Quality Control for Digitization Projects

Quality Control (QC) is a process that verifies the quality, accuracy, and consistency of digital images to ensure the products match the specifications outlined in the project plan. The Quality Control process is particularly important for scanning projects where paper records will be destroyed after scanning.

Quality control standards should be developed and widely shared at the beginning of the project and need to remain consistent throughout. Additionally, the specific QC responsibilities of individual staff members must be clearly communicated.

DEVELOPING A QC PROGRAM:

- *Review the project goals*
  The goal of the scanning project will impact the QC criteria. For example, if the goal is to create a replica of the paper record, the digital images should look as close as possible to the original item. However, if the goal of the project is to create the best quality digital images, then ensuring the digital images accurately represent the originals will not be part of the QC process.
- *Identify the products to be evaluated*
  Determine if QC will be performed on the master files and any copies, printouts, metadata, and/or the OCR'ed (Optical Character Recognition) files.
- *Control the QC Environment*
  - A number of factors can impact the QC process including viewing conditions, monitor calibration, and color management.
  - View scanned documents on a monitor and as print outs during the QC process. Documents must be legible both ways.
- *Determine what percentage of files will be reviewed*
  - At the outset, review 100% of files, after the proof of concept many projects review 10% of files.
  - Clearly define baseline characteristics for "acceptable" and "unacceptable" digital images.
  - Determine the number of errors that will render a batch unsuccessful.

QUALITY CONTROL CHECKLISTS

There are two stages of the QC process. The initial is performed by the scanner operator as a regular part of their workflow. The second QC process should be performed by separate staff, after a batch is scanned, at a dedicated QC station.

Scanner Operator QC Responsibilities

- File Specifications
  - Is the file name correct?
  - Is the file format correct?
  - Are the pages in the correct order?
- Image Quality
  - Is the image rotated or backwards?
  - Is the image skewed or off center?
  - Is the content readable?
  - Is the scan "clean"? (no hair, dust, specks of paper, or finger prints visible in the scan)

Secondary Quality Control

- Image Quality
    - Are there moiré patterns? (wavy lines or swirls, usually found in areas where there are repeated patterns)
    - Are digital artifacts present? (such as very regular, straight lines across picture)
    - Is the image too light or too dark?
    - Are details lost in highlight or shadows?
    - Does the image have clean edges, clear contrast, and legible text?
    - Are individual pixels apparent to the naked eye?
- File Properties
    - Is the image the correct size/resolution in long dimension?
    - Is the bit depth correct?
- Is the OCR error free?
- Are all files and pages accounted for?
- Were any items skipped between scanning sessions?
- Is the metadata accurate and complete?

Resources referenced for this document

FADGI - Still Image Working Group. *Technical Guidelines for Digitizing Cultural Heritage Materials*. 2016. http://www.digitizationguidelines.gov/guidelines/digitize-technical.html (accessed June 2017).