Pixels to Purpose:
Transforming a Rights & Reproductions Department to Support a Sustainable Digital Collection
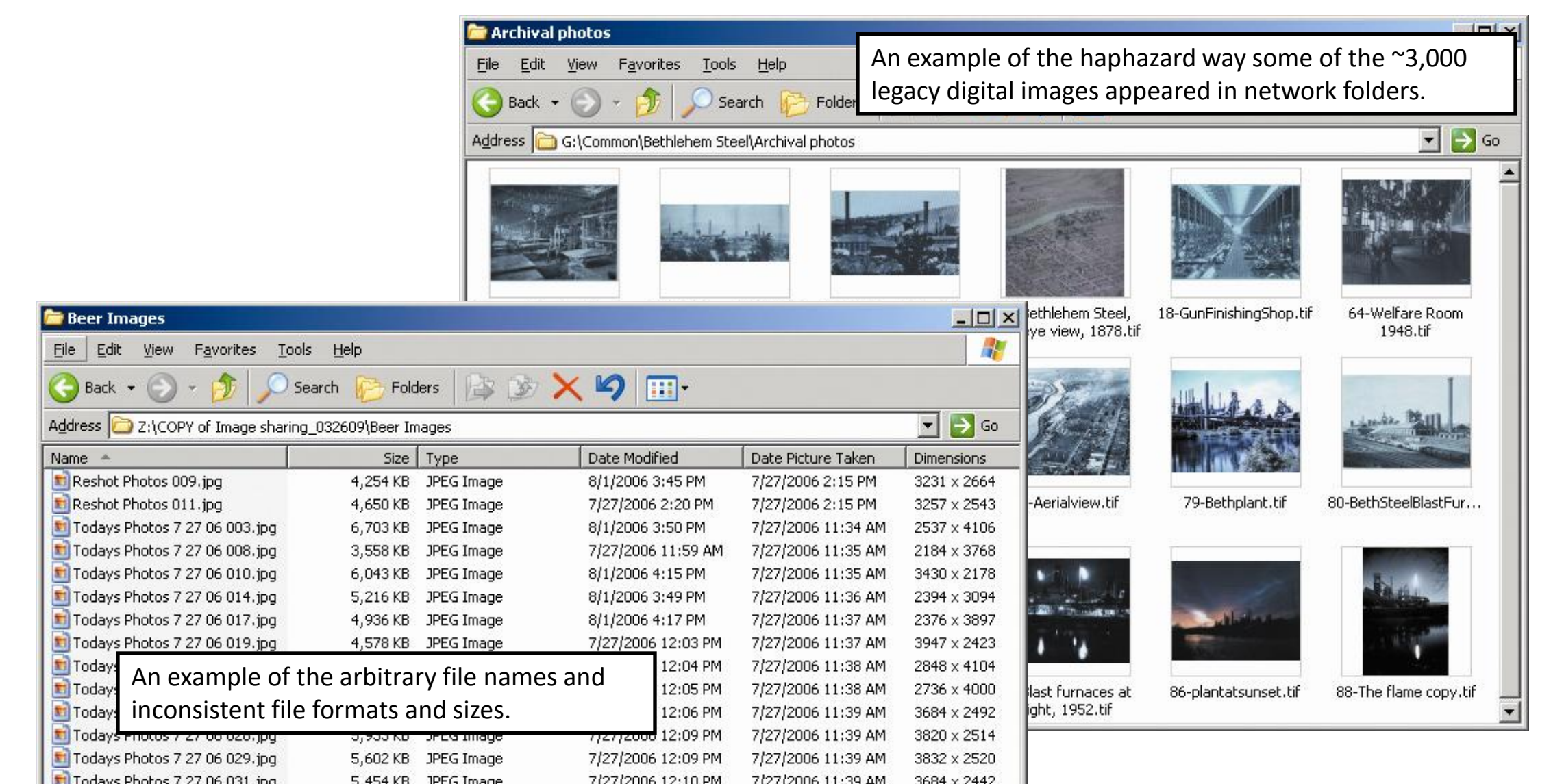### by Dana M. Lamparello

The Historical Society of Pennsylvania

Objective

To build a sustainable digital collection by repurposing digital images originally created to fulfill internal and external reproduction orders executed by the Historical Society of Pennsylvania's (HSP) Rights & Reproductions (R&R) Department.

Challenges

❖ Because R&R was not treating the amassing digital files as their own valuable collection, ~3,000 (out of over 4,500 total) legacy digital surrogates suffered from:
  - Inconsistent image quality
  - Little to no metadata
  - Arbitrary or unformatted file names
  - Inconsistent file formats
  - Lack of file storage structure

❖ Striking a balance between what is "appropriate" or minimal quality for continuing to fulfill ongoing R&R orders and what is "ideal" quality for building a sustainable digital collection.

An example of the haphazard way some of the ~3,000 legacy digital images appeared in network folders.

An example of the arbitrary file names and inconsistent file formats and sizes.

A record from the legacy MS Access database that minimally cataloged only ~1,500 of over 4,500 total images; note the catchall "Item Name" field that will require mining.

