

Advisory 19-01, Quality Control and Content Verification of Digital Images

This Advisory must be used in conjunction with the New York State Archives [Digital Imaging Guidelines](#), (2014) (Updated 2019)

Scanning or digitizing is a useful means to increase, and make more efficient, access to records. Although the benefits of scanning can be great, the process of reformatting is labor intensive.

Scanning projects have three main components or steps

1. **Preparation.** This consists of purging duplicates and obsolete documents, putting the records in the order that best facilitates access, removing fasteners such as staples or paper clips, repairing torn papers, flattening folded or rolled documents, and photocopying faded pages to darken them and make them more legible.
2. **Scanning - Indexing.** This is the process of using a document scanner to convert each page into a digital image and linking the images to indexing terms that can be used to retrieve the image later.
3. **Quality Control/Content Verification.** This is the last step in the scanning process and the subject of this advisory. It is the process verifying that all pages were scanned, each is legible and usable, and is linked to the correct indexing terms.

The terms Quality Control (or QC) and Content Verification consist of two components: Quality Control is the process of reviewing scanned images with the human eye to ensure that the images are of acceptable quality and meet all technical specifications. Content Verification is the process of checking the scanned images against the original record (paper or microfilm) to ensure each page of each document was scanned. These two components are typically performed simultaneously.

You Must Verify Your Own Images

Imaging vendors will perform quality control on at least a portion of the scanned images to try to ensure they are producing a quality product and meeting the standards set forth in the client contract. However, you cannot depend on the assurances of a vendor that all your records were scanned and are legible and usable and contain the correct indexing terms. (If scanning inhouse, ensure that someone other than the persons scanning and indexing are conducting verification) It is the responsibility of every state agency or local government to ensure that all their records are accessible until they meet their legal retention requirements.

Determining the Method of Verification

There are two methods for verifying your images: 100% verification and sampling.

The method of verification you choose will mainly be determined by what you decide to do with the original records once they are scanned. If you are destroying the original records, you must conduct 100% verification of all images against the original document. Verification *must be performed in batches* during the scanning process and *not at the completion of the scanning of all records* Therefore, state

agencies and local governments must stipulate how frequently and by what transfer method a vendor will return batches for verification in their vendor contracts.

100% Verification

100% verification, as its name suggests, means visually inspecting each image while comparing it to the original document to ensure that each paper document was scanned and is legible with no loss of information.

Sampling

Sampling, as its name suggests, means taking a sample of a batch of images to visually inspect. If that sample is acceptable, the whole batch is deemed acceptable.

If you choose to maintain the original records as a backup for the images, then you may use one of two sampling methods:

Local governments and state agencies may use the ANSI/AIIM standard AIMM TR34: *Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management and Micrographic Systems* (36 CFR 1237.28 (d)(2)) developed by the Association for Information and Image Management. This method uses complex statistical formulas to determine sample sizes and the acceptable number of errors per sample size.

Local governments and state agencies may also use the following simpler method developed by the New York State Archives:

1. Review the first one hundred images in the batch.

If there is an error rate of 3% or less,

2. then review every 25th image in the batch to determine if the whole batch appears to have a consistent quality. If you find an error while checking every 25th image be sure to check a few images immediately before and after the error to determine if the problem is more widespread or just a lone error.
 - a. If it appears the errors are more widespread then return to checking clusters of 100 images at the point you found the error. If the error rate exceeds the accepted threshold for that cluster, then review another cluster of 100. If it that one exceeds the accepted error threshold, then reject the whole batch. If it doesn't, return to checking every 25th image.
 - b. If you detect few, if any, errors on every 25th image after that, then you may accept the whole batch. If you find more errors while checking every 25th image then reject the whole batch.

If the error rate is over 3%,

3. then verify the next cluster of 100 images. If the error rate is still over 3%, then review one more cluster of 100 images, and if it exceeds the error rate, then reject the whole batch. If the third cluster does not exceed the error rate, return to checking every 25th image and follow the instructions from 2b above.
4. If the error rate falls below 3% for the next cluster of 100 image, then return to reviewing every 25th image and follow the instructions from 2b above.

How to Verify Images

Whether doing 100% verification or sampling verification, it should always be done in batches. Do not wait until all records are scanned, especially if doing 100% verification. First, many standard imaging vendor contracts provide a limited window of opportunity to rescan images for free so doing verification in batches will ensure you can comply with such provisions. Second, if you wait until all images are scanned to conduct verification, it will make the process much more challenging.

Let's look at an example. The County of Foster scanned 430,000 pages of tax rolls and waited until scanning was complete to begin verifying images. Using the standard verification rate of 300 pages per hour it will take 1,433 hours or 40 weeks for one person to verify the images. And this assumes it is feasible for someone to verify images 7 hours a day (assuming a one-hour lunch). We must consider factors such as eye fatigue, sustained concentration level, and sheer boredom. Even if several people are doing verification it is not feasible to expect them to do it seven hours a day and do an accurate job.

If doing 100% verification, the Archives recommends:

- Do it in batches throughout the scanning phase of the project.
- Have more than one person work on verification
- Allow the verifiers to take frequent breaks and/or assign them other duties between verification sessions

Error Detection

When conducting image verification/quality control the verifier must know what types of errors to look for. Among the types of errors that must be corrected are

- the page was not scanned
- portion of the page not captured
- speckle on image background
- incorrect orientation

For a more complete list of errors, see section 23.6 of the [*Digital Imaging Guidelines*](#)