



The following document describes the Quality Control procedures implemented by **Purdue University Libraries Archives & Special Collections** to ensure and maintain the consistent accuracy, quality, and overall integrity of the digital archives. Quality Control is a critical step in the digital archiving process; these steps are necessary to verify that each digital image is an accurate representation of the original physical document.

Archived items are scanned according to certain image specifications, and the files are saved to the hard drive as uncompressed TIFF files. Within 48 hours of this digital capture, the files are subjected to the following procedures, verifying the accuracy and quality of each image, before backup to both CD and hard drive.

Purdue University Libraries Archives & Special Collections
Quality Control Procedures
for digital archives



<http://www.lib.purdue.edu/spcol>

© 2006 Purdue University, West Lafayette, Indiana
Last Revised: July 20, 2006

Prepared for in-house use by Liz Erlewine under the supervision of Carl Snow, Digital Initiatives Librarian, and Sammie Morris, Archivist.

Quality Control Procedures must be executed within 48 hours of scanning.

Please follow the instructions very carefully.

Digitization standards must be followed.

IMAGE SPECIFICATIONS

- All images must be 100% of their original size.
- Each image must be scanned at 600 dpi. (Usually resulting in a minimum of 3000 dots along the longest edge.)
- Text-only documents are scanned at 300 dpi.
- Images are scanned in 24 bit color (8,8,8 bits per sample).
- Images must be saved as uncompressed Tiff Images.

ITEM ID CONVENTIONS

Notice that ID and file names represent the collection, box, folder, and item numbers associated with that document.

Example: AEPb2f38i5

AEP is the collection code for the Amelia Earhart at Purdue collection

b = Box

f = Folder

i = Item

So the ID AEPb2f38i5 refers to Amelia Earhart at Purdue Box 2, Folder 38, item 5

MULTIPLE PAGES

If a document is composed of multiple pages or parts, then an additional number is added to the file name by adding the letter **P** (for page or part) and then the sequential number.

Example: b17f10i3p1 refers to Box 17, Folder 10, Item 3, Page 1

COPIES

If you come across multiple copies of the same document, number each item, but use the letter **C** (for copy) and then the sequential number. **You do not need to number the original document as a copy.**

Example: b17f10i4 refers to Box 17, Folder 10, Item 4 original
b17f10i4c2 refers to Box 17, Folder 10, Item 4, Copy 2
b17f10i4c3 refers to Box 17, Folder 10, Item 4, Copy 3

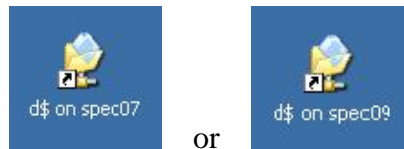
Remember, duplicate photographs may be printed at different sizes. Mark the original with the original ID number and all others as copies, regardless of size.

If the document appears to be numbered incorrectly, check with the Digital Initiatives Librarian or Archivist before making changes.

GETTING STARTED

Image files for Quality Control must be accessed from the two scanning stations: **SPEC07** and **SPEC09**.

To access one of these stations, you can click on the shortcut on your desktop (images shown below). If no shortcut is found, access the desired drive by going to **My Computer\My Network Places**.



Search through the folders on the **D:** drives of either **SPEC07** or **SPEC08** to find images that are ready for Quality Control.

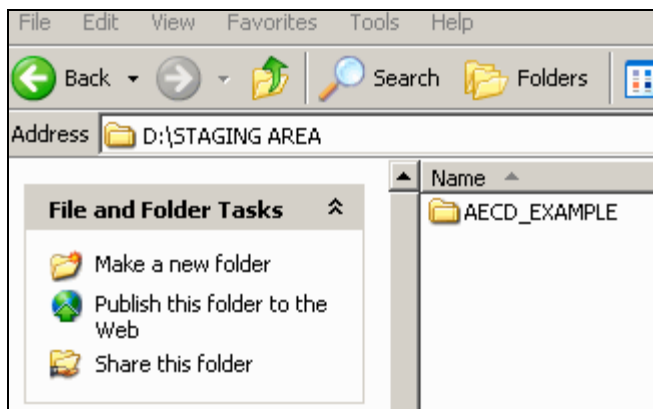
- **If the folder has TIF images in it, that means it has not been tested and is ready for Quality Control.**
- **If a folder *only* has the Inventory XLS sheet in it, then it has already been checked.**

Once you have found files to check, you need to move them to a **STAGING AREA** to begin Quality Control.

