The greatest control with the fewest steps: Automated TMS and DAM media ingest
Keep it Simple Stupid
Office of Digital Assets and Infrastructure

Digital Asset Management System – Phase 1

Yale University Art Gallery (TMS)
Yale Center for British Art (TMS)
Yale Peabody Museum (Emu)

Yale Libraries
Other Yale Content Providers

collective imagination
New York 2010
Objectives - Compromises

DAM - open APIs – connect Yale systems

Scalable enterprise system – grow with Yale’s needs

1 Standardize - 3 museums and future partners

DAM interface
Ingest process (automate TMS image linking/ DAM ingest)
Metadata models

Eliminate campus data silos – images/ metadata in the DAM

2 Business rules (YUAG/ YCBA)

TMS is the database of record for metadata
Metadata flows from TMS to the DAM, not back
DAM will be the Image source/ repository
Selected DAM – Tech Foundation

1. Artesia by Open Text - scalable enterprise system
   Open APIs, published API documentation – Yale systems connection
   Web based interface, written in java
   1 Image server, 1 SQL server for metadata, video requires additional server
   Artesia servers housed in central campus ITS server farms

2. 160 TB Isilon on-line storage with 160 TB on-line mirrored storage at remote site for nightly replication (reduce campus data silos).
   Museum access to DAM over 100mb network
   Artesia DAM managed by ODAI
   Servers/ storage managed by Yale ITS

Contracts signed 6/30/09
Reality bites....

Initial ingest design

Yale University Art Gallery (TMS)

Yale Center for British Art (TMS)

Yale University DAM ON-LINE
Digital Image storage
Metadata storage

Yale Peabody Museum
(Emu)
Yale University Art Gallery (TMS)

YUAG Photo Studio

YCBA Photo Studio

Yale Center for British Art (TMS)

Final TMS/ DAM ingest design

Yale University DAM
160 TB ON-LINE
Digital Image storage
Metadata storage
plus 160 TB on-line
mirrored
Capacity to grow
1. Java
2. jdbc sql driver
3. Tomcat
4. Firefox
5. ImageMagick
6. Adobe Flash 10
7. Adobe Flex
8. Robocopy
9. Velocity xml template
10. TMS media paths
11. New SQL login
12. New text types
13. New SQL views/tables in TMS db
14. Standardized image file name
15. Calculate checksums for error checking – MD5
17. Nightly Metadatasync
Standardized image file name

ag-obj-29883-003-pub.tif

only lower case

1 Museum: ag
2 TMS module: obj
3 ObjectID/ ExhibitionID: TMS sequential # field in table
4 Index: 001, 002, 003, 004, 021, 233, 0001, 0002, 0003, 0004, 0021, 0233
5 File type: bar, color corrected/ color bars included, mas, raw master for repository, pub, color corrected/ cropped; publishing
6 Suffix: jpg, tif
**Ingest steps/ workflow**

1. **Photographer**
   - Capture/ name image
   - Images
   - BridgeUI web page - Review image names
   - TMS

2. **Photo Tech**
   - Process images
   - Move to input folder
   - Input folder
   - Images stay here until marked and ingested
   - After IngestUI Processing, Images ready for ingest

3. **Designated Staff**
   - Process images for ingest
   - IngestUI web page – Images queued for TMS/ DAM ingest

4. **Thumbnailer**
   - Ingester
BridgeUI web page

Photographer
Capture image, name image

BridgeUI web page - Review image names

TMS

1. Review image names
2. TMS Lookup
   - Object: 6463
   - Id: 6463 / B1975.4.1382
   - Nymphs at a Roman Bath

<table>
<thead>
<tr>
<th>RP</th>
<th>Rendition</th>
<th>View</th>
</tr>
</thead>
<tbody>
<tr>
<td>6*</td>
<td>ba-obj-6463-0001-pub</td>
<td>See Artesia DAM</td>
</tr>
<tr>
<td>1</td>
<td>BA-TMS-6463-0001-DSP</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ba-obj-6463-0001-bar</td>
<td>See Artesia DAM</td>
</tr>
<tr>
<td>4</td>
<td>ba-obj-6463-0003-bar</td>
<td>See Artesia DAM</td>
</tr>
<tr>
<td>5</td>
<td>ba-obj-6463-0004-pub</td>
<td>See Artesia DAM</td>
</tr>
<tr>
<td>7</td>
<td>ba-obj-6463-0002-pub</td>
<td>See Artesia DAM</td>
</tr>
</tbody>
</table>

collective imagination
New York 2010
1. **Photographer**
   - Capture/ name image
   - BridgeUI
     - web page - Review image names
     - TMS

2. **Photo Tech**
   - Process images
   - Move to input folder
   - Input folder
     - Images stay here until marked and ingested
   - After IngestUI Processing, Images ready for ingest

   "Images queued for TMS/ DAM ingest"

   "Images ready for ingest"

