From Digitization to Repository - A Case Study in Creating a Managed Environment

Steven T. Puglia, Preservation and Imaging Specialist, U.S. National Archives and Records Administration, Washington, DC, USA; Kevin De Vorsey, Preservation Analyst, Preservation Programs, U.S. National Archives and Records Administration, Washington, DC, USA; Erin Rhodes, Digital Imaging Specialist, U.S. National Archives and Records Administration, Washington, DC, USA

Abstract

This article explores the idea of extending the concept of a managed environment, beyond the digital repository and forward in time to include the entire digitization process, for an extremely large and complex institution. Rather than managing digitization independently of the digital repository, this article suggests that it is advantageous to work from the beginning of the digitization effort towards the end goal of long-term preservation of digital objects and related metadata in the digital repository. In particular, this article briefly examines the U.S. National Archives and Records Administration's developing approach to managing digitization projects, including: characteristics of digital objects produced during the digitization process; working data models and approaches for collecting all types of metadata to be done locally in the digitization labs and in centralized environments; where this information is held; digitization activities and influence/impact across functional areas; and discussion of how tools and processes that have been implemented at the local lab level can inform the definition and development of a centralized, institution-wide information technology infrastructure to support digitization at NARA.

Introduction

The National Archives and Records Administration (NARA) is an independent agency in the Executive Branch of the United States Government. NARA is responsible for preserving and making available approximately ten billion permanently valuable records of the Federal Government and Presidential records, overseeing twelve Presidential Libraries (with another being planned), operating fourteen Regional Archives, and has also established ten Affiliated Archives located around the country. Finally, NARA provides records management guidance to Federal Agencies, operates fifteen Federal Records Centers nationwide, and manages the records of the Legislative Branch in the Center for Legislative Archives.

As expected, NARA's holdings are diverse, geographically dispersed, and complex. The holdings include large numbers of textual records on paper; records on media such as photographic film, magnetic tapes, wax cylinders, etc.; and increasingly, records in electronic form on computer readable media such as compact disks and hard-drives. NARA has decades of experience developing policies and designing facilities for the storage and preservation of traditional physical archival holdings. This paper describes efforts by NARA to address the analogous needs for the creation of and management of digital copies of physical records, and for the preservation of new digital copies in a managed environment for as long as they are needed.

NARA has experience digitizing most forms of physical records, and the digital copies are primarily used for remote research access, publication, and exhibition. The traditional media used to preserve records via reformatting, such as magