Federal Agencies
Digital Guidelines Initiative

September 2016

Technical Guidelines for Digitizing Cultural Heritage Materials

Creation of Raster Image Files

Document Information

Title	***Editor***
Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Files	Thomas Rieger
Document Type	Technical Guidelines
Publication Date	September 2016

Source Documents

Title	***Editors***
Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files *http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf*	Don Williams and Michael Stelmach
Document Type	Technical Guidelines
Publication Date	August 2010
Title	**Authors**
Technical Guidelines for Digitizing Archival Records for Electronic Access: Creation of Production Master Files – Raster Images *http://www.archives.gov/preservation/technical/guidelines.pdf*	Steven Puglia, Jeffrey Reed, and Erin Rhodes U.S. National Archives and Records Administration
Document Type	Technical Guidelines
Publication Date	June 2004

INTRODUCTION

The *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Files* represents shared best practices followed by agencies participating in the Federal Agencies Digital Guidelines Initiative (FADGI) Still Image Working Group for digitization of cultural heritage materials. This group is involved in a cooperative effort to develop common digitization guidelines for still image materials (such as textual content, maps, and photographic prints and negatives) found in cultural heritage institutions.

This revision of the 2010 FADGI Guidelines incorporates new material reflecting the advances in imaging science and cultural heritage imaging best practice, and adds a section on newspaper digitization. The format has changed. Relevant information specific to defined imaging tasks has been grouped into individual chapters, along with guidance on how to apply FADGI to specific tasks. These *Guidelines* are intended to be used in conjunction with the DICE (Digital Imaging Conformance Environment) targets and software developed by the Federal Agencies Digital Guidelines Initiative and the Library of Congress. Together, these *Guidelines* and the DICE testing and monitoring system provide the foundation for a FADGI compliant digitization program.

This revision also recognizes and is compatible with the Metamorfoze guidelines, released in January 2012, and is consistent with ISO standards currently under development.

SCOPE

The focus of the *Guidelines* is on historical, cultural and archival materials. The scope is limited to digitization practices for materials that can be reproduced as still images, e.g., printed matter, manuscripts, maps, and photographic prints, negatives and transparencies.

The *Guidelines* are intended to be informative, not prescriptive. We acknowledge that this document does not address the entire range of image quality parameters, but these topics will be incorporated as the Still Image Working Group identifies recommendations in these areas. The Working Group has produced a "Gap Analysis" document that identifies and prioritizes digitization activities that are not currently defined within existing agency guidelines, or are not adequately addressed by existing guidelines. The Gap Analysis contains topics that the Working Group intends to investigate and provide as updates and recommendations in future versions of these *Guidelines*.

The current Gap Analysis can be found on the FADGI website at:

http://www.digitizationguidelines.gov/guidelines/Gap_Analysis.pdf

We hope to provide a technical foundation for digitization activities, but further research will be necessary to make informed decisions regarding all aspects of administrative, operational, and technical issues surrounding the creation of digital images. These *Guidelines* provide a range of options for various technical aspects of digitization, primarily relating to image capture, but do not recommend a single approach.

The following topics are addressed in this document:

- Digital image capture for still images – creation of raster image files, image parameters, digitization environment, color management, etc.
- Color encoding accuracy – color space, color temperature for imaging and viewing, quality of linear vs. area arrays, and quality of different interpolation algorithms.
- Digital image performance – development of operational metrics and criteria for evaluating digital image characteristics for purposes of investigation or for quality control purposes, including metrics and criteria for resolution, noise, color encoding, mis-registration, etc., and for measuring system performance capabilities.
- Example workflow processes – includes guidelines for image processing, sharpening, etc.

- Minimum metadata – we have included a discussion of metadata to ensure a minimum complement is collected/created so master image files are renderable, findable, and useable.
- File formats – recommended formats, encodings of master and derivative files.
- Approaches to file naming.
- Basic storage recommendations.
- Quality management – quality assurance and quality control of images and metadata, image inspection, acceptance and rejection, and metrology (ensuring devices used to measure quality or performance are giving accurate and precise readings) among others.
- Optical character recognition.

