OFFICE *of the*
CHIEF RECORDS
OFFICER

DIGITIZATION QUALITY MANAGEMENT GUIDE

National Archives and Records Administration
October 2023

FOREWORD

Digitization has revolutionized the way we preserve and access information. The benefits of digitization are undeniable - from increased accessibility and improved preservation to making it easy to share images of documents, photographs, and artifacts. However, digitization also introduces unique challenges, particularly in the area of quality management (QM). There are many opportunities for information to be lost or distorted during digitization. Records might be overlooked, images might not capture sufficient detail, or lack of sufficient metadata could render the resulting digital images difficult to work with.

The need for effective QM practices in digitization cannot be overstated. As we move towards a more digital world, it is critical that we prioritize prevention over correction when it comes to quality issues.

This guide was written to help agencies learn about the various aspects of QM in digitization, including quality assurance and quality control, as well as the role of objective testing and automation in optimizing quality control and inspection processes.

This guide supports the National Archives and Records Administration's (NARA) regulations concerning digitization standards for permanent records (36 CFR 1236 Subpart E) and digitization standards for temporary records (36 CFR 1236 Subpart D). NARA's regulations for the digitization of permanent records provide an example of effective QM practices in digitization. By integrating the requirements for records management practices, digital imaging quality standards, quality control, inspection, and validation steps, NARA supports the goal that the resulting digital surrogates fulfill the same legal and evidentiary purposes as the source records.

As agencies work to digitize records, adopting effective quality management practices is critical. This guide emphasizes the importance of prioritizing QM practices in digitization projects to ensure that digitized records remain authentic, accurate, and usable for future generations.

Digitization Quality Management

> *"Inspection with the aim of finding the bad ones and throwing them out is too late, ineffective, and costly. Quality comes not from inspection but from improvement of the process."* W. Edwards Deming

Introduction

The goal of quality management (QM) in a digitization project is to prevent defects before they happen. Prevention of defects is the most efficient, cost-effective, and productive activity in a digitization workflow. Digitization is an image manufacturing process that relies upon quality assurance (QA) to establish specifications and requirements, and a quality control (QC) process to test for and inspect defects. A well-run digitization operation will be proactive rather than rely upon QC inspection to achieve quality.

Many agencies use Federal Agencies Digital Guidelines Initiative (FADGI) or Metamorfoze guidelines to inform their QM practices. Both FADGI and Metamorfoze define image quality parameters and tolerances as well as describe testing and analysis processes to evaluate the digital capture device performance. The underlying assumption is that through evaluating the capture device's performance against known metrics, the files created by that device conform to a standard.

NARA's digitization regulations could be considered a total quality management document that brings together records management practices, digital imaging quality standards, quality control, inspection, and validation steps. Each phase of a digitization project affects the other parts. For example, establishing intellectual and physical control contributes to creating metadata as well as identifying missing records. The quality control and inspection steps have been optimized to rely on automated processes where possible and limit human inspection to the phases that cannot be automated. By relying on objective testing and analysis for image quality, we eliminate wasteful subjective inspection of attributes. Finally, the validation phase is a high-level review to verify that all the requirements of a project have been met and the digital surrogates can serve the same legal and evidentiary purpose as the source records.

What Is Quality Management

Quality management (QM) is the overall management function and underlying activities that determine quality policies, objectives, and responsibilities, and implement them through planning, control, assurance, and improvement methods within the quality system.

Quality assurance (QA) is a proactive quality management (QM) activity focused on preventing defects by ensuring that a particular product or service achieves certain requirements or specifications. A QA program is heavily dependent on quality control (QC) data to search for patterns and trends. QA activities also include controlled experiments, design reviews, and system tests. QA programs can improve quality by creating plans and policies or creating and conducting training.

Quality control (QC) encompasses activities that examine products through inspection or testing to determine if they meet their specifications. The purpose is to detect defects (deviations from predetermined requirements) in products or processes.

