips, plus prints, slides
    ■ 1280+ photo CDs
    ■ 201,000+ JPGs (before deduplication)
  ○ Plus additional University images (on CD and server) held by PR - all duplicates?

● **PR Photos**
  ○ Physical/digitized/digital hybrid (slides, prints, born digital)
  ○ Digital photos on PR server, physical slides held by University Archives

● **Athletics Photos**
  ○ Digital photos on CDs held by Archives, also Athletics server - some duplicates?
The physicality of born digital
Beginner-Level Automation

- Ripstation!
  - Can create disk image or copy files only
  - Automate folder naming based on CD identifiers
Beginner-Level Automation

- File renaming with **FileRenamer**
  - Add CD identifier as prefix
  - Remove/replace special characters
- Deduplication with **VisiPics**
Digital Preservation

- Local copies synced to [DuraCloud](#) repository
  - 2 offsite backups
  - Amazon S3 and Glacier
Embedded Metadata

- **Adobe Bridge**
- Default EXIF data from camera
- **Date Taken! :D**
- Match sets of images to photo assignments
Not-So-Embedded Metadata

<table>
<thead>
<tr>
<th>Assignment Identifier</th>
<th>tpc-ja2005-008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title of Assignment</td>
<td>Fall Convocation and Awards, 2003</td>
</tr>
<tr>
<td>Description</td>
<td>Photographs of the University of Scranton's fall Convocation ceremony, including the presentation of Sursum Corda Awards (to Cheryl Boggs, Peggy Burke, and Valentine Gerigo) and the John L. Earl III Distinguished Service Award (to Joseph A. Fusaro). The event was held in the Royal Theatre.</td>
</tr>
<tr>
<td>Approximate Count</td>
<td>64 digital photos</td>
</tr>
<tr>
<td>Client</td>
<td>University of Scranton</td>
</tr>
<tr>
<td>Creator</td>
<td>Connors, Terry</td>
</tr>
<tr>
<td>Assignment Date</td>
<td>2005-09-09</td>
</tr>
<tr>
<td>Academic Year</td>
<td>2000-2005</td>
</tr>
<tr>
<td>Decade</td>
<td>2000-2009</td>
</tr>
<tr>
<td>Geographic Location</td>
<td>Scranton (Pa.)</td>
</tr>
<tr>
<td>Subject</td>
<td>Photographs; University of Scranton -- Convocations; University of Scranton -- Employees -- Awards; University of Scranton -- Faculty -- Awards; Fusaro, Joseph A.; Ballie, Harold W., 1950--; Royal Theater (Joseph M. McDade Center for Literary and Performing Arts, Scranton, Pa.); Boggs, Cheryl Y.; Filorzi, Scott R., 1959--; Stained glass windows;</td>
</tr>
</tbody>
</table>
Access

- Photos available!
  - ...upon researcher request

  - Novena with Bishop Martino, 2005
    - 2005-07-26
    - 30 digital photos
    - Digital photos may be made available upon researcher request. To submit a request, contact archives@scranton.edu with the Assignment Identifier.

  - Vascular Lab, 2005
    - 2005-07-27
    - 14 digital photos
    - Digital photos may be made available upon researcher request. To submit a request, contact archives@scranton.edu with the Assignment Identifier.

  - Trauma Center donation, 2005
    - 2005-07-28
    - 6 digital photos
    - Digital photos may be made available upon researcher request. To submit a request, contact archives@scranton.edu with the Assignment Identifier.
Access Issues: Bonus Round

● Browsability
  ○ What’s the best way to browse through hundreds of photos of the same thing?

● Selectivity
  ○ Which photo was the final PR-approved, publicized version?

● Privacy
  ○ Photos of community members at Mass, social events
  ○ Less-than-photogenic head shots, goofy candid shots
  ○ Face recognition

● Sensitive subject matter

● Rights (non-University clients)
Somehow, Someday, Somewhere

- Better, faster, automated (authenticated?) access
- More automated description
- Better integration with digitized photographs, negatives, slides
- Closer collaboration with PR on incoming photographs (in progress)
Building a Unified Image Delivery Service

Chris Prom
University of Illinois at Urbana-Champaign

@chrisprom
prom@illinois.edu
University of Illinois Archives

- Established 1963
- Unit of University Library
- Mixed Repository: Digital Holdings today
  - ~12 TB digital under active management (processed)
  - ~ 3 TB digital backlog
  - Short term commitment to bit level preservation
  - Format profiling and testing
- History of series ("collection") level control to deal with bulk
Aerial Photograph of Memorial Stadium During a Game

Title: Aerial Photograph of Memorial Stadium During a Game
Date: Oct. 19, 1963
Description: Aerial photograph of Memorial Stadium looking northwest with the surrounding area visible including Assembly Hall

Found in RS: 39/2/20, Box: BUI -Stadium, Folder: BUI- Stadium, 1960-63
Negative Number: 13323, Record Series 39/2/22.