The following aspects of digitization projects are not discussed in these *Guidelines*:

- Project scope – defining goals and requirements, evaluation of user needs, identification and evaluation of options, cost-benefit analysis, etc.
- Selection – criteria, process, approval, etc.
- Preparation – archival/curatorial assessment and prep, records description, preservation/conservation assessment and prep, etc.
- Descriptive systems – data standards, metadata schema, encoding schema, controlled vocabularies, etc.
- Project management – plan of work, budget, staffing, training, records handling guidelines, work done in-house vs. contractors, work space, oversight and coordination of all aspects, etc.
- Access to digital resources – web delivery system, migrating images and metadata to web, etc.
- Legal issues – access restrictions, copyright, rights management, etc.
- IT infrastructure – determine system performance requirements, hardware, software, database design, networking, data/disaster recovery, etc.
- Project assessment – project evaluation, monitoring and evaluation of use of digital assets created, etc.
- Digital preservation – long-term management and maintenance of images and metadata, etc.
- Digitization of audio/visual and moving image materials.
- Management of “born-digital” materials.

A companion FADGI document, *Digitization Activities – Project Planning and Management Outline*, provides a conceptual outline of general steps for the planning and management of digitization projects, and addresses some of the topics listed above. This document is available at

http://www.digitizationguidelines.gov/guidelines/DigActivities-FADGI-v1-20091104.pdf

The intended audience for these *Guidelines* includes those who will be planning, managing, and approving digitization projects, such as archivists, librarians, curators, managers, and others, as well as practitioners directly involved in scanning and digital capture, such as technicians and photographers. The topics in these *Guidelines* are inherently technical in nature. For those working on digital image capture and quality control for images, a basic foundation in photography and imaging is essential. Generally, without staff with a good technical foundation, achieving the appropriate level of quality as defined in these *Guidelines* is problematic. Cultural heritage digitization is a specialization within the imaging field that requires specific skills and experience. The FADGI has compiled these specific recommendations and best practices as they are practiced at participating institutions. Implementation of these recommendations should be accomplished by personnel with appropriate experience or in consultation with institutions or experts experienced in implementation of FADGI compliant digitization programs.

Revisions:

These *Guidelines* reflect current best practices shared by members of FADGI. We anticipate they will change as technology and industry standards, as well as institutional approaches, improve over time. As the technical arenas of conversion, imaging, and metadata are highly specialized and constantly evolving, we envision these *Guidelines* to be a continually evolving document as well. The *Guidelines* will be collectively reviewed by participating agencies at regular intervals and updated as necessary.

We welcome your comments and suggestions.

Please note that the **online version** of the *Guidelines* is considered to be the official document.

The FADGI Star System

FADGI defines four quality levels of imaging, from 1 star to 4 star. Higher star ratings relate to more consistent image quality, but require greater technical performance of both operator and imaging system to achieve. The appropriate star performance level for a particular project should be carefully considered in the planning stage of the project.

Conceptually the FADGI four star system aligns with the Metamorfoze[1] three tier system, with a fourth tier (1 star) on the lower end of the performance scale. Both FADGI and Metamorfoze trace their metrics to emerging ISO standards efforts. While similar, there are differences.

This revision of the *Guidelines* more fully expands the use of colorimetric measures like L*a*b* color and ΔE(a*b*) 2000 measurements. These changes align FADGI with ISO TC42/WG18 protocols and Metamorfoze guidelines. Count values metrics (e.g. white balance +/- 3 counts) shown are for reference to the original FADGI specifications and are not precisely matched to the new values.

The star system ratings are summarized below. It is important to understand that the star ratings system is an indicator of acceptability of higher error relative to an aim value (i.e. accuracy). Four star requires much less error tolerance relative to an aim than a one star requirement.

- One star imaging should only be considered informational, in that images are not of a sufficient quality to be useful for optical character recognition or other information processing techniques. One star imaging is appropriate for applications where the intent is to provide a reference to locate the original, or the intent is textual only with no repurposing of the content.

- Two star imaging is appropriate where there is no reasonable expectation of having the capability of achieving three or four star performance. These images will have informational value only, and may or may not be suitable for OCR.