Quality management represents the addressable corrections that can be made during a digitization project. Validation is a high level review to ensure the project and deliverables meet the legal and evidentiary requirements. Therefore, NARA's digitization standards include different requirements for quality management (§ 1236.46) versus validation (§ 1236.56).

Quality Management: Plans and Documentation

The purpose of a quality management plan is to define the requirements to be met so that the digital products conform to the specifications of the project. A quality management plan should define the work activities relating to the digitization of source records, and outline a generic sequence of high-level planning and management activities. In this regard, "digitization" is defined as the complete process that includes: selection, assessment, prioritization, project management and tracking, preparation of source records for digitization, metadata collection and creation, digitizing, quality management, data collection and management, submission of digital resources to delivery systems and into a repository environment, and assessment and evaluation of the digitization effort. The processes involved in a digitization workflow generally fall into four main phases:

- Project planning
- Processes occurring before digitization
- Digitization
- Post-digitization work

The plan should define the policies and procedures relating to digitization activities and the management of digitization projects, as well as define how the policies will be implemented to ensure compliance. There is not a single approach to digitization and metadata

creation/management for all projects. The differences in the media type, quality, and condition of the source record, nature of information, or preservation risk level will determine the approaches for digitization and metadata requirements for each project. The essential characteristics (significant properties) of the source record should be identified to determine the best approach to capturing the information of the record.

NARA's regulations with digitization standards contain requirements for documentation. Agencies must create documentation when digitizing source records. They must retain this documentation in association with the digitized records. For permanent records, agencies must include a quality management plan as part of this documentation. A quality management (QM) plan must be created that ensures the project meets the quality assurance (QA) objectives and quality control (QC) inspection procedures.

The quality management plan must include the policies, functions, roles, responsibilities, requirements, and objectives of the project. The quality assurance component of the QM plan must include documentation of: the image quality performance parameters selected to capture the information present in the source records; the equipment and device acceptance testing methods and results; design reviews; and training conducted. The quality control component of the QM plan must document the procedures used to inspect image quality; the procedures used to inspect metadata quality; the corrective actions taken to mitigate deviations throughout all phases of the project; and the procedures used to verify that digitized records conform to the requirements.

Quality Assurance: Specifications and Standards

The quality assurance component of a quality management program includes the mandatory requirements, specifications, and standards that a digitization project must meet. The major components of a digitization product are image quality, metadata quality, records management quality, and file format compliance. QA focuses on the processes used to create project deliverables to ensure that products conform to the standards and specifications. The aim of a QA program is consistency and accuracy. The goal of a QA program is to prevent deviations by performing systematic checks and analyses throughout the production cycle and implementing corrective actions. QA relies upon quality control testing and analysis to identify conformity or deviations from specifications.

Agencies must meet the image quality performance parameters specified in NARA's regulations for the digitization of permanent records by verifying that their equipment achieves the aimpoints and tolerances in § 1236.50. They should not rely on advertised equipment specifications, such as scanner ppi settings or camera sensor megapixels, to ensure digital image quality. Agencies must use QA processes to:

- Quantify scanner or camera performance before selecting the equipment by scanning a reference target and measuring the results with analytical software to determine if the equipment meets the technical parameters;
- Evaluate internal or external vendor imaging systems against image quality performance parameters;
- Monitor equipment performance by quantifying scanner or camera performance during digitization; and
- Verify that resulting digital files meet project specifications.

Image quality is the perceived or objective measurement of a digital image's overall accuracy in faithfully reproducing a source record. A high-quality digital image has been created to a high degree of accuracy, meets or exceeds objective performance attributes, and has minimal defects. The International Organization for Standardization (ISO) ISO 19264-1:2021 standard codifies a range of image quality attributes and defines various quality levels to be achieved through testing and analysis. FADGI describes methodologies to measure the performance of digitization equipment against the ISO tolerances and aimpoints through a Digital Image Conformance Program (DICE).

