

*Federal Agencies Digitization Guidelines Initiative*

# DIGITIZATION ACTIVITIES

## Project Planning and Management Outline

Version 1.0  
November 2009

**Document Information and Version History**

<i>Title</i>	Digitization Activities
<i>Author</i>	Federal Agencies Digitization Guidelines Initiative – Still Image Working Group
<i>Document Type</i>	Technical Guideline
<i>Publication Date</i>	2009

<i>Date of Revision</i>	<i>Author(s)</i>
September 2009	FADGI v1
August 2008	U.S. National Archives and Records Administration ( <a href="http://www.archives.gov">www.archives.gov</a> ) internal draft by Steven Puglia, Erin Rhodes, and Kevin DeVorsey

## **Table of Contents**

### ○ **Introduction** - 4

Defines the purpose of this document by outlining high-level activities related to the digitization of cultural heritage materials for planning and management purposes. We have defined a generic sequence of activities and workflow for digitization. We relate these activities to library/archival issues, imaging and conversion work, and IT infrastructure in particular. We discuss guidance and related frameworks, including record keeping principles and sustainability factors that should be considered.

### ○ **Assumptions** - 7

Provides general assumptions regarding digitization, including aspects that may differ depending on the rationale for digitizing, as well as on the nature of the originals. As there is no single approach to digitization, specific work processes should be tailored to each individual digitization effort.

### ○ **General Policy Issues** - 8

Lists policy-related issues to be addressed from an institutional perspective, including: policies and procedures relating to digitization, recommendations for sustainable data formats, creation and management of metadata schema, definition of essential characteristics, appropriate information capture and quality levels for digitization, verifying authenticity of digital copies, and ensuring appropriate records management of digital resources, if applicable.

### ○ **Approaches to Digitization Chart** - 9

Illustrates a complement of different approaches to and purposes for digitization, including: improving access to collections, preservation reformatting, digitization for exhibits and reference requests, conservation documentation, classified records review, and digitization of current business records.

### ○ **Digitization Technical Requirements** - 10

Lists components that form the basis of digitization specifications. Technical requirements should address: digital conversion parameters for original materials, recommended data formats and digital object types, assignment of copy or record status, identification of work processes, responsibilities, roles, resources, metadata, identifiers, quality management, etc.

### ○ **Digitization Activities** - 11

Describes the nature and sequence of high-level activities necessary to plan, conduct and manage digitization projects, including: selection, assessment, and prioritization; project planning, management, and tracking; copy status and records management; preparation of originals for digitization; metadata creation and management; digitization; data collection and management; and assessment and evaluation.

### ○ **Digitization Activities Chart** - 14

Provides a visual overview of the relationships and sequencing of digitization activities.

### ○ **Detailed Outline** - 15

Provides an in-depth outline illustrating the management, operational, and assessment activities within each of the four phases of digitization workflow.

### ○ **Infrastructure Issues** - 22

Provides a brief outline of IT infrastructure needs to support and facilitate digitization projects.

### ○ **Resources** - 23

List of resources consulted for this document.

## **Introduction**

This document attempts to define work activities relating to the digitization of original cultural materials. The intention is to outline a generic sequence of high-level activities for planning and management purposes. In this document, we present a conceptual outline that accommodates most digitization approaches, but one that acknowledges all steps may or may not apply, may occur in different sequences, or may occur simultaneously depending on the type of project and the context of the digitization. The activities described and the general sequencing of the work processes were identified using project management outlines from a number of organizations that had significant experience with the digitization of cultural materials.

In this document, we define “digitizing” as the creation of digital objects from physical originals. So-called “born digital” objects (both born-digital content and current business records) are not in scope in this document at this time, although many of the activities outlined in this document will likely apply to born digital materials as well.

We define “digitization” as a complete process that broadly includes: selection, assessment, prioritization, project management and tracking, preparation of originals 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. This document divides the processes involved in a digitization workflow into four main phases:

- Project planning
- Processes occurring prior to digitization
- Digital conversion
- Post-digitization work

Project planning and management, data collection, and quality management are considered to be some of the ongoing activities throughout all four phases of the digitization workflow.

The activities described within each phase address library/archival issues, imaging and conversion work, and IT infrastructure issues. Library and archival issues include preparation of originals for digitization, indexing, collection and creation of metadata of all types, and quality control of the digital versions, indexing data, and other metadata. Imaging/conversion work includes digitization, creation of derivative versions for access, quality control, and metadata creation. IT infrastructure issues include: collection and transfer of data to other systems, networked and Web services, databases, and managed storage and backup. Additional IT infrastructure issues include: short-term/intermediate data storage, backup of digital resources for disaster recover, and safeguards and checks to protect against data loss and to ensure data integrity.

*The Archives New Zealand S-6 Digitisation Standard* provides a similar framework for digitization projects as this document, and lists a set of mandatory requirements for digitization processes. A partial list of those requirements, appropriate for cultural heritage institutions, is listed below – some of the requirements have been reworded to be more generic. We have listed them here as they dovetail nicely with concepts outlined in the four phases of digitization workflow as described earlier. For a more detailed discussion of the Archives New Zealand requirements, go to <http://continuum.archives.govt.nz/files/file/standards/s6.pdf>.

- All digitization and digitization processes must be planned, scoped and documented
- An appropriate digitization approach must be selected, documented and implemented
- Technical specifications aligned to the digitization requirements must be selected, documented and implemented
- Equipment and software aligned to the digitization requirements must be implemented
- Systems to support management of the digital output of digitization must be in place
- Guidelines for the preparation of original collections/records must be documented and implemented
- All digital objects created must be assigned metadata to document digitizing processes and to support ongoing business processes

- Quality assurance and quality control procedures must be defined, documented and implemented
- Digital storage and disaster recovery procedures for digital objects and metadata must be defined, documented and implemented
- Systems for the long-term management of digital objects and metadata must be documented and implemented
- Preservation strategies and processes for digital objects and metadata must be defined, documented and implemented

A large component of post-digitization activities includes content management and preservation within a digital repository environment. This document does not detail those processes, but instead leaves off after submission of digital objects into a managed repository environment. In Open Archival Information System (OAIS) parlance, the document addresses “pre-ingestion” activities, or those activities that take place prior to submission of digital resources to a digital repository for long-term management. This document does not address the processes, procedures, and actions surrounding management of Archival Information Packages (AIP) and subsequent creation and dissemination of Digital Information Packages (DIP). Creation of Submission Information Packages (SIP) is only discussed generically, as it is assumed that different repositories will have different ingest requirements or procedures.