STAGING AREA

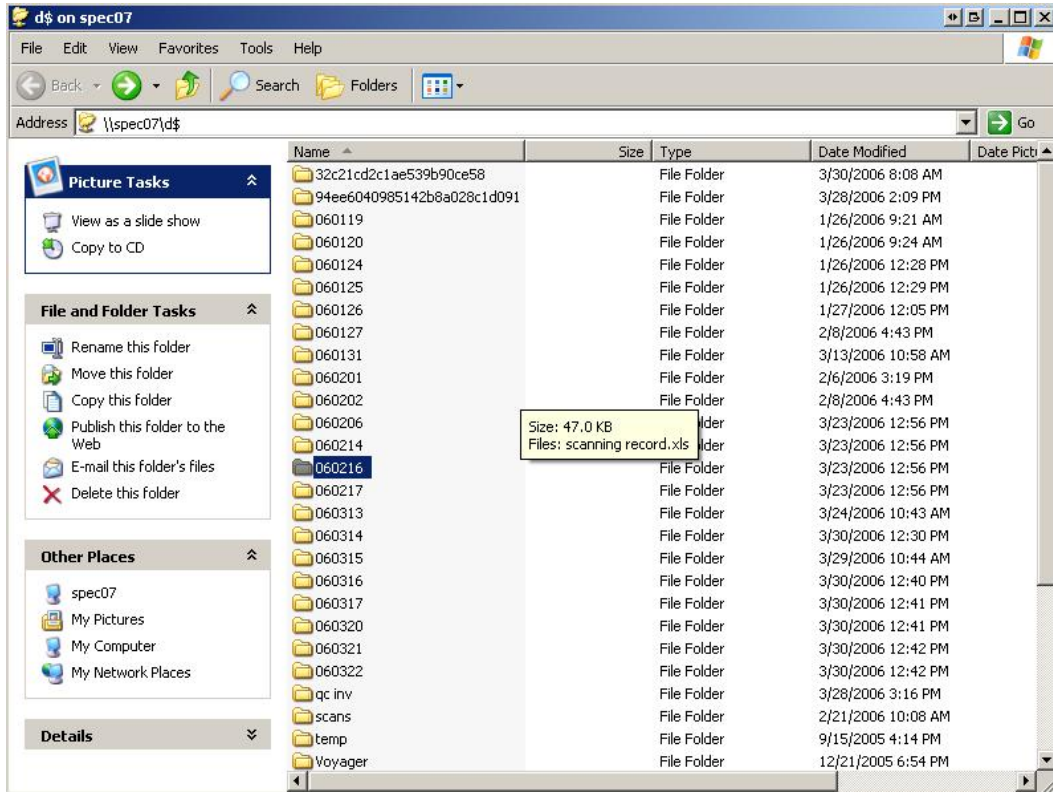
Look for a folder called **STAGING AREA** on local drive **D:** and open it. (If this folder does not exist, go ahead and create it.)

Example of STAGING AREA:



Move the image folder that needs to be checked into your **D:** drive **STAGING AREA**, and into either the **SPEC07** or **SPEC09** folder (depending where you got the images from). Moving might take a few minutes, so be patient.

Example of what Spec07 might look like:



Notice the folder naming convention:

Folders are named according to the date on which the images were scanned.

Example Folder Name: **060216**

February 16, 2006.

06=2006

02=February

16=16th

Do not forget to copy the folder from one of the scanning computers to your local **D:\STAGING AREA**.

You are now ready to use *AsTiffTagViewer* to verify the image specifications.

ASTIFFTAGVIEWER

AsTiffTagViewer will allow you to verify that all images conform to the required technical standards.