FADGI has created a set of measurement parameters for specific media types such as modern paper, photographic prints and film, bound volumes, and microforms. The DICE program provides the measurement and monitoring component of a FADGI-compliant digitization program and consists of two components: image targets, both reflective and transmissive, and analysis software. The evaluation parameters establish quality and performance goals for the desired quality level. Digital image conformance testing tools, when used with appropriate testing targets, provide the user with precise and repeatable analysis of the imaging variables that comprise FADGI star ratings.

A hierarchical classification of image quality attributes has been developed to form a taxonomy of digital imaging performance. The taxonomy provides context and a framework for the array of commonly used terms and the appropriate imaging standards for the evaluation of digital image files. The tables of image performance parameters in NARA's digitization regulations have been adopted from FADGI 3-star performance standards. The tables consist of attribute categories such as Color Accuracy, Lightness Uniformity, Sampling Frequency, Spatial Frequency Response (SFR), Tone Response, and White Balance, among other attributes. Each parameter has a set of defined metrics and tolerance levels. The ISO-compliant test targets have regions such as slant edge targets, tone scale, color patches, and other measurement references that relate to the attributes to be measured. By digitizing conformant test targets and analyzing the results,

we can measure the performance of the image capture system and signal processing of the device.

Some of the digitization attributes that must be measured during the QA process include:

Color accuracy is measured by computing the color difference (ΔE2000) between the digital imaging results of the standard target patches and their pre-measured color values. By imaging an appropriate target and evaluating through the software, variances from known values can be determined, which is a good indicator of how accurately the system is recording color. Analytical software measures the average deviation of all color patches measured (the mean).

Color channel misregistration measures the spread of red, green, and blue light in terms of pixel misregistration. This parameter is used to evaluate lens performance. The vernacular term for this is called color fringing.

Lightness uniformity measures how evenly a lens records the lighting of neutral reference targets from center to edge and between points within the image.

Modulation transfer function (MTF)/spatial frequency response (SFR) is the modulation ratio between the output image and the ideal image. SFR measures the imaging system's ability to maintain contrast between increasingly smaller image details. Using these two functions, a system can make an accurate determination of resolution related to the sampling frequency.

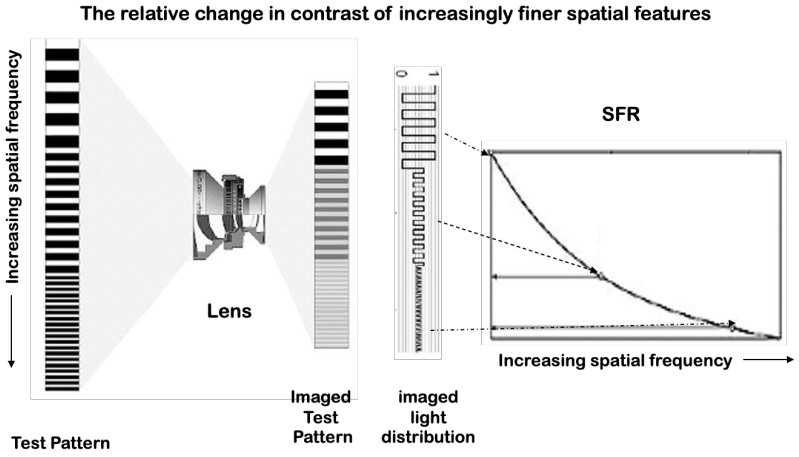

Figure 1. Graphical Illustration of SFR using a resolution test target showing how the contrast between image details worsens as spatial frequency increases

Noise is one or more undesirable image artifact(s) in a digitized record that is not part of the source material.

Spatial resolution determines the amount (for example, quantity, ppi, megapixels) of data in a raster image file in terms of the number of picture elements or pixels per unit of measurement, but it does not define or guarantee the quality of the information. Spatial resolution defines how finely or widely spaced the individual pixels are from each other. The actual rendition of fine detail is more dependent on the SFR of the scanner or digital camera.