In many organizations, digitization processes may create official “record” copies. If this is the case, policies, workflow, and infrastructure should be designed to ensure the integrity and authenticity of record copies is maintained during the digitization process and through submission of digital copies into various management and access systems and repositories. ARMA International has developed a set of “Generally Accepted Recordkeeping Principles” that suggests records should be created, managed, and maintained according to these principles. These *Generally Accepted Recordkeeping Principles*, <http://www.arma.org/garp/>, include:

- Principle of accountability
- Principle of integrity
- Principle of protection
- Principle of compliance
- Principle of availability
- Principle of retention
- Principle of disposition
- Principle of transparency

Also, we believe factors developed around the sustainability of file formats can inform the conceptualization of digitization activities, related workflow, and supporting IT infrastructure. Several organizations have developed these criteria, including the Library of Congress; the National Library of the Netherlands; the State and University Library, Aarhus, and the Royal Library, Denmark; and the National Archives UK. These sustainability concepts may be applied to the digitization process as a whole, not just to the creation and management of data formats. Factors to consider may include:

- Library of Congress - (Sustainability of Digital Formats, Planning for Library of Congress Collections - <http://www.digitalpreservation.gov/formats/index.shtml>)
  - Disclosure
  - Adoption
  - Transparency
  - Self-documentation
  - External dependencies
  - Impact of patents
  - Technical protection mechanisms

- KB – (National Library of the Netherlands, *Evaluating File Formats for Long-term Preservation* - [http://www.kb.nl/hrd/dd/dd\\_links\\_en\\_publicaties/publicaties/KB\\_file\\_format\\_evaluation\\_method\\_270\\_22008.pdf](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/KB_file_format_evaluation_method_270_22008.pdf))
  - Openness
  - Adoption
  - Complexity
  - Technical Protection Mechanism (DRM)
  - Self-documentation
  - Robustness
  - Dependencies
- KB – Denmark (Denmark, the State and University Library and the Royal Library, *Handling File Formats*, 2004 - <http://netarchive.dk/publikationer/FileFormats-2004.pdf>)
  - Openness
  - Portability
  - Quality
  - Monitoring obsolescence
- National Archives UK - (National Archives, UK, *Selecting File Formats for Long-Term Preservation* - [http://www.nationalarchives.gov.uk/documents/selecting\\_file\\_formats.pdf](http://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf))
  - Open standards
  - Ubiquity
  - Stability
  - Metadata Support
  - Feature Set
  - Interoperability
  - Viability
  - The following additional criteria should be considered for migration:
    - Authenticity
    - Processability
    - Presentation

Note: These sustainability factors are based on file format characteristics and use and do not reflect empirical data obtained through controlled studies.

## Assumptions

- Cultural heritage organizations digitize materials for many purposes:
  - To facilitate access
  - For preservation
    - To prevent loss of information due to:
      - Obsolescence
      - Deterioration
      - Handling
      - Theft or destruction
  - For exhibits, publications, and web use
  - For researcher/patron requests
  - To support current business processes
- There is not a single approach to digitization and metadata creation/management for all projects
  - Type of original, media type, quality and condition of originals, nature of information, preservation risk level, etc., will determine the complement of approaches for digitization and metadata requirements that are matched to the originals
  - The specific work process should be tailored to each individual digitization effort
    - Due to differences in the nature of and the media type of the originals and existing copies
    - Due to the extent and nature of existing descriptive or bibliographic information
    - For maximum efficiency and cost-effectiveness
  - If there is pre-existing descriptive or bibliographic information (metadata), in either hardcopy or electronic form, that is related to candidate materials, it should be collected and come into the digitization work process as early as possible, ideally in the pre-digitization phase
  - More descriptive and technical information (metadata) will be created during the conversion (digitizing) process
  - Descriptive work will need to be done post-digitization but prior to the completion of the project, if the metadata does not exist prior to conversion
  - Approach to digital object identifiers may vary by class of digital object
- These activities encompass all types of originals (manuscripts, books, still photography, maps, plans, artifacts, audio, video, motion pictures, aerial photography, etc.)
- Long-term preservation functionality for digitized versions of originals should be provided by a managed repository environment
  - Accordingly, digitization activities should be aligned with a managed repository and its work processes and requirements
- Digital resources will require metadata at the appropriate level - the level and completeness of the descriptive and technical metadata will vary depending on the original and media type. The level and extent of metadata provided should be determined prior to conversion and should be created according to defined criteria/standards, whenever possible