   "After IngestUI Processing, Images ready for ingest"
3 IngestUI web page

1. Transferring data from bac@yale.yale.edu...

2. New Asset(s)
   - TMS ID: 1483
   - Module: obj
   - Include in queue: 
     - [ ]

3. Viewing item 4 of 5 entities, 14 assets total.

4. [Delete Selected Items]

5. [Rescan input folder] [Queue batch for ingest]
**Batch Queue is complete!**

14 assets queued for ingest in D:\Registrar_Images\ingest\input\queue.1267455833971

Close this browser window, or reload the page to try again.
Ingest steps

1. TMS
2. Primary
3. Public access
4. DAM

TMS
- Append media records
- Insert data
- Generate image derivatives
- Using ImageMagick
- Place screen & thumbnail in appropriate folders
- Link screen & thumbnail images to media records
- Mark PA box in media record, mark as primary image

DAM
- Build xml document from TMS Metadata, calculate checksum
- Build bagit bag
- Move bag to DAM hot folder
- Bag sits in DAM hot folder until DAM ingest process reads folder, ingests bag

Clean-up
- Move image to photo Studio archive folder
- Delete settings files & Temp. thumbnails
- Clean-up

Ingester checks for queue file every 3 minutes, process images using settings files.
TMS image storage

1. Open the folder E:\TMSmedia\Images\Screen

2. Navigate to the subfolder obj

3. Calculate: \text{obj} + 00 + 6463 = \text{obj} \ 006 \ 463

4. The images are stored in the subfolder \text{obj} \ 006 \ 463
Photo Studio Archive Copy

1. [Image]
2. [Image]
3. \text{obj} + 00 + 6463 = \text{obj \ 006 \ 463}
4. [Image]

[Collective Imagination New York 2020]
Guidelines for creating and moving standardized digital containers called ‘bags’.

Checksums – MD5
<?xml version="1.0" encoding="iso-8859-1" ?>
<yaleObject>
  <workOrder>
    <repositoryCode>YCBA</repositoryCode>
    <repositoryID>ba-obj-117-0007-bar.tif</repositoryID>
  </workOrder>
  <securityPolicies>
    <securityPolicy>ingest</securityPolicy>
  </securityPolicies>
  <culturalObject>
    <objectID>117</objectID>
    <accessionNumber>
      <![CDATA[ D1973.1.11 ]]>  
    </accessionNumber>
  </culturalObject>
  <title> Jama Masjid, Delhi </title>
  <date> 1811 </date>
  <beginYear>1811</beginYear>
  <endYear>1811</endYear>
  <creditline> Paul Mellon Collection </creditline>
  <objectOwner> Yale Center for British Art </objectOwner>
  <department> entomology </department>
  <classification> Visual Works </classification>
  <subclassifications> Paintings </subclassifications>
  <sourceClassification> Painting </sourceClassification>
</yaleObject>
1. Ingester Log

2. Avg #seconds per asset

3. Average asset size

4. Move bag to DAM

5. Total TMS/DAM

6. TMS ingest/ build bag

Total minutes: 3.7

Total time: 0 seconds

- Working on 1 of 10: ba-obj-5525-0001-bar.tif
  - Size: 2340044140
  - Checking for existing renditions matching ba-obj-5525-0001-bar
  - TMS Ingest
    - inserting data into media tables, module=obj, id=5525, primary=0
    - DAM Ingest
      - building bag 'ba-obj-5525-0001-bar.tif' in D:\Registrar_Images\ingest\work
        - contents: ba-obj-5525-0001-bar.tif, ba-obj-5525-0001-bar.xml
        - elapsed time: 6 seconds
        - executing pico copy "D:\Registrar_Images\ingest\work\1270747630704" "\artesia\dam01.it"
        - elapsed time: 84 seconds
      - moving asset to RR Queue 'D:\Registrar_Images\ingest\dam\obj\005\525'
      - cleaning up D:\Registrar_Images\ingest\input\ca-obj-5525-0001-bar.settings
      - cleaning up D:\Registrar_Images\ingest\input\ca-obj-5525-0001-bar.xml

- Working on 2 of 10: ba-obj-5525-0001-pub.tif
  - Size: 2340044140
  - Checking for existing renditions matching ba-obj-5525-0001-pub
  - TMS Ingest
    - inserting data into media tables, module=obj, id=5525, primary=1
    - DAM Ingest
      - building bag 'ba-obj-5525-0001-pub.tif' in D:\Registrar_Images\ingest\work
        - contents: ba-obj-5525-0001-pub.tif, ba-obj-5525-0001-pub.xml
        - elapsed time: 6 seconds
        - executing pico copy "D:\Registrar_Images\ingest\work\1270747630704" "\artesia\dam01.it"
        - elapsed time: 50 seconds
      - moving asset to RR Queue 'D:\Registrar_Images\ingest\dam\obj\005\525'
      - cleaning up D:\Registrar_Images\ingest\input\ca-obj-5525-0001-pub.settings
      - cleaning up D:\Registrar_Images\ingest\input\ca-obj-5525-0001-pub.xml