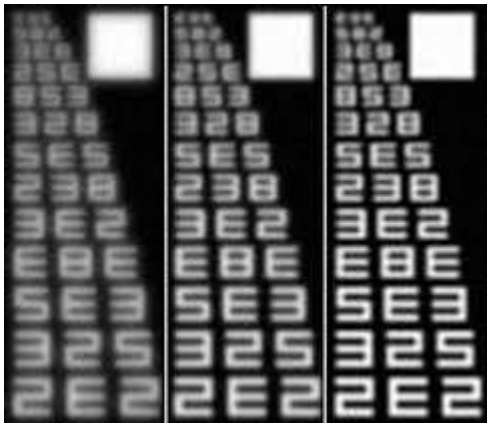

Figure 2. Test target (R.I.T Alphanumeric Target) demonstrating different spatial resolutions

Tone response or optoelectronic conversion function (OECF) is a measure of how accurately the digital imaging system converts light levels into digital pixels.

White balance error is a measurement of the digital file's color neutrality. The definition of neutral is not universal. RGB workflows that use digital count values encode neutral as defined by the International Color Consortium (ICC) color space chosen. L*a*b* workflows define neutral as 0 on the a* axis and b* axis, with the lightness recorded from 0-100 on the L* axis.

Quality Control: Testing and Analysis

The purpose of Quality Control (QC) is to find and eliminate sources of quality non-conformities using tools and equipment to ensure the customer's expectations for the quality of the project deliverables are systematically and continually met. QC activities or techniques are used to achieve and maintain product quality, process, and service. QC is the responsibility of the team that tests the product for defects. Testing early and often is a basic quality control strategy.

During the digitization process, agencies must perform QC testing and analysis to identify malfunctioning or improperly configured digitization equipment, improper software application settings, incorrect metadata capture, or human error, and take corrective actions. An image quality analysis process must be established and use reference targets to verify that digitization devices conform to imaging parameters.

Basic operations include:

- Scanning reference targets that contain a grayscale, color chart, and accurate dimensional scale at the beginning of each workday;
- Replacing reference targets as they fade or accumulate dirt, scratches, and other surface marks that reduce their usability; and
- Frequently testing equipment to ensure scanners and digital cameras/copy systems are performing optimally.

Image quality analysis software must be used to verify that the performance evaluation specifications are being met. Equipment must be tested with the specific software/device driver combination(s) used and re-tested after any changes to the workflow. Ensure that equipment operation, settings, and image processing actions are the same as those used to evaluate the test target. Disable auto settings in the capture equipment such as “auto exposure” that may affect the target evaluation or the resulting image files.

The “OpenDICE” program developed by FADGI is an open-source, free image analysis software program. The software is used to analyze scans of reference targets selecting the material to be digitized and the type of reference target to be analyzed.