## **General Policy Issues**

- Policies and procedures relating to digitization activities and the management of digitization projects
  - Define policies and procedures
  - Implement and ensure compliance with policies and procedures
  - Different procedures selected depending on class or category of project
- Determine complement of digital objects, file types, and file formats for both preservation and access digitization projects
- Recommendations for sustainable formats for digital archival records and digital master copies
  - Digital object characteristics (i.e., nature of raster image files, digital audio and video files, etc.)
  - Digital conversion parameters (i.e., technical specifications followed for the creation of digital objects)
  - Minimum complement of metadata required
- Creation and management of metadata schema(s)
  - Define the complement of metadata needed for preservation of digital objects
  - Define appropriate complement of various metadata to ensure management of assets for desired retention time period (short or long term)
  - Determine when in workflow what metadata is added or created
  - Determine where metadata will be stored (embedded or in external system or both)
  - Determine in what formats metadata will exist
  - Determine relationship to identifiers of digital objects
- Definition of essential characteristics (significant properties) of the original – curatorial/archival and technical
- Determination of appropriate approach and quality levels for digitization
  - Approaches to digitization
  - Conversion specifications for preservation reformatting
  - Establish quality management approach
  - Establish metrological approach for scanner and digital conversion equipment performance
- Ensure authenticity of digital copies
  - Verification procedures for digital copies
  - Comparison and review of digital copy to original record from the library/archival/curatorial perspective to ensure digital copy satisfies requirements for authentic digital versions
  - Document chain of custody – both original records during digitization and the digital copies
  - Audit-trail – history of actions on the digital copies and related metadata, from creation through final submission to digital repository
  - Verification of fixity information such as checksums and digital signatures
  - Relationship to identifiers of digital objects
- If applicable, ensure appropriate records management of the digital resources to be created
  - Define records management issues related to the digital copy
  - Define class/status of new copies (e.g., records, copies, non-records, master files, access/distribution/derivative files) in order to determine digitization approaches and management of file types and content over time
  - Identify party/parties responsible for managing the various digital versions and ensure appropriate records management depending on status of digital copies
  - Define procedures and methods for accessing the digital copies for most or all use requests



**Approaches to Digitization**

This table illustrates various approaches to digitization and the purposes for conducting digitization work. The actual work activities may vary for each approach.

Purposes for Digitization of Cultural Heritage Materials							Digitization of Current Business Records
Improved Access to Collections/Records		Classified Records Review (Federal Agencies)	Preservation Reformatting	Exhibits / Publications / Web	Fee and Reference Requests	Conservation Documentation and Object Inventory	Support for Current Business Processes
External Partnerships	Internal Projects						
<p>Copy/Record Status –</p> <ul style="list-style-type: none"> <li>Assumed to be the same as the internal access projects</li> </ul> <p>For potential partnerships, attention should be given to the following prior to start of project:</p> <ul style="list-style-type: none"> <li>Chain of custody methodology, including key identifier</li> <li>Acceptable file formats and quality/compression level</li> <li>Embedded and external metadata</li> <li>Digital image performance requirements – metrics and tolerances</li> <li>Transfer methodology</li> <li>Validation method and confirmation of acceptance</li> </ul>	<p>Copy/Record Status –</p> <ul style="list-style-type: none"> <li>As determined by institution – digital master copies and digital access copies to be managed for the long-term to protect institutional investment – for archives, may treat digital copy as if a permanent record copy, but does not have record status</li> </ul>	<p>Copy/Record Status –</p> <ul style="list-style-type: none"> <li>May be different status depending on the records and projects</li> <li>Assumed to be the same as the internal access projects; when the digital copy replaces original, digital copy is archival record copy (record of originating agency); or when digital copy is preservation copy, retain and manage as if record copy, can become record copy if original deteriorates or is obsolete</li> </ul>	<p>Copy/Record Status –</p> <ul style="list-style-type: none"> <li>When the digital copy replaces original, digital copy is permanent or archival record copy (record of originating agency)</li> <li>When digital copy is preservation copy – retain and manage as if permanent/record copy, can become permanent/record copy if original deteriorates or is obsolete</li> </ul>	<p>Copy/Record Status –</p> <ul style="list-style-type: none"> <li>Assumed to be the same as the internal access projects</li> </ul>	<p>Copy/Record Status –</p> <ul style="list-style-type: none"> <li>Digital Master Copies or Digital Access Copies, unless it is determined preservation reformatting is desirable</li> </ul>	<p>Copy/Record Status –</p> <ul style="list-style-type: none"> <li>May vary</li> </ul>	<p>Copy/Record Status –</p> <ul style="list-style-type: none"> <li>May vary</li> </ul>
<ul style="list-style-type: none"> <li>Digitization of complete groups of holdings</li> <li>Digitization may be done to reference quality levels – for example, see the alternative minimum requirements in NARA’s 2004 <i>Technical Guidelines</i>*</li> <li>Preservation reformatting information capture levels only needed if records are at risk – for example, see the recommended requirements in NARA’s 2004 <i>Technical Guidelines</i>*</li> <li>Minimal complement of metadata to facilitate access</li> <li>Both appropriate digitization information capture levels and the complement of appropriate metadata are necessary to protect institutional investment in the partnership</li> </ul>	<ul style="list-style-type: none"> <li>Digitization of complete groups of holdings</li> <li>Digitization may be done to reference quality levels – for example, see the alternative minimum requirements in NARA’s 2004 <i>Technical Guidelines</i>*</li> <li>Preservation reformatting information capture levels only needed if records are at risk – for example, see the recommended requirements in NARA’s 2004 <i>Technical Guidelines</i>*</li> <li>Minimal complement of metadata to facilitate access</li> <li>Both appropriate digitization information capture levels and the complement of appropriate metadata are necessary to protect institutional investment in the digitization of records</li> </ul>	<ul style="list-style-type: none"> <li>Risk assessment and prioritization recommended – if risk is low, digitization may be done to reference quality levels to facilitate review process – for example, see the alternative minimum requirements in NARA’s 2004 <i>Technical Guidelines</i>*</li> <li>Preservation reformatting information capture levels only needed if records are at risk – for example, see the recommended requirements in NARA’s 2004 <i>Technical Guidelines</i>*</li> <li>Minimal complement of metadata to facilitate access</li> <li>Both appropriate digitization information capture levels and the complement of appropriate metadata are necessary to protect institutional investment in the digitization of records</li> </ul>	<ul style="list-style-type: none"> <li>Requires formal risk assessment and prioritization</li> <li>Digitization of complete groups of holdings</li> <li>Digitization to be done to appropriate information capture levels for preservation – for example, see the recommended requirements in NARA’s 2004 <i>Technical Guidelines</i>*</li> <li>Collection of appropriate metadata for the long-term management and preservation of the digital copies</li> <li>Both appropriate digitization information capture levels and the complement of appropriate metadata are necessary to prevent loss of information</li> </ul>	<ul style="list-style-type: none"> <li>Driven by exhibit, publication, and web schedules</li> <li>Highly selective digitization of records from across institution</li> <li>May require higher quality digitization than other activities</li> <li>May require the collection of more metadata than other digitization efforts - due to the high potential demand and use</li> <li>Both appropriate digitization information capture levels and the complement of appropriate metadata are necessary to protect institutional investment in the digitization of records</li> </ul>	<ul style="list-style-type: none"> <li>Driven by researcher and customer demand. Anniversaries of historic events can have an influence on requests.</li> <li>Range from highly selective individual documents to complete groups</li> <li>Digitization quality ranges based on customer needs</li> <li>Minimal complement of metadata to facilitate access</li> </ul>	<ul style="list-style-type: none"> <li>Should meet a minimum information capture level</li> </ul>	<ul style="list-style-type: none"> <li>Should meet an appropriate minimum information capture level</li> </ul>

\* National Archives and Records Administration (NARA), *Technical Guidelines for Digitizing Archival Materials for Electronic Access – Creation of Production Master Files – Raster Images*, June 2004 (<http://www.archives.gov/preservation/technical/guidelines.pdf>)

## **General Components for Digitization Technical Requirements**

### *Technical Guidelines*

Requirements for capture/conversion of original materials, production workflow, editing and processing operations, use of targets, equipment characterization and performance, quality control, etc.

For example, NARA 2004 *Technical Guidelines for the Digitization of Archival Materials* –

Available online at <http://www.archives.gov/preservation/technical/guidelines.pdf> -

- Types of originals addressed – textual, microfilm, still photographs, aerial photographs, maps, architectural and engineering plans, and objects and artifacts
- Digitization information capture levels-
  - Recommended requirements – preferred minimum requirements for preservation reformatting
    - For textual records – generally considered by the digital library community to be equivalent of 35mm microfilm
  - Alternative minimum requirements – appropriate for reference quality scanning
    - Equivalent to the Transfer Guidance for Federal Agencies
    - Specification used for NARA’s pilot Electronic Access Project

### *Definition of Digital Object Classes and Data Formats*

- Recommendations for digital object types for different types of originals and the corresponding data formats for the digital objects

### *Classification of Digital Copies Produced from Physical Records*

- Define the copy or record status of digital objects, the purpose and type of digital files, and the generation of digital file (master or derivative). Copy or record status is a characteristic of a digital file, and the copy or record status can change over time. The purpose and type of digital file relates to the initial intended use and related information capture levels for digitization. The combination of copy/record status, purpose and type of file, and generation determines both retention as well as management requirements to ensure data integrity, authenticity, disaster recovery, preservation, and access over time.

### *Identification of Work Processes, Responsibilities, Roles, Resources*

- Identification of units and staff that will perform different digitization related activities and
- Identification of current and future IT systems/applications that may be used for different digitization related activities, including storage

### *Metadata for Digitization*

- Defines the recommended complement of metadata that could be considered for a digitization project, and identifies a list of minimum metadata elements for digitization projects, the category of metadata to which they belong, and the level at which they might apply. In general, these elements encompass metadata that documents: the original record, the digital resource, the process of creating the digital resource, changes made to the digital resource over time, and the content
- Identifies methods, tools, systems for implementing metadata in the digitization workflow
- Defines standards, formats, and schemas followed

### *Identifiers*

- Determination of approach to identifiers, whether identifiers are system-assigned, actionable, descriptive, or all of the above; at what level identifiers are applied, what is their role in production workflow, metadata creation, and fixity information; how are they crosswalked across systems or structures; are they local or standardized; etc.

### *Quality Management*

- Defines all activities that determine quality policies, objectives, and responsibilities, as well as the implementation of these activities by processes such as quality planning, quality control, quality assurance, and quality improvement. See ISO 8402, *Quality Management and Quality Assurance – Vocabulary*.

## **Digitization Activities and Phases**

### *Activities*

The processes necessary to appropriately manage digitization projects include the following high-level activities, which can be grouped into a sequence of general phases (see below):

- Selection, Assessment, and Prioritization (to determine what materials will be selected, what projects will be approved, and how they will be prioritized)
  - Determination of access/use restrictions or copyright, condition of records, copy status of digital resources, and approach to digitization and metadata
  - Review and approval process
    - Of digitization projects
    - Of technical and metadata approaches
- Project Planning, Management and Tracking
  - Determine what resources are required and available for projects
  - Communication about and coordination of digitization projects
  - Tracking of records throughout the process
- Digital Copy Status and Records Management
  - Review reasons for digitization and evaluate status of originals
  - Determine copy and record status for digital objects and for metadata to be created
  - Manage and document process appropriately to ensure authenticity of digital copies
  - Finalize status of digital copies and related metadata
  - Update status of original records if needed
- Preparation of Originals for Digitization
  - Bibliographic or archival preparation, preservation preparation, etc.
- Metadata
  - Collection, creation, management, and reuse in other systems of all types of metadata (not just descriptive)
  - Quality assurance and quality control of metadata
  - Validation and verification of metadata (both technical and curatorial)
- Identifiers
  - Determination of format, use in workflow and systems, standardization
- Digitization
  - Digital reformatting
  - Quality management, quality assurance, and quality control of digital copies
  - Metrological assurance and device conformance
  - Validation and verification - curatorial verification of the digital copies; technical verification of digital objects to technical approach
- Submission of Digital Resources to
  - Access and delivery systems
  - To digital repository
    - May include creation and ingestion of submission information package
- Data Collection and Management
  - Entry, collection, import, export, etc. of digital copies and metadata
  - Links to all appropriate IT systems

- Manage and make available digital copies and related metadata
  - May include management of archival information package (content preservation) and provision of access to dissemination information package to end-users
- Assessment and Evaluation
  - Per project
  - Assessment of impact of digitization on other activities –such as business processes
  - Data collection
  - Continual process improvement
  - Project assessment, evaluation, and reporting

### ***Phases***

Digitization can be broken out into a sequence of project phases. In all phases, activities (above) can be grouped into management, operational, and program assessment categories. The phases follow a general sequence of steps, and can be grouped into:

- Project planning activities
- Pre-digitization activities (selection, assessment, and prioritization; preparation, and metadata collection and creation)
- Digitization activities
- Post-digitization activities (submission to delivery and repository systems, data collection and management; making digital copies and associated metadata available; assessment and evaluation). Post digitization activities also include creation and ingestion of a submission information package into a repository, management and preservation of the archival information package, and provision of access to the dissemination information package to end users; however, these activities are not specifically addressed here.