- Three star imaging defines a very good professional image capable of serving for almost all uses.

- Four star defines the best imaging practical today. Images created to a four star level represent the state of the art in image capture and are suitable for almost any use.

Our mission is to define what is practical and achievable today, and provide you with the knowledge and the tools to achieve your intended FADGI compliance level.

Generally, in order to avoid future rescanning and given the high costs and effort for digitization projects, FADGI does not recommend digitizing to less than three-star. This assumes availability of suitable high-quality digitization equipment that meets the assessment criteria described (see the section on Quantifying Scanner/Digital Camera Performance), and produces image files that meet the minimum quality described in the *Technical Guidelines*. If digitization equipment fails any of the assessment criteria or is unable to produce image files of minimum quality, then it may be desirable to invest in better equipment or to contract with a vendor for digitization services.

[1] http://www.metamorfoze.com/english/digitization

The FADGI Conformance Program

The FADGI digitization program consists of three elements:

- Technical Guidelines and Parameters
- Best Practices
- Digital Imaging Conformance Evaluation (DICE)

These three elements, when implemented together, form a FADGI compliant digitization environment. FADGI conformance is a process of continuous validation to known and accepted standards, best practices, and adhering to the technical guidelines as detailed in this document. While it is possible to create FADGI compliant images in a physical environment that does not conform to the recommendations in this document, conformance to FADGI recommendations related to the physical environment is highly recommended.

Digital Imaging Conformance Evaluation (DICE) Process Monitoring

The Digital Imaging Conformance Evaluation program (DICE) provides the measurement and monitoring component of a FADGI compliant digitization program. DICE consists of two components:

- Image Targets, both reflective and transmissive
- Analysis Software

The DICE targets have been designed to comply with various ISO specifications, and the parameters as defined in the FADGI program have been validated through years of use at participating Federal agencies.

There are other targets and measurement programs available, but these have not been evaluated and cannot be substituted for use in a FADGI compliant environment.

Certification of FADGI conformance must be measured using DICE targets and analysis software.

DICE Evaluation Parameters

The FADGI guidelines establish quality and performance goals for the four levels of the star ranking system. The DICE conformance testing tool, when used with appropriate testing targets, provides the user with a precise and repeatable analysis of the imaging variables that comprise FADGI star ratings.

The following parameters are evaluated by the DICE program[2]:

Sampling Frequency

This parameter measures the imaging spatial resolution, and is computed as the physical pixel count in pixels per inch (ppi), pixels per mm, etc. This parameter informs us about the size of the original and also provides part of the data needed to determine the level of detail recorded in the file. ISO 12233:2014 defines the resolution measurements.

Tone Response (OECF)

Opto-Electronic Conversion Function (OECF) is a measure of how accurately the digital imaging system converts light levels into digital pixels. ISO 14524:2009 defines the OECF measurement.

White Balance Error

This is a measurement of the color neutrality of the digital file. Ideally, an image of a white reflective object would be recorded digitally as even values across red, green and blue channels, with a value

[2] The implementations can be seen in Appendix A.

approaching the limit of the file format to define white. These specific values are defined in each section of the guidelines.

Illuminance Non-Uniformity

Both lighting and lens performance contribute to this measurement. Ideally, there should be a perfectly even recording of a neutral reference from center to edge and between points within the image. ISO 17957:2015 defines the shading measurements. Specific values are defined in each section of the guidelines.

Color Accuracy

There is no perfect imaging system or perfect method of color evaluation. Color accuracy is measured in DICE by computing the color difference (ΔE2000) between the imaging results of the standard target patches and their pre-measured color values. By imaging the DICE target and evaluating through the DICE software, variances from known values can be determined, which is a good indicator of how accurate the system is recording color. Dice measures the average deviation of all color patches measured (the mean). Refer to ISO 13658:2000 for additional documentation on color accuracy measurement.

Color Channel Mis-Registration

All lenses focus red, green, and blue light slightly imperfectly. This parameter measures the spread of red, green, and blue light in terms of pixels of mis-registration. This parameter is used in the evaluation of lens performance.