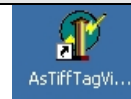
Remember:

- All images must be **100%** of their original size.
- Each image must be scanned at a resolution of **600 dpi**, usually resulting in a minimum of **3000 dots along the longest edge**. (If the original document is small in size, it may not meet this minimum. Verify resolution and size.)
- Text-only documents are scanned at **300 dpi**.
- Images are scanned in **24 bit color** (8,8,8 bits per sample).
- Images must be saved as **uncompressed Tiff** Images.

Check each and every image for the above specifications. **If an image does not conform, it must be rescanned.**

USING ASTIFFTAGVIEWER

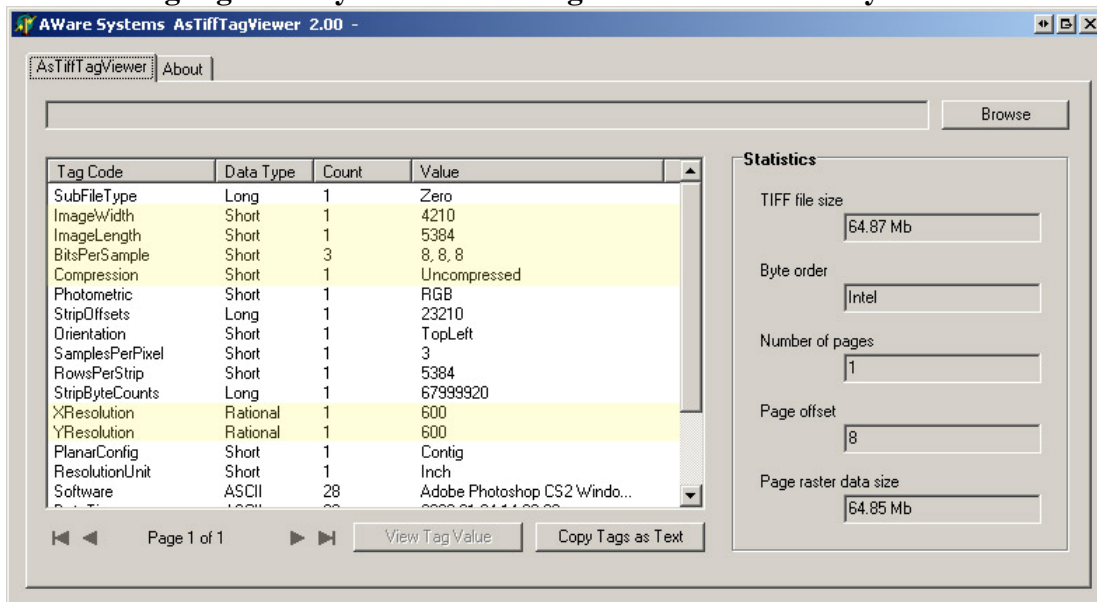
Simply drag and drop the image file onto the *AsTiffTagViewer* shortcut on your desktop.



This will open *AsTiffTagViewer*, displaying the statistics for the selected file (see example below). Verify that the following tags meet the necessary requirements:

- ImageWidth, ImageLength:** One should meet the **3000 minimum** (unless the image is small)
- BitsPerSample:** **8,8,8** (24 bit color)
- Compression:** **Uncompressed**
- XResolution:** **600** (see Exceptions above)
- YResolution:** **600** (see Exceptions above)

The lines highlighted in yellow in the image below are the ones you need to check.



Once you have verified the image standards, you can check the image files with the **Scanning Record**.