- Working on 3 of 10: ba-obj-60395-0001-bar.tif
  - Size: 235012544
  - Checking for existing renditions matching ba-obj-60395-0001-bar
  - TMS Ingest
    - inserting data into media tables, module=obj, id=60395, primary=0
    - DAM Ingest
      - building bag '1270747735470' in D:\Registrar_Images\ingest\work
        - contents: ba-obj-60395-0001-bar.tif, ba-obj-60395-0001-bar.xml
        - elapsed time: 6 seconds
        - executing pico copy "D:\Registrar_Images\ingest\work\1270747735470" "\artesia\dam01.it"
        - elapsed time: 33 seconds
      - moving asset to RR Queue 'D:\Registrar_Images\ingest\dam\obj\000\395'
      - cleaning up D:\Registrar_Images\ingest\input\ca-obj-60395-0001-bar.settings
      - cleaning up D:\Registrar_Images\ingest\input\ca-obj-60395-0001-bar.xml

- Working on 4 of 10: ba-obj-60395-0001-pub.tif
  - Size: 235012544
  - Checking for existing renditions matching ba-obj-60395-0001-pub
  - TMS Ingest
    - inserting data into media tables, module=obj, id=60395, primary=1
    - DAM Ingest
      - building bag '1270747784327' in D:\Registrar_Images\ingest\work
        - contents: ba-obj-60395-0001-pub.tif, ba-obj-60395-0001-pub.xml
        - elapsed time: 3 seconds
Ingest log email

<table>
<thead>
<tr>
<th>Filename</th>
<th>Status</th>
<th>Size (MB)</th>
<th>Secs</th>
<th>TMS?</th>
<th>DAM?</th>
<th>Bagnname</th>
</tr>
</thead>
<tbody>
<tr>
<td>ba-exb-500-0001-dsp.jpg</td>
<td>OK</td>
<td>1.9</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1271871901883</td>
</tr>
<tr>
<td>ba-exb-500-0002-dsp.jpg</td>
<td>OK</td>
<td>2.1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1271871903883</td>
</tr>
<tr>
<td>ba-exb-500-0003-dsp.jpg</td>
<td>OK</td>
<td>2.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1271871905446</td>
</tr>
<tr>
<td>ba-exb-500-0004-dsp.jpg</td>
<td>OK</td>
<td>2.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1271871907102</td>
</tr>
<tr>
<td>ba-obj-60395-0001-bar.tif</td>
<td>OK</td>
<td>251.0</td>
<td>50</td>
<td>1</td>
<td>1</td>
<td>1271871914461</td>
</tr>
<tr>
<td>ba-obj-60395-0001-pub.tif</td>
<td>OK</td>
<td>125.5</td>
<td>25</td>
<td>1</td>
<td>1</td>
<td>1271871961633</td>
</tr>
<tr>
<td>ba-obj-60395-0002-bar.tif</td>
<td>OK</td>
<td>251.0</td>
<td>48</td>
<td>1</td>
<td>1</td>
<td>1271871989149</td>
</tr>
<tr>
<td>ba-obj-60395-0002-pub.tif</td>
<td>OK</td>
<td>125.5</td>
<td>25</td>
<td>1</td>
<td>1</td>
<td>1271872034712</td>
</tr>
<tr>
<td>ba-obj-8036-0001-bar.tif</td>
<td>OK</td>
<td>267.9</td>
<td>56</td>
<td>1</td>
<td>1</td>
<td>1271872066837</td>
</tr>
<tr>
<td>ba-obj-8036-0001-pub.tif</td>
<td>OK</td>
<td>267.9</td>
<td>56</td>
<td>1</td>
<td>1</td>
<td>1271872122494</td>
</tr>
<tr>
<td>ba-obj-8036-0002-bar.tif</td>
<td>OK</td>
<td>267.9</td>
<td>58</td>
<td>1</td>
<td>1</td>
<td>1271872178541</td>
</tr>
<tr>
<td>ba-obj-8036-0002-pub.tif</td>
<td>OK</td>
<td>267.9</td>
<td>59</td>
<td>1</td>
<td>1</td>
<td>1271872236885</td>
</tr>
</tbody>
</table>
MetadataSync wakes up once every night, probes the DAM SQL db, extracts list of the auto-ingested assets, populates temp table in TMS.

MetadataSync reads 1st record in temp table, builds metadata xml document, calculates checksum, compares it to stored checksum from DAM.

If the checksums don’t match, MetadataSync builds a bag with only the xml document, sends the bag to the DAM.

If checksums match, MetadataSync discards the xml document and moves to the next record in the temp table.

Bags sit in DAM hot folder until DAM ingest process reads folder, ingests bags, updates DAM metadata.
Artesia went ‘live’ 4/6/2010, the Peabody is ingesting assets.


YUAG production ingest not yet determined.

tim.speevack@yale.edu
Yale University Art Gallery

david.parsell@yale.edu
Yale Center for British Art
Keep It Simple Stupid