MTF/SFR (Modulation Transfer Function / Spatial Frequency Response)

Modulation Transfer Function is a measurement of the contrast difference between the original image and the digital image. MTF is defined as the modulation ratio of the output image and the ideal image. Spatial Frequency Response measures the imaging systems ability to maintain contrast between increasingly smaller image details. Using these two functions, an accurate determination of resolution can be made as it relates to sampling frequency. ISO 12233:2000, ISO 16067-1:2003, and ISO 16067-2:2004 define MTF/SFR measurement.

Reproduction Scale Accuracy (Future Implementation)

This parameter measures the relationship between the size of the original object to the size of that object in the digital image. This parameter is measured in relation to the pixels per inch (ppi) or pixels per mm (ppmm) of the original digital capture. For example, capturing an image of a ruler at 400 ppi will digitally render at the correct size when displayed or printed at 400 ppi. It is critically important in cultural heritage imaging to maintain the relationship to the original size of the object.

The original size of microfilmed documents can only be determined if the filming “reduction ratio” is known. The scale is referred to as 8x or 10x (or other) reduction, indicating that the magnification of the image on the microfilm is 1/8th or 1/10th the size of the original document. This may or may not be known when digitizing microfilm. Unless noted in metadata, the scale of the original will be lost when microfilm is digitized. Microfilm is digitized at the same ppi resolution, regardless of the original “reduction ratio.”

Photographic film cannot be related to a reproduction scale, unless there is a physical measurement in the image to scale to. Photographic film is digitized to appropriate resolutions relative to the size of the film.

Sharpening

Almost all digital imaging systems apply sharpening, often at a point where the user has no control over the process. Sharpening artificially enhances details to create the illusion of greater definition. DICE

testing quantifies the level of sharpening present. There are three major sharpening processes in a typical imaging pipeline: capture sharpening (through camera setting adjustment), image sharpening in post processing, and output sharpening for print or display purposes. Sharpening is usually implemented through image edge enhancement, such as filtering techniques using unsharp masks and inverse image diffusion.

Noise

Digital images contain artifacts that do not relate to the original image, in much the same way that photographic film grain did not relate to the original scene. There are many sources for digital image noise. DICE measures all visible noise and provides a single measurement value.

Skew (Future Implementation)

This parameter measures how straight the image is in the file. This is important because rotating an image by anything other than 90 degree increments involves interpolating every pixel in the image, reducing the effective spatial resolution and integrity of the file.

Field Artifacts (Future Implementation)

Ideally, digitization should only capture what is in the original. However, dust, dirt and scratches almost inevitably find their way into digital files. This parameter quantifies the physical non-image artifacts in a digitization system.

Geometric Distortion (Future Implementation)

Critically important to faithful reproduction of an original is the management of geometric distortion in image capture. Typical camera taking lenses are poorly corrected for this, and create images which exhibit significant distortion even under ideal conditions. If the imaging system is not correctly aligned, other distortions are introduced to the image as well. Cultural heritage imaging requires high quality optics designed for close imaging applications. Typically, these are identified as macro lenses, although that term more correctly refers to their close focus ability. Recently digital correction processes have been created to “correct” for the distortions created by lenses from many manufacturers. While these software approaches are interesting and often effective, they interpolate pixels which cause significant loss of image integrity. High quality lenses designed for copy work will generally have very well controlled geometric distortion, and should not be corrected further through software for master files. ISO 17850:2015 defines the geometric distortion measurements for digital cameras.

File Formats

Master File Format

The choice of master file format is a decision which affects how digitized materials can be used and managed. There is no one correct master file format for all applications, all format choices involve compromises between quality, access and lifecycle management. The FADGI star system tables list the most appropriate master file formats for each imaging project type. Selection of the most appropriate format within these recommended choices is an important decision that should be consistent with a long term archive strategy.

One or more digital master files can be created depending on the nature of the originals and the intended purpose of digitization. Digitization should be done in a “use-neutral” manner, and should not be geared for any specific output. If digitization is done to meet the recommended image parameters and all other