Activities such as project management and tracking, quality management, process improvement, as well as metadata management/collection (of all types of metadata) are ongoing processes that will continue throughout the entire digitization project.

Specific work activities within these phases may take place in a single phase or in more than one phase. As an example, activities like collection/creation of descriptive information may take place at different and/or multiple points during the metadata collection and creation process, the digitization process, and/or the data collection and management process.

Therefore, many activities may take place at different points in the chronology and/or repeat at different stages of the workflow depending on the:

- Type of original, media type, and physical copies available to be digitized
- Condition and usability of the originals and/or copies to be digitized
- Nature of the digitization effort, see approaches listed in the table on page 4
- State and extent of processing done for the originals being digitized
- Nature and extent of descriptive information available in hardcopy and/or electronic form

Many activities are also likely to occur concurrently, rather than sequentially.

Planned digitization projects will likely start at the beginning of the sequence of phases, while other efforts (like exhibits, fee requests, reference requests, etc.) will probably start in the middle of the sequence of phases. For example, exhibits, fee requests, and reference requests may start directly with digitization.

On average\* –

- One third of the effort will be project planning, preservation preparation, management, and oversight
- One third of the effort will be archival description and indexing
- One third of the effort will be the actual digitization

The actual percentages of time spent on various activities will vary depending on the project.

\*For more information on the above estimates see *The Cost of Digital Imaging Projects*, by Steven Puglia, RLG DigiNews, Volume 3, Number 5, October 15, 1999, at:  
<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=0000070511:000006278991&reqid=7049#feature>

**Overview and Sequencing of Activities for Digitization Projects – see the following outline for detailed listing and description of activities**

Sequence of Project Phases –		1. Project Planning	2. Pre-Digitization	3. Digitization	4. Post-Digitization	
<b>Activities –</b>	<b>Management</b> - generally to be done prior to other activities in each phase	<b>Selection, Assessment, and Prioritization</b>	<ul style="list-style-type: none"> <li>- Consider nominations, proposals, and priorities for digitization</li> <li>- Collect information and evaluate all needs – curatorial and preservation</li> <li>- Review and approve projects/work based on defined criteria</li> <li>- Review and resolve restrictions and permissions issues, copyright, etc.</li> <li>- Etc.</li> </ul>			<ul style="list-style-type: none"> <li>- Process improvement – see Assessment and Evaluation (Assess+Eval)</li> </ul>
		<b>Project Management and Tracking</b>	<ul style="list-style-type: none"> <li>- Communication and coordination</li> <li>- Define project parameters</li> <li>- Establish timeline</li> <li>- Track, manage, and document activities</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- All activities continue</li> </ul>	<ul style="list-style-type: none"> <li>- All activities continue</li> </ul>	<ul style="list-style-type: none"> <li>- All activities continue</li> <li>- Process improvement – see Assess+Eval</li> </ul>
		<b>Copy Status and Records Management</b>	<ul style="list-style-type: none"> <li>- Review reasons for digitization and evaluate status of original records</li> <li>- Preliminary determination of copy and record status for digital objects and for metadata to be created</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Finalize recommendations for copy status of digital objects and status of metadata</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Manage and document process appropriately to ensure authenticity of digital copies</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Finalize status of digital copies and related metadata</li> <li>- Update status of original records if needed</li> <li>- Process improvement – see Assess+Eval</li> <li>- Etc.</li> </ul>
	<b>Operational</b>	<b>Preparation</b>	<ul style="list-style-type: none"> <li>- Identify and evaluate needs for records preparation</li> </ul>	<ul style="list-style-type: none"> <li>- Archival/Curatorial prep</li> <li>- Preservation prep</li> </ul>	<ul style="list-style-type: none"> <li>- Perform any additional preparation if needed</li> </ul>	<ul style="list-style-type: none"> <li>- Process improvement – see Assess+Eval</li> </ul>
		<b>Metadata – Collection, Creation, Management and Use</b>	<ul style="list-style-type: none"> <li>- Identify approach for all types of metadata</li> <li>- Define metadata scheme/template</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Identify available descriptive metadata</li> <li>- Collect/create appropriate descriptive metadata</li> <li>- QA/QC on metadata</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Collect/create appropriate metadata – descriptive, technical, administrative, structural, etc.</li> <li>- QA/QC on metadata</li> <li>- Manage metadata</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Collect/create appropriate metadata – desc., tech., admin., struct., etc.</li> <li>- Manage metadata</li> <li>- Link/aggregate digital objects and metadata</li> <li>- QA/QC on metadata</li> <li>- Process improvement – see Assess+Eval</li> <li>- Etc.</li> </ul>
		<b>Digitization – Creation of Digital Objects</b>	<ul style="list-style-type: none"> <li>- Identify technical approaches for digitization</li> <li>- Define technical approach/template</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Finalize technical approach for digitization</li> <li>- Establish QA/QC procedures for project</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Digital conversion</li> <li>- QA/QC on digital objects and conversion</li> <li>- Document conversion</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Process improvement – see Assess+Eval</li> </ul>
		<b>Data Collection and Management</b>	<ul style="list-style-type: none"> <li>- Identification and analysis of existing data</li> </ul>	<ul style="list-style-type: none"> <li>- Access needed data in other IT systems</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Collect and manage new data in appropriate central and local systems</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Update data in other IT systems</li> <li>- Submit digital objects/metadata to managed environment (repository) and other IT/access systems</li> <li>- Process improvement – see Assess+Eval</li> <li>- Etc.</li> </ul>
	<b>Program Assessment</b>	<b>Assessment and Evaluation</b>	<ul style="list-style-type: none"> <li>- Establish criteria and begin collecting data</li> </ul>	<ul style="list-style-type: none"> <li>- Data collection</li> </ul>	<ul style="list-style-type: none"> <li>- Data collection</li> </ul>	<ul style="list-style-type: none"> <li>- Data collection</li> <li>- Project assessment, evaluation, and reporting</li> <li>- Assessment of impact on other activities</li> <li>- Process improvement</li> <li>- Etc.</li> </ul>