Found in: Photographic Subject File, 1868
Rights: Copyright University of Illinois.

Copyright of this image is managed by the University Archives or we have been unable to identify the copyright holder. Please contact us if you would like to purchase a high-res copy of the image, to inquire about permissions for publication, or if you can help use identify the copyright holder.

Unique ID: 0001041

Additional Information:
Repository: University of Illinois Archives
Born Digital Experimentations

http://library.illinois.edu/archives/gallery
Meta Issues

- Bifurcated Descriptive and Access systems (Library-wide and in Archives)
- B-D Photos only one (smallish) issue
- Real need:
  - Method to rapidly ingest and provide access to multiple file types (photos and others)
  - Automated descriptive processes (to extent possible)
Emergent Description/Preservation/Access Environment

ARCHON™ UNIVERSITY OF ILLINOIS

ArchivesSpace

UNIVERSITY LIBRARY
Medusa

Digital Repository Service

Digital Collections

Digital Collections
Working with Creators

- Public Information Photograph File
  - 50,000+ analog photographs
  - Since 2004 → Born digital

- ACES Photograph File
  - College of Agriculture, Consumer, Environmental Sciences
  - 1.17 TB; ~ 50,000 photographs
  - David Riecks

- System Development: Repository Group
The Medusa collection registry provides a web-accessible management interface to preservation microservices built upon an enduring storage and management environment for the University of Illinois at Urbana-Champaign Library’s digital collections. Access to the collection registry is restricted to digital content producers at the University of Illinois Library. For more information, read the FAQ.

Quick Links

- Deposit Files
- Create a Collection
- Request Training
- Feedback

Learn More

- Policies
- Technology
- Staff

More information: Medusa at 217-333-4648 or medusa@library.illinois.edu or http://medusa.library.illinois.edu
<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Envelope Record Version</td>
<td>4</td>
</tr>
<tr>
<td>Coded Character Set</td>
<td>UTF8</td>
</tr>
<tr>
<td>Application Record Version</td>
<td>4</td>
</tr>
<tr>
<td>Object Name</td>
<td>ab00104a</td>
</tr>
<tr>
<td>Category</td>
<td>AB</td>
</tr>
<tr>
<td>Supplemental Categories</td>
<td>Slide Scan</td>
</tr>
<tr>
<td>Keywords</td>
<td></td>
</tr>
<tr>
<td></td>
<td>environment</td>
</tr>
<tr>
<td></td>
<td>ecology</td>
</tr>
<tr>
<td></td>
<td>ecosystem</td>
</tr>
<tr>
<td></td>
<td>environmentalism</td>
</tr>
<tr>
<td></td>
<td>scenery</td>
</tr>
<tr>
<td></td>
<td>nature</td>
</tr>
<tr>
<td></td>
<td>water</td>
</tr>
<tr>
<td></td>
<td>foam</td>
</tr>
<tr>
<td></td>
<td>pollution</td>
</tr>
<tr>
<td></td>
<td>pollute</td>
</tr>
<tr>
<td></td>
<td>polluted</td>
</tr>
<tr>
<td></td>
<td>water pollution</td>
</tr>
</tbody>
</table>
Embedded Metadata: Early Dublin Core Mapping

- IPTC Author: creator
- IPTC Credit: creator
- IPTC Creator: creator
- IPTC Date Created: dateCreated
- IPTC Description: description
- IPTC Caption: description
- IPTC Copyright: rights
- IPTC Copyright Notice: rights
- IPTC Keywords: subject
- IPTC Headline: title
- IPTC Title: title
Mapping Issues and Resources

- **EXIF vs. ITPC/IIM vs. XMP vs. Photoshop**
- Each photographer’s use different
- Metadata Stripping--common among sharing and image management services
- [http://photometadata.org/](http://photometadata.org/)
Descriptive Info

Creator
- David Riecks
- UIUC-ACES-ITCS: Riecks

Date
- 2012-01-18T00:00:00Z
- 2012:01:18

Description
A swirl of white and brown foam on top of water in a tank.

Rights
- UIUC-ACES-ITCS: Riecks
- UIUC-ACES-ITCS: Riecks

Subject
- abstract
- design
- conceptual
- concept
- concepts
- horizontal
- horizontals
- close-up
Digital Collections

Descriptive Info

Creator: David Riecks, UIUC-ACES-ITCS

Date: 2012-01-18T00:00:00Z
Date Created: 2012-01-18

Description: A swirl of white and brown foam on top of water in a tank.