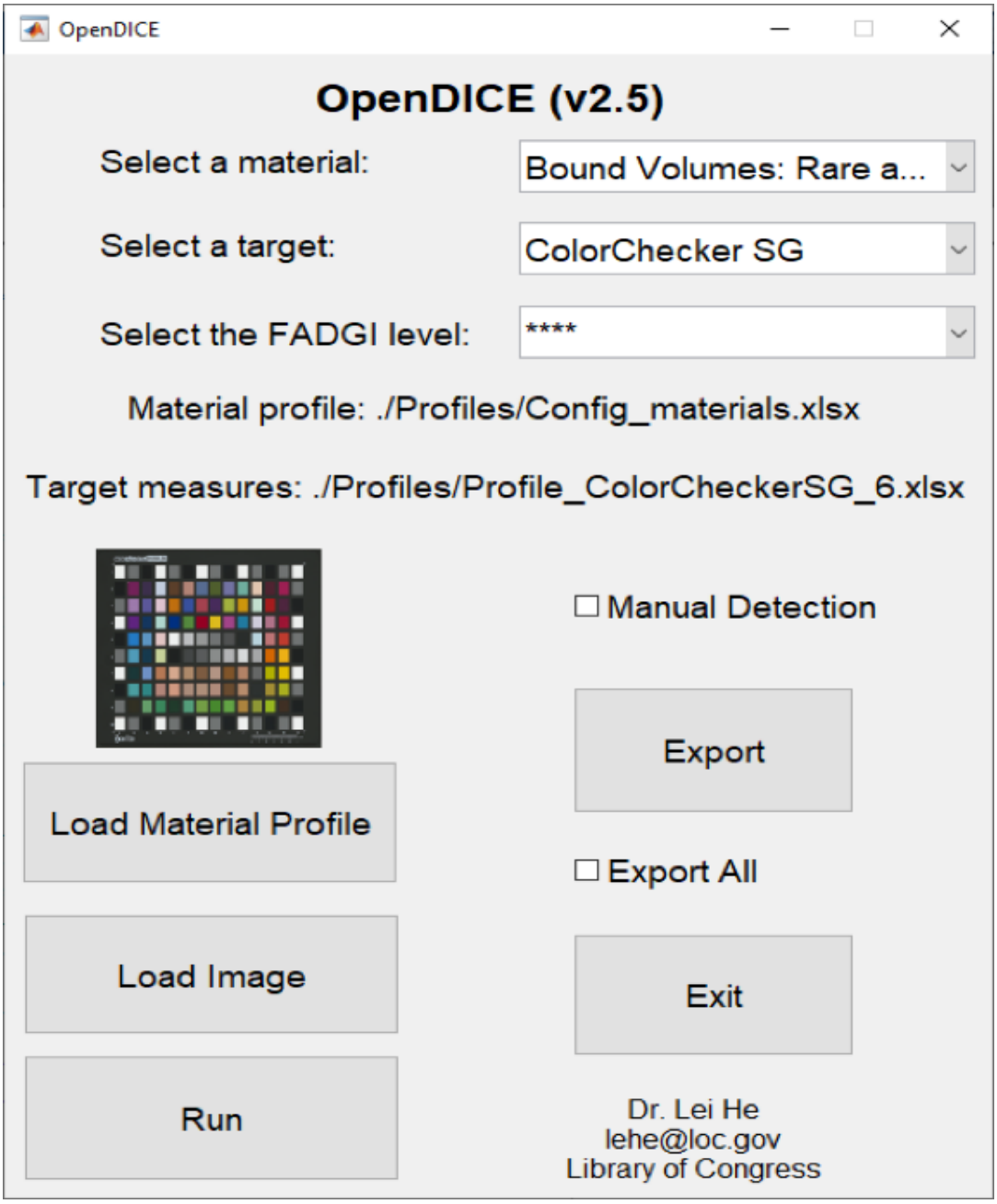

Figure 3. Library of Congress Open DICE software input screenshot

The capture device performance is evaluated by the software by reading the values of the test target and comparing the results against the parameters of the FADGI material type.