**Detailed Outline-** (Incomplete, additional detail to be determined, and will vary from one organization/institution to another - see *Overview and Sequencing of Activities for Digitization Projects* chart on previous page for the scope of activities)

## 1. Project Planning

Selection, assessment, and prioritization

- Selection, assessment, and prioritization of digitization projects/candidate materials
  - Includes both external partnerships and internal access projects
    - Nominations for proposals/priorities for digitization informed by agency/institutional priorities and researcher/public interest
      - Priorities to provide enhanced access to high value and/or high use collections/records
      - Scale, scope, comprehensiveness may be factors in selection and drive priorities for external/partnership digitization
    - Collect information, evaluate needs, and analyze collections/records selected to determine status/extent of description, cataloging, and processing, access/use restrictions and permissions, best format for digitization (images, full text, etc.), physical characteristics (bound/disbound, foldouts, etc.), physical condition, restrictions and copyright, etc.
    - Will the whole series/collection, a selected segment, or a cross section be digitized? Will access be provided via bibliographic records, a finding aid, or some other means? If the collection is treated in selective fashion, will the access tool also be selective or will it describe the entire collection or content body?
    - Digitization done in-house, by a partner, or by a contractor
      - Development of contract, RFP, procurement, etc.
    - Preservation reformatting
      - Priorities for digitization informed by formal risk assessments
    - Classified Records Review (for federal agencies)
      - Priorities for digitization informed by formal risk assessments
    - Exhibits and Publications
      - Priorities may be driven by Exhibits schedule
      - Evaluation of whether digitizing more than is required for an exhibit would benefit an existing digital collection or future digitization projects
    - Fee and Reference Requests (Keep digital copies produced across institutions? May be a decision to be made by staff based on importance of collections/records or illustrates a type of original?)
    - Conservation Treatment Documentation and Object Inventory
    - Digitization of Current Business Documents
- Approve projects and work based on defined decision-making criteria, including available resource and capacity analysis

Project management and tracking

- Project management
  - Communication and coordination
    - Coordinate digitization activities across institution
    - Define organizational roles and responsibilities
    - Plan and establish staff resources
    - Identify any constraints and challenges - relating to technical, staffing, financial, and scheduling issues
  - Acceptance and review of project proposals
    - Access driven projects
    - Risk assessments for preservation reformatting projects

- Identification and definition of project parameters and approaches to digitization and metadata (consult existing metadata schema if possible)
- Determination of identifiers and/or file naming approaches
- Identification of any IT issues: databases, software, storage needed, Web issues, etc.
- Identification of available bibliographic records, finding aids, and existing descriptive metadata, plan for processing/cataloging
- Review/approval of procedures for conversion/reformatting, metadata, digital copies, etc.
- Ensure compliance with specification/guidelines for digital resource creation and metadata, and with pre-defined templates and profiles
- Creation and management of appropriate metadata schema
- Manage workflow for all activities
- Management of project documentation, including compliance with industry standards
- Identification of available resources – staff, supplies, equipment, etc.
- Determine if any staff training is needed
- Address issues of conversion site if applicable (particularly with partner projects) in terms of transfer of originals, security, etc.
- Establish project timeline
  - Project timeline may be influenced by other requirements; if so, resources required to complete the digitization effort may need to be changed to meet specific deadlines
- Project tracking
  - Track, manage, and document activities and inventory projects
  - Document all procedures and processes

#### Copy status and records management

- Review reasons for digitization and evaluate status of original records/originals
- Preliminary determination of copy and/or record status for digital objects and for metadata to be created

## 2. Pre-Digitization

- Project management continues
- Project tracking continues
  - Tracking location of originals during pre-digitization processes
- Data assessment and aggregation
  - Establish access to any existing metadata, documentation, cataloging, or archival description to be used to facilitate the digitization process and the intellectual organization of the digital resources
    - Identification of bibliographic records, finding aids, indices, folder lists, inventories, etc. in both hard copy and electronic format
    - Identification of electronic metadata held in management systems and access systems
    - Quality assurance, quality control, verification and validation
- Preparation
  - Curatorial/archival preparation of physical originals/records
    - Analysis of originals (formats, organization, condition, copies, size, etc.)
    - Physical and intellectual organization