Rights: UIUC-ACES-ITCS: Riecks, UIUC-ACES-ITCS: Riecks

Subject: abstract, design, conceptual, concept, concepts, horizontal, horizontals, close-up

https://kumquat-dev.library.illinois.edu
Tools

ExifTool by Phil Harvey

Photo Mechanic v5

International Image Interoperability Framework

Cantaloupe

High-performance dynamic image server in Java
Access Next Steps--Develop Metadata Profile and link to collection/aggregate description
Lessons

● Move to integrated system complex but doable
● Doing the right work in the right order really makes a difference; simple to complete
  ○ images, embedded MD, then externally managed MD
● Partnerships critical. Thanks go to:
  ○ Kyle Rimkus, Bill Ingram, Bethany Anderson, David Riecks, Alex Dolski, Patricia Lampron, MJ Han.
Greg Wiedeman

University Archivist
University at Albany, SUNY
@gregwiedeman
Born-Digital Photography at UAlbany

• Campus Photographer in Digital Media Department
  – 134 events in 2014, around 3-300 images per event

• Camera raw files (.NEF, .CR2)

• JPG derivatives

• Images go back to 1999
Disks in Boxes

- 4 boxes, 598 DVDs and CD-Rs
- 1.8 TB
- In folders by Job Number
- Subfolders have minimal description
- 1999-2008 Access Database
  - Has descriptions
- 2008-2012 REST DB
  - Dates, no descriptions
Born-Digital Photography at UAlbany

• Implemented SmugMug service in 2012
  – Online public photo database
  – Over 19,000 images

• Uploads and enters metadata in SmugMug
Immigration Symposium at Rockefeller Institute of Government


This symposium considers opportunities and challenges presented by immigration and national immigration policy reforms with a focus on states and localities, particularly NYS and New York’s cities. Featured guests include Michele Waslin, Pew Center, Neil Ruiz, Brookings Institution; and Jorge Montalvo, NYS Office of New Americans.
Principles

• Automation
  – Need to scale
  – No metadata creation, must describe themselves

• Standardization
  – Format-independent tools and utilities for born-digital records

• Transparency
  – Researchers need context

• Access
  – No restrictions, immediate public access
Crawling SmugMug

• Develop crawler for SmugMug
  – Download all images
  – Periodically crawl for updates
  – Hash index to see if already downloaded
  – Package into standard SIPS with metadata
  – After approval, automatically incorporate into EAD files and make publically available

github.com/UAlbanyArchives/ua395
def driveListen(drivePath, disk):

    while disk == False:
        try:
            # check if disk has size
            get_file_size(drivePath)
            disk = True
        except:
            disk = False

    # do stuff with disk
    print "Disk found, ripping disk"

    # image disk
    start = time.time()
    subprocess.call(['"dd", "if=/dev/" + driveName, 
                     end = time.time()
    elapsed = end - start
    print str(elapsed) + " seconds"
    print str(elapsed / 60) + " minutes"

    print "eject disk"
    subprocess.call(['"eject", driveName], shell=False)

    print "listen again"
    disk = false
    driveListen(drivePath, disk)

driveListen(drivePath, disk)
Carve files with fiwalk and icat (TSK)

Audit against fiwalk output

- Batch 1: 49646 of 50212 – 98.87%
- Batch 2: 47574 of 48030 – 99.05%
- Batch 3: 22436 of 24530 – 91.46%
- Batch 4: 49646 of 50212 – 98.87%
- Total: 169302 of 172984 – 97.87%

Convert with ImageMagik

Issues with Disk Imaging at Scale
Appraisal Decisions

• Not accept camera raw
  – Large, hard to make available
  – Proprietary

• Convert all files to JPG prior to accessioning
  – .CR2 Canon raw lossless or lossy JPG compression
  – .NEF Nikon proprietary lossless or lossy
  – 1.8 TB to 274 GB
  – Not using compression is not a preservation strategy

• Not spend time recovering files
Access

- New public access system
- Drupal, XTF, and static pages
- Bootstrap 3
- Schema.org
- Public domain
- Over 180,000 images

http://meg.library.albany.edu:8080/archive/view?docId=ua395.xml
Narrow Search

<table>
<thead>
<tr>
<th>Type of Materials</th>
<th>Subjects</th>
<th>People</th>
<th>Organizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital Files</td>
<td>Albany, New York</td>
<td>Blanchard, Duncan C</td>
<td>State University of New York, Atmospheric Sciences Research Center</td>
</tr>
<tr>
<td>Subjects</td>
<td>Conservation and the Environment</td>
<td>Schaefer, Vincent J</td>
<td>General Electric Company</td>
</tr>
<tr>
<td></td>
<td>Schenectady, New York</td>
<td>Vonnegut, Bernard</td>
<td>State University of New York at Albany</td>
</tr>
<tr>
<td></td>
<td>Literature</td>
<td>Moore, Charles B., 1920-2010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>more</td>
<td>more</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collection</th>
<th>Date Coverage</th>
<th>Extent</th>
<th>Subjects</th>
<th>Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duncan Blanchard Papers</td>
<td>1789-2003</td>
<td>35.8 cubic ft.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

http://meg.library.albany.edu:8080/archive/view?docId=ua395.xml
Digital Files

State University of New York at Albany, Office of Communication and Marketing

Repositories

University at Albany M. E. Grenander Department of Special Collections and Archives