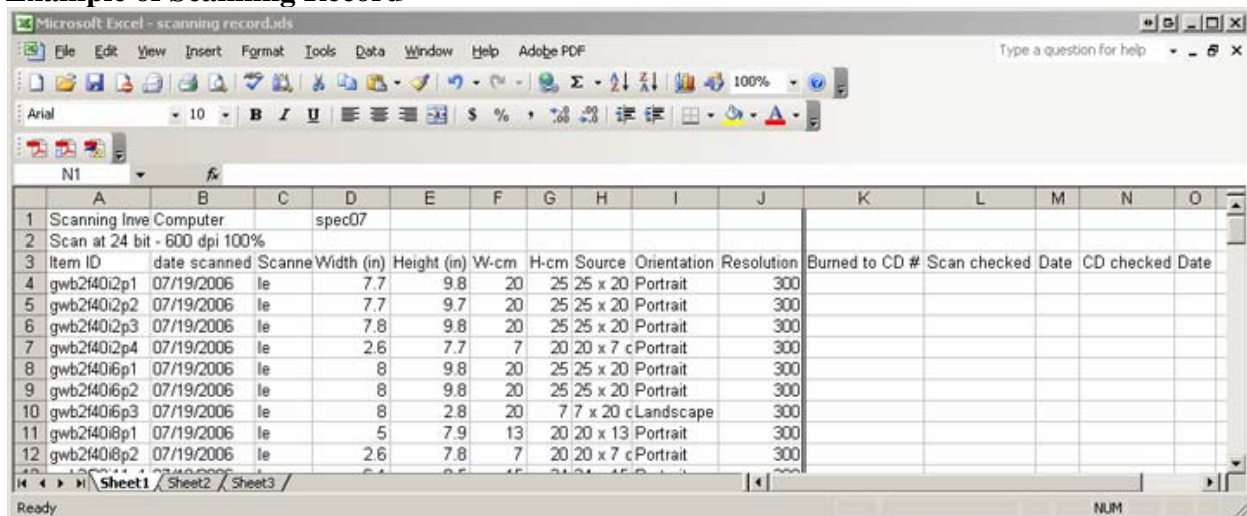
SCANNING RECORD

Each folder of images contains a **Scanning Record** (usually named **scanning record.xls**). This file is an *Excel* spreadsheet listing information for all of the images found in that folder. You need to make sure that this record accurately describes the images in the folder.

The Scanning Record should contain the following information about each of the scanned images:

- **Item ID**
- **Date Scanned (MM/DD/YYYY)**
- **Scanned By**
- **Width (in inches)**
- **Height (in inches)**
- **Width (in cm)**
- **Height (in cm)**
- **Source**
- **Orientation**
- **Resolution**
- **Burned to CD**
- **Scan Checked**
- **Date**
- **CD Checked Date**

Example of Scanning Record



Item ID	date scanned	Scanned By	Width (in)	Height (in)	W-cm	H-cm	Source	Orientation	Resolution	Burned to CD #	Scan checked	Date	CD checked Date
gwb2f40i2p1	07/19/2006	le	7.7	9.8	20	25	25 x 20	Portrait	300				
gwb2f40i2p2	07/19/2006	le	7.7	9.7	20	25	25 x 20	Portrait	300				
gwb2f40i2p3	07/19/2006	le	7.8	9.8	20	25	25 x 20	Portrait	300				
gwb2f40i2p4	07/19/2006	le	2.6	7.7	7	20	20 x 7 c	Portrait	300				
gwb2f40i6p1	07/19/2006	le	8	9.8	20	25	25 x 20	Portrait	300				
gwb2f40i6p2	07/19/2006	le	8	9.8	20	25	25 x 20	Portrait	300				
gwb2f40i6p3	07/19/2006	le	8	2.8	20	7	7 x 20 c	Landscape	300				
gwb2f40i8p1	07/19/2006	le	5	7.9	13	20	20 x 13	Portrait	300				
gwb2f40i8p2	07/19/2006	le	2.6	7.8	7	20	20 x 7 c	Portrait	300				

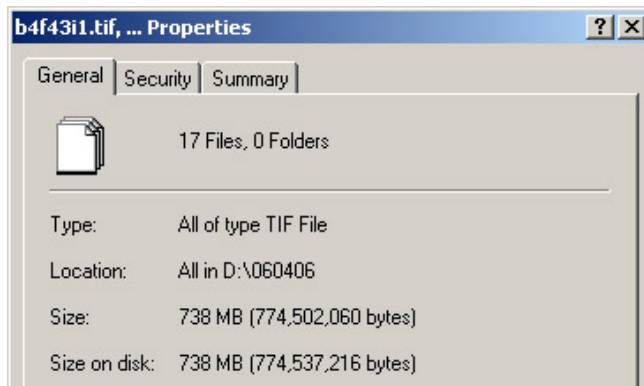
QUANTITY

Also check to make sure the number of files in the folder matches the number of records listed in the spreadsheet.