- Collect and record a more detailed level of descriptive metadata during the course of curatorial/archival preparation work to enhance description in existing systems
- Create, assign, and record appropriate records management/administrative metadata for new digital resources
- Batch records for conversion
- Preservation preparation
  - Evaluation of physical condition and readiness for scanning
  - Holdings maintenance, if needed
  - Conservation prep, if needed
  - Batch records for conversion
- Requirements review
  - Define metadata requirements for different collections/groupings/classes of resources and determine **minimal** level of appropriate metadata to provide adequate access to and long-term preservation of digital copies
    - Identification of appropriate metadata schema or templates
    - Identification of appropriate minimum complement of metadata
    - Descriptive, administrative, technical, and structural metadata
    - Indexing– if done before scanning
    - Records management metadata, if applicable
  - Determination of identifiers and/or file naming approach
    - In registry
    - Assigned by system or repository
    - Actionable at file level, resolved by system
    - Identifiers used by/in descriptive systems
    - Identifiers used in production workflow
    - Original identifiers
    - Role of identifiers in fixity, authenticity auditing and reporting
    - Application of identifiers at what level (collection, series, item; digital object, file)
    - Crosswalking of identifiers across systems
    - Standardized identifiers or local identifiers
  - Definition of essential characteristics – curatorial/archival and technical
    - Define legal admissibility/authenticity of digital copies of records, if applicable
  - Determination of appropriate approach and quality levels for digitization
    - Approaches to digitization (including image capture specifications, testing and evaluation, workflow, header information, image processing, compression, file naming, file directory structure, file formats for archiving and for presentation, etc.)
    - Conversion specifications for preservation reformatting
    - Determine any special production needs
  - Versions - determination of types and number of digital versions to be created during project
    - Define copy type/record status for resources being created (such as preservation master, production master, derivative files, etc.)
    - Define levels of access and storage for copy types
  - Determination of naming and directory structure schemes

- Determination of file storage needs
- Establish QC/QA Procedures
- Define user interface and digital resource delivery requirements, if necessary
- Copyright and privacy issues
  - Review and identify use and access restrictions on collections/items
  - Resolve and restrictions and permissions issues
- Digital copy status and records management
  - Finalize recommendations for copy status of digital objects and status of metadata
- Determine and assign responsibility for managing the digital objects and metadata
  - May vary depending on copy/file types for versions of the same resources

### 3. Digitization

- Project management continues
- Project tracking continues
  - Monitor status and products of all activities
  - Check in and check out of items to production unit
- Data entry
  - Record any pre-existing metadata needed to begin conversion (may include job tracking information, descriptive metadata, etc.)
- Digital Conversion
  - Capture done according to specifications in-house, by partners, and/or by contractors
  - Image target use for performance verification
  - Device conformance testing and calibration
    - Initial and on-going testing of digital image quality and equipment based on established benchmarks and specifications
  - Digitization of existing documentation, if not already in electronic form
    - Digitization of descriptive information, finding aids, indices, folder lists, inventories, etc. if not in electronic format
  - Perform any correction/editing/processing to digital files
  - Image evaluation – objective and subjective
  - Create and track production metadata
  - OCR and text conversion/mark-up, rekeying, etc.
- Technical, structural, administrative, and descriptive metadata creation and collection
  - Define requirements for and record metadata for different collections/groupings/classes of resources at different levels
  - Create and record/embed metadata into appropriate systems/headers
  - Auto characterization and manual and automated collection of technical and other metadata to carry forward as files are moved into other systems
- Indexing – minimal intellectual organization of digital objects to match the appropriate level within the archival descriptive hierarchy or to match the intellectual organization of the collection. Indexing is primarily geared towards describing and organizing large groups of digital versions of physical records. Indexing provides a level of association and organization of digital resources so they can be effectively searched and retrieved.
  - Role of identifiers

- Quality management - quality assurance and quality control of digital copies and metadata to ensure conformance to guidelines
  - As with any manufacturing process, exceptions or defects can consume an inordinate amount of resources; the further downstream the error detection, the greater the resource use to correct
  - Defect identification and inspection and verification of files
    - Automated quality assurance/quality control for both digital objects and for related metadata (all types of metadata – including technical, administrative, descriptive, etc.)
    - Follow up by staff on problems identified by automated checks
    - Statistically valid sampling checks by staff, automated identification of resources to be checked
  - Rework for error identification
  - Ensure compliance with templates/profiles
  - Follow established metrology protocols and document certifications, or correct and replace as required
  - Documentation of quality assurance/quality control process
  - Create and record QC/QA metadata
- Data entry/import
  - Import technical, structural, descriptive, production, administrative, rights, QC/QA metadata into appropriate systems on local level
  - Import assets into appropriate systems on local level
  - Collect and manage new data in central and local systems
- Version control
  - Define and record relationship between types of files (such as preservation master, production master, derivative files, etc.)
  - Automate production of derivative files and versions
  - Automation of metadata into and out of header tags and files (such as XMP, IPTC, etc.)
  - Perform inspection and verification of derivative files and versions
  - Create and apply checksums to appropriate versions
  - Create batches
  - Aggregate multiple versions, files, and metadata files into a “package” for submission/delivery into storage
  - Role for identifiers
- Copy status and records management
  - Manage and document process appropriately to ensure authenticity of digital copies

#### **4. Post-Digitization**

- Project management continues
- Project tracking continues
- Copy status and records management
  - Finalize status of digital copies and related metadata
  - Update status of original records/originals if needed
- Complete bibliographic/archival description and collection and creation of any additional appropriate metadata (descriptive, structural, administrative, technical) not collected in earlier processes
- Manage metadata

- Finalize the complement of metadata needed
  - Appropriate complement of various metadata to ensure management of assets for desired retention time period (short or long term)
- Quality assurance and quality control of metadata and digital objects
  - Conformance to standards, data types, templates/profiles
  - Accuracy
  - Defect identification and error correction
    - Automated quality assurance/quality control for both digital objects and for related metadata (all types of metadata – including technical, administrative, descriptive, etc.)
    - Follow up by staff on problems identified by automated checks
    - Statistically valid visual checks by staff (i.e., color and tone accuracy), automated identification of resources to be checked
    - Record actual rework/defect correction efforts
    - Documentation of quality assurance/quality control process
- Curatorial/archival validation and verification of digital versions in comparison to originals from curatorial/archival perspective to ensure digital copies satisfy requirements for authentic digital versions
- Technical validation to industry specifications for well-formed digital objects and data formats; and assessment of digital objects to verify they meet local profile and requirements
- Make digital objects and metadata available to staff and researchers
  - Deliver digital objects via web-based/delivery systems for research
  - Deliver high-quality digital products via the web and via optical media
- Aggregate and associate digital objects and metadata files for packaging and transfer
  - Create and associate multiple low resolution derivative files
  - Assign checksums
  - Export – flexible packaging of both digital objects and metadata for delivery into other systems using different metadata schema
    - Submit resources to access/delivery systems and make resources available online
    - Submit resources to digital repository
  - Export metadata in different formats to other systems
  - Export digital files to other systems
  - Acceptance/confirmation of export/submission of digital objects and metadata into other systems
- Update metadata in other management and access systems as needed to synchronize or replace with new metadata generated during digitization projects
  - Linking of metadata between systems
- Provide routine reference to digitized records via on-line systems
- Track and associate new digital/analog versions to the physical originals
- Manage digital resources in appropriate actively managed storage environment
  - After submission of completed digital objects and related metadata to long term digital repository
    - Ensure provenance and authenticity of digital resources
    - Ensure data integrity
    - Ensure disaster recovery