Open to the public Monday-Friday, 9am-5pm
Closed for state holidays and winter intersession
Located on the top floor of the Science Library on the Uptown Campus
Directions

Digital Media Department Collection
Date Coverage: 2002-2016
Extent: 219.75 GB 164854 Digital Files
Types: Digital Files
Matches: 1 hit

...Bldg/ Red Room / Kurt Vonnegut 1/22/00 @ 5 PM reception...
http://meg.library.albany.edu:8080/archive/view?docId=ua395.xml
NYS Writer's Institute: 1/22/00 @ 3:30 PM State Author/Poet Award Ceremony Capitol Bldg/ Red Room / Kurt Vonnegut 1/22/00 @ 5 PM reception Elda's on Lark St. Photo session 1/22/00 @ 7 PM Public Reading/Reception Page Hall color print

Processing Details

2016 July 21 - Imaged from optical media with dd, carved from image with fswalk and icat, converted all raw to jpg
2016 July 21 - description record extracted from photographer's Microsoft Access database by Job number

http://meg.library.albany.edu:8080/archive/view?docId=ua395.xml
Molly Tighe

Archivist & Public Services Librarian
Chatham University
Collecting & Preserving Documentation of Campus-wide Green Building Initiatives for Teaching & Research

Working with Creators & Donors
Embedded Metadata
Automation and Tools
Access

Molly Tighe, Archivist & Public Services Librarian, Chatham University
@archimolly  mtighe1@Chatham.edu
Founded **1869** as Pennsylvania Female College

1890: Pennsylvania College for Women

1955: Chatham College (now University)

Enrollment: **2200**
Commitment to sustainability & green living.

Inspired by Rachel Carson (Class of 1929).
Completed
Phase 1 of 8

Expected completion: 2032 (20 year plan)
As the home of Falk School of Sustainability, Eden Hall Campus serves as a classroom on green living, green building, & sustainable technology.
Consortium on Digital Resources for Teaching & Research
Shared Shelf is a Web-Based Digital Asset Management System
How do you work with donors and/or creators of material?
Department of Finance & Administration

Contractors & Subcontractors

Progress Records and Reports
Challenges

It is not standard practice for contractors to receive requests from University Archives to obtain copies or their work records for use in classroom.
Chatham University is looking to forming strategic partnerships to ensure a broad base of documentation.

One such partnership is with the Carnegie Mellon University Create Lab, who use high-resolution, panoramic images of Eden Hall Camps taken at regular intervals by 180-degree cameras and stitch them together to create time-lapse videos of the campus construction.
So far…

579 aerial photographs taken at monthly intervals between 2009 and 2014

745 assorted images of construction progress & building infrastructure

…many still to collect.
Grappling with Embedded Metadata

**Extract:**
- Capture Date
- Measurements
- Anything good!

**Embed:**
- Visual Resources Association (VRA)
- VRA Import/Export Tool
- VRA Info Panel Plugin
VRA Info Panel Plugin

- **Creator**
- **Title**
- **Capture Date**
- **Repository Info**
- **Rights**
- **Description**
Import to Shared Shelf

Step 1: Upload Media

Step 2: Upload Data via Excel

In Development: Extract Metadata on Import
Linked & Local Vocabularies
Shared Shelf

Customizable metadata

Field-level user permissions

NDSA Level 1
How do you provide access?
Publish to ...

- DPLA
- Shared Shelf Commons
- IR
- OAI-PMH
- Omeka
- All ArtStor Subscribers
- Limit to Institutional Users
- Personal Collections
Customizable Metadata Mappings from Shared Shelf
Integration with Artstor Classifications
New Course: Eden Hall Field Experience for Interior Architecture Students
IAR231: Green and Sustainable Design
IAR630: Building Systems
IAR631: Design for Sustainability
SUS607: Applied Green and Social Innovation
SUS605: Leadership for Transitions to Sustainability
SUS302: Social Justice and Sustainability
FST640: Sustainable Community Development
CEM503: Green Chemistry
Thank you!

Molly Tighe, Archivist & Public Services Librarian, Chatham University
@archimolly  mtighe1@Chatham.edu
Questions and Discussion

(We’re here until 3:15pm)