One way to do this is to select all .tif files in folder (click on the first .tif file, hold SHIFT, and then click on last .tif file).

Now right click on the selected files and select Properties from the menu that appears.

This screen will tell you the number of selected files. You can then compare this to the number of entries in your **scanning record.xls**.



If the numbers don't match, check to see if there are any missing files or entries in the spreadsheet, or if there are any duplicates. Make corrections as necessary.

Remember: the spreadsheet will include a few extra lines of data at the top that are not part of your inventory.

FILE NAME

It is also necessary to make sure there are no errors in the image file name. Files should be named according to their ID numbers as found on the original document. Verify that the file name matches the ID found in the Scanning Record. Discrepancies may require examination of the original document.

COMPARE WITH ORIGINAL DOCUMENT

The scanned images must be checked against the original folder to:

- **Make sure all appropriate originals have been scanned.**
- **Verify that the digital files match the physical files.**

To access the original documents, you must ask a member of the Special Collections staff to retrieve the necessary box for you. *Remember, the box number is the first number in the ID and file name.*

- **You should only be in possession of one box at a time.**
- **Boxes should not be left unattended.**
- **Documents should be handled in their plastic sleeves. If it is necessary to remove the document from the protective sleeve, use the white gloves.**

IMAGE

Verify that the scanned image is indeed the document with that ID number. You can use Windows Thumbnail view or filmstrip to glance at the scanned image in comparison with the actual document, or use *Photoshop* or the *Windows Picture and Fax Viewer* for a larger view. If the image does not match the physical document, correct the file name or and/or rescan the document as necessary.

ID NUMBERS

Look at the ID entries in the Scanning Record and compare them to the numbers found on the actual document. Note that Old and New ID numbers may be found anywhere on the front or back of a document (see examples below).

Old ID found in the upper right corner of the document.



New ID on the right edge of the document.



Errors in the digital file names or Scanning Record are easily corrected by renaming the file or updating the spreadsheet.

Errors found on the original documents must also be corrected.

See the Digital Initiatives Librarian or Archivist before making any changes to original documents. Identification instructions are also found in the Scanning Guidelines.

MISSING ITEMS

If an item appears to have been accidentally skipped and not scanned, you should make sure it has an ID and scan it now and add it to the collection.

If you are unsure if the item should or should not be scanned, check with the Digital Initiatives Librarian or Archivist before identifying and scanning the item.

HARD DRIVE BACKUP

Once you have finished the checks described above and have verified the accuracy and quality of the files, they can be added to the collection folder on **SPEC08 Bartbackup**.

If no folder exists for the collection you are working on, you can create a new folder, naming it after the collection

Example: Amelia Earhart at Purdue

IMAGE FILES

Copy all Quality Checked **.TIF** files to this collection folder on Bartbackup.

INVENTORY.XLS

In addition to the image files, the folder must also include an Inventory file (**inventory.xls**).

This file includes the information from your Scanning Record spreadsheet (see **Scanning Record** above).

If an **inventory.xls** file already exists, simply copy and paste the entries from the Scanning Record as additional lines in the Inventory file.

If no **inventory.xls** exists, copy your **scanning record.xls** file and rename it **inventory.xls**

FINAL CHECK

Now that the files have been stored to the Hard Drive, verify the backups.

VIEW FILES ON HARD DRIVE

Check the images stored on **SPEC08** one more time by viewing the index in Thumbnail format. Also open up 10% sampling of the images with *Photoshop* to make sure the files are read.

QUALITY CONTROL IS COMPLETE

The files are now ready for to be prepared for loading into *Content DM*.

See **File Load Preparation Procedures**.