- Project assessment, reporting and evaluation
  - Project Assessment
  - Web, Image File, and Database Usage Analyses
  - Cost-Benefit Analyses
- Assessment of impact on other activities
  - Assess effects of digitization on traditional reference activities (e.g., online access, in-person access, send all source analog content offsite?) and researcher requests, and update procedures
- Identify and correct problems and errors relating to both digital objects and related metadata
  - Correct problems/deficiencies on a routine basis for all categories of digitization
- Lessons learned
  - Unexpected results, scoping errors, etc.
- Process improvement – as needed update workflows, tools, procedures, policies, etc.

## **IT Infrastructure Needs**

### ***Access:***

Access to both digital files and metadata will be needed by both the public (online) and by internal staff for the purposes of research, exhibits, publications, sale, etc.

- Provide a centralized IT workspace so copies are accessible to all staff during the work process in order to complete description, quality control work, etc.
- Versions will be moved into other systems for access, display, presentation, etc.
  - Identify new digital versions in management systems to reduce duplication of digitization efforts and to minimize handling of records
- Metadata will be moved into other systems for access, display, presentation, etc.
- System to assign/register identifiers
- System to resolve identifiers, if actionable

### ***Managed Storage:***

Infrastructure to store, manage, and provide access to digital copies

- Data storage - master files, access files, metadata, and data migration issues
- Need to ensure the- data integrity, disaster recovery, and authenticity of the digital resources created
- Provide minimal bit preservation activity
- Accept packages of digital files, versions, metadata, and information about submission process (verification and validation information, etc.)
  - Ensure viability of data and maintenance of essential characteristics
  - Incorporate checksums, validation, and verification functionality
  - Monitoring
  - Track change history to digital objects
  - Authenticity/provenance chain
- Perform backups and redundancy to appropriate levels to ensure data integrity and disaster recovery
- Define server requirements - develop configuration management plan
- Network security issues
- Systems documentation
- Plan and budget for systems upgrades
- Site licenses and hardware/software maintenance contracts
- Move objects into one or more long-term destinations
- Ability to transfer digital objects and metadata into other systems for access purposes
- Ensure appropriate intellectual control of digital resources
  - Synchronization of metadata and digital objects
  - Updating of metadata and digital objects
  - Manage relationships and associations between versions/multiple components, parent-child relationships, etc.
- At some appropriate point, digital resources and metadata move into digital repository

## Resources

Federal Agencies Digitization Guidelines Initiative (FADGI)  
<http://digitizationguidelines.gov>

FADGI Glossary  
<http://www.digitizationguidelines.gov/glossary.php>

Library of Congress, *NDLP Project Planning Checklist*.  
<http://memory.loc.gov/ammem/prjplan.html>

*NC Echo Guidelines for Digitization 2007 Revised Edition*  
<http://www.ncecho.org/dig/digguidelines.shtml>

Washington State Library, *Digital Best Practices*  
<http://digitalwa.statelib.wa.gov/newsite/best.htm>

Northeast Document Conservation Center, *Handbook for Digital Projects: A Management Tool for Preservation and Access*, 2000.  
<http://www.nedcc.org/resources/digitalhandbook/dman.pdf>

Bibliographic Center for Research, Collaborative Digitization Project, *Best Practices and Publications*  
<http://www.bcr.org/dps/cdp/best/index.html>

*NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, 2002.  
<http://www.nyu.edu/its/humanities/ninchguide/>

*The Archives New Zealand S-6 Digitisation Standard*  
<http://continuum.archives.govt.nz/files/file/standards/s6.pdf>

ARMA International, *Generally Accepted Recordkeeping Principles*  
<http://www.arma.org/garp/>

Library of Congress, *Sustainability of Digital Formats, Planning for Library of Congress Collections*  
<http://www.digitalpreservation.gov/formats/index.shtml>

National Library of the Netherlands (KB), *Evaluating File Formats for Long-term Preservation*.  
[http://www.kb.nl/hrd/dd/dd\\_links\\_en\\_publicaties/publicaties/KB\\_file\\_format\\_evaluation\\_method\\_27022008.pdf](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/KB_file_format_evaluation_method_27022008.pdf)

The State and University Library, Aarhus, Denmark, and the Royal Library, Copenhagen, Denmark, *Handling File Formats* (2004)  
<http://netarchive.dk/publikationer/FileFormats-2004.pdf>

National Archives, UK, *Selecting File Formats for Long-Term Preservation*.  
[http://www.nationalarchives.gov.uk/documents/selecting\\_file\\_formats.pdf](http://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf)

National Archives and Records Administration, *Technical Guidelines for the Digitization of Archival Materials for Electronic Access – Creation of Production Master Files – Raster Images*, June 2004.  
<http://www.archives.gov/preservation/technical/guidelines.pdf>

Puglia, Steven. *The Cost of Digital Imaging Projects*. RLG DigiNews, Volume 3, Number 5, October 15, 1999.  
<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=0000070511:000006278991&reqid=7049#feature>