Steps in Assessing HSP's Legacy Digital Collection

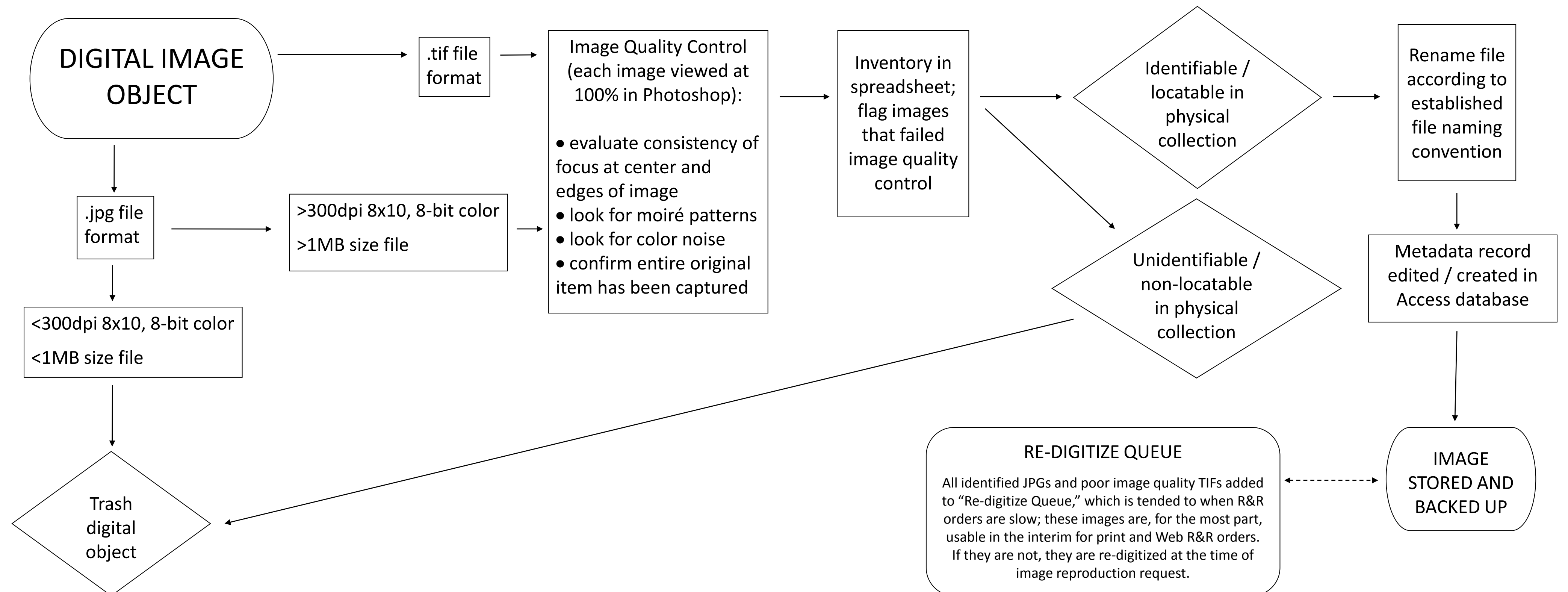1. Set quality standards for digital images:
   - **Minimal Model** → Quality appropriate for 75% of R&R orders (i.e., print and Web use)
     300dpi at ~ 8 x 10 inches, 8-bit color, JPG or TIFF

   - **Ideal Model**
     400-600dpi at ~ 8 x 10 inches, 24-bit color, TIFF

2. Design expanded Dublin Core-based metadata schema and modify legacy Microsoft Access database—which minimally cataloged ~1,500 of over 4,500 total images—to reflect edits.

3. Set standards for metadata capture:
   - **Minimal Model**
     Includes collection-level descriptive metadata and basic technical metadata.

   - **Ideal Model**
     Includes maximum descriptive, administrative, and technical metadata.

4. Inventory network folders and employee hard drives for usable digital image objects based on criteria detailed in flow chart:

The legacy digital image inventory spreadsheet. Only those files that met the criteria outlined in the flowchart were recorded; all others were permanently trashed.

Revised MS Access database that supports both minimal and ideal metadata capture.

Flow chart:
- DIGITAL IMAGE OBJECT
  - .tif file format → Image Quality Control (each image viewed at 100% in Photoshop):
    - evaluate consistency of focus at center and edges of image
    - look for moiré patterns
    - look for color noise
    - confirm entire original item has been captured
    → Inventory in spreadsheet; flag images that failed image quality control
  - .jpg file format
    - >300dpi 8x10, 8-bit color, >1MB size file
    - <300dpi 8x10, 8-bit color, <1MB size file → Trash digital object

- Identifiable / locatable in physical collection → Rename file according to established file naming convention → Metadata record edited / created in Access database → IMAGE STORED AND BACKED UP
- Unidentifiable / non-locatable in physical collection → RE-DIGITIZE QUEUE

RE-DIGITIZE QUEUE
All identified JPGs and poor image quality TIFs added to "Re-digitize Queue," which is tended to when R&R orders are slow; these images are, for the most part, usable in the interim for print and Web R&R orders. If they are not, they are re-digitized at the time of image reproduction request.

Next Steps

1. Tend to "Re-Digitize Queue," changing minimal model legacy digital images to reflect ideal quality standards (ongoing).

2. Edit and mine legacy metadata to reflect new metadata schema (ongoing).

3. Export metadata and digital images to CollectiveAccess, an open source digital asset management system, to improve internal and public access (January 2010).

Results

❖ 13% of the ~3,000 images (400 images) met above criteria and were not duplicates or low-resolution derivatives of master files.

❖ 400 usable images were added to digital collection and are ready to be searched and/or ordered for reproduction.