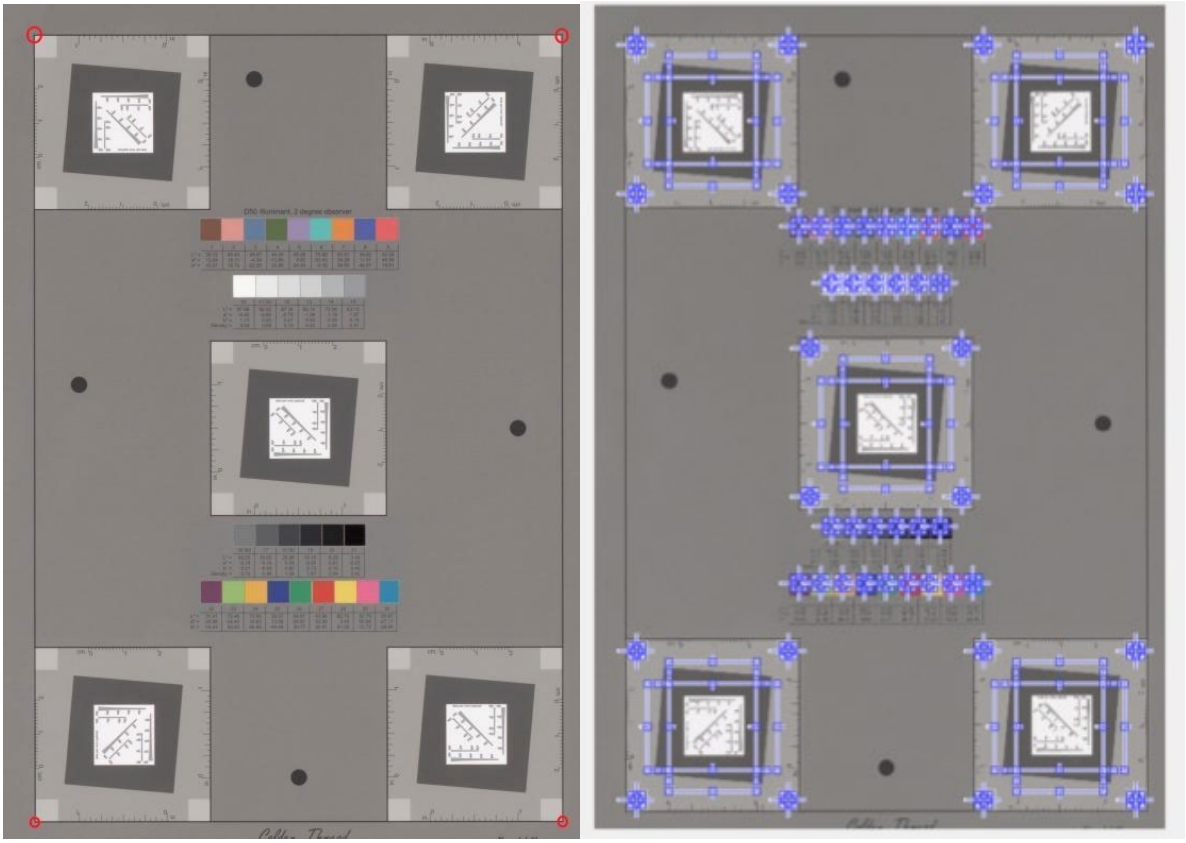

Figure 4. FADGI Compliant reference target (right). Highlighted regions of interest

The quality of the image is determined by the level of compliance with the parameters specified by the FADGI performance level. The software creates reports for the different measurement parameters. The reports are used to evaluate the image device performance and to identify regions where the device is out of compliance with the parameters.

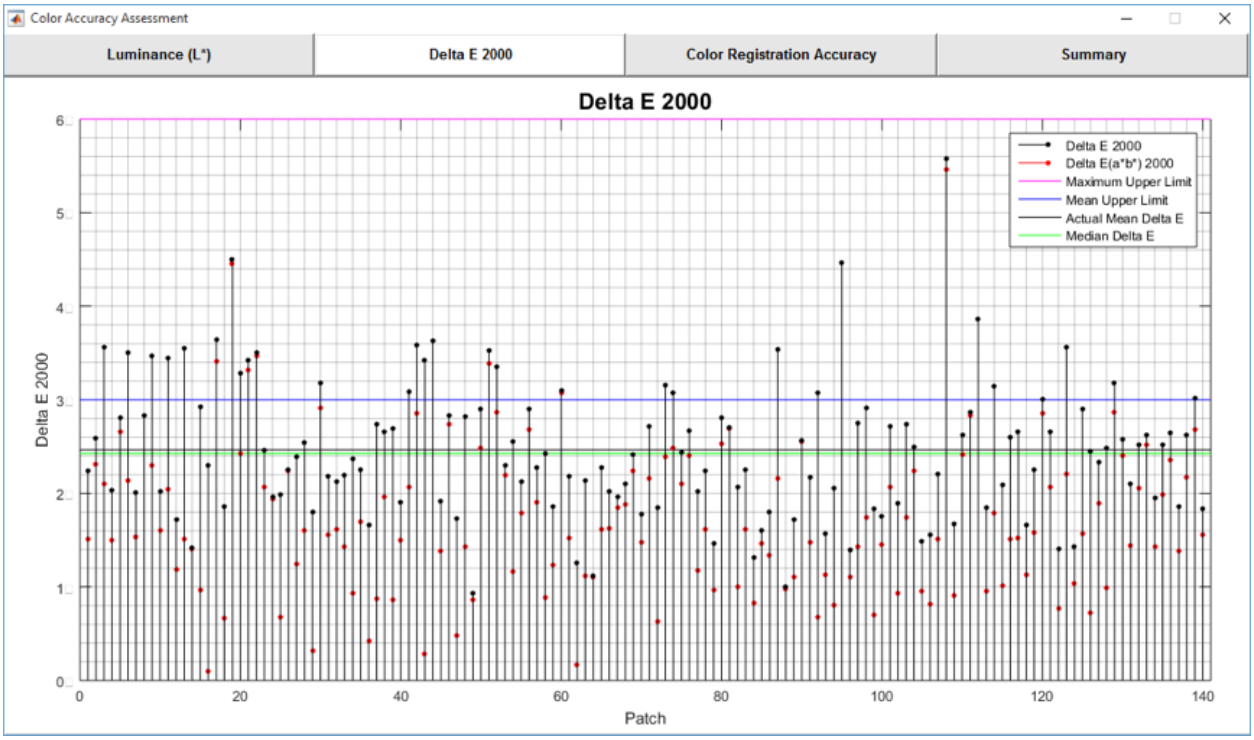

Figure 5. Example of color accuracy data analysis results.

Image testing and analysis are critical components of digitization quality control. Tests should be performed on equipment before purchase to determine if a particular device can meet the performance goals of the project, and capture the information contained in the source material. The testing and analysis should also be used during the digitization production process to monitor for any deviations in the workflow.

Quality Control: Inspection

The image files and metadata created by the digitization project must be inspected to detect any errors. Any deviations from the specifications must be corrected before the files may be accepted. The digital versions must be compared to source records and verified that they meet the standards in the regulation. The inspection process uses objective measurements, as well as subjective methods to ensure quality.

Inspection is the process of measuring, examining, testing, or otherwise comparing the product with the requirements. The inspection process incorporates sampling techniques to identify the percentage of errors that occur. It is not desirable to conduct a visual review of every file that has been digitized. In most large projects visual inspection of all files is impractical. Therefore, a statistically valid sample of digital files from each batch must be inspected.

Automated quality control checks may be performed using a variety of tools and scripts. For example, a script may be able to check and verify files for correct file names, proper file sequence, comparison of checksums, correct resolution, bit-depth, compression, color mode, correct file format, or embedded color profile.

Automated checks are usually conducted as the “first pass” of quality control and all files must pass these automated checks. The frequency of automated checks may vary across projects. Checks may be run with every batch, at specific time intervals, or as files move into a specific directory or sequence in the workflow.

Human-conducted visual quality control checks often form the second pass of the inspection process. The effectiveness of the evaluation of subjective factors, such as image quality, metadata accuracy, and legibility rely upon the skill and experience of the person performing the inspection. This forms the more challenging part of the quality control process, as evaluation is not based on quantifiable criteria. Therefore, care must be taken when analyzing the quality of subjective factors.

One major factor is to evaluate the “completeness” of the digitized records. Completeness includes ensuring all the physical records have been accounted for in the digital version, that there are no missing pages, and most importantly that the information in the source record has been captured accurately.

Errors or defects identified during inspection will require re-inspection and re-digitization. A proactive quality management program is less costly and a more efficient practice.

Inspection Checklist

QC inspections of image files for compliance with the technical parameters and specified criteria must be performed to ensure that 100% of the image files:

- can be opened and displayed;
- are encoded with a compression type and in a format specified in § 1236.48; and
- have the resolution, color mode, bit depth, and color profile specified in the regulation.

Agencies must perform a visual inspection by inspecting a random sample of a minimum of ten image files or 10% of each batch of image files, whichever is larger, on a calibrated graphics workstation or alternatively by employing a statistically valid technique. Visual inspection must be conducted on a monitor at 100% magnification for the following image quality characteristics:

- the image tone, brightness, contrast, and color accuracy match the specifications in the regulation;
- images are free from clipping (missing detail lost in highlights or shadows);
- images are free from color channel misregistration, or quantization errors;
- images are free of any image artifacts that may compromise the informational content of the record, such as dust, Newton's rings, missing pixels, scan lines, drop-outs, flare, or over-sharpening; and
- images have the expected dimensions and orientation (landscape/horizontal or portrait/vertical), and images are not flipped, inverted, improperly cropped, or skewed.

Corrective Measures

If errors are detected during inspection, perform the following steps until the specifications and requirements have been met:

- If 1% or more of examined records fail to meet any of the criteria, determine the source and scope of any errors, correct or re-digitize affected records, and reinspect the images by following the requirements until a 100% success rate is achieved for the sample set;
- If less than 1% of examined records fail to meet any of the criteria, determine the source and scope of any errors and correct or re-digitize the affected records.

Inspection for other quality aspects

Agencies must inspect the resulting files to verify that they meet the metadata and records completeness requirements.

Metadata quality. The accuracy of metadata must be evaluated. This may be done using automated techniques if appropriate. Otherwise, manual QC inspections must be conducted to evaluate the accuracy of metadata content. Agencies must ensure that:

- files are named according to project specifications; and
- correct administrative, descriptive, and technical metadata are captured in a recordkeeping system and image files.

Records completeness. Agencies must employ automated and visual inspection processes to verify the completeness and accuracy of digitization to:

- verify that all records have been accounted for by referring to box lists, folder title lists, or other inventories;
- compare source records with their digitized versions to verify that 100% of the informational content has been captured;
- compare source records with their digitized versions to verify the digitized records are in the same order as the source records;
- examine records for related envelopes, notes, or other forms of media to verify that all sources of record information have been digitized;
- verify that any mixed-media records that cannot be digitized are associated with the digitized records using the "Relation" metadata elements in § 1236.54(c); and
- confirm that missing pages or images have been noted in the project documentation.

Validation

Validation is the human-conducted process where agencies formally declare that their digital records are complete, accurate, comply with the standards in the regulations, and can be used for the same business purposes as the source records.

Validation is different from quality management. Quality management occurs throughout the digitization process and involves the inspection of the various attributes of each image such as their file format, resolution, and metadata. Validation is a higher level review of an agency's digital records and the process that were used to create them. Validation occurs when digitization is complete and must be conducted by staff that were not involved in the quality control inspections conducted during the digitization process.

Conclusion

Digitizing paper and analog records is an essential step for federal agencies transitioning to digital government. However, simply converting paper-based records to digital formats is not enough. Quality management plays a crucial role in ensuring the accuracy, completeness, and reliability of digital records.

This guide has explained how effective quality management practices can help agencies minimize errors, reduce costs, improve productivity, and enhance customer satisfaction with reliable, authentic, and usable records. By implementing a robust quality management system,

organizations can establish standardized processes for digitizing records, monitoring performance, and continuously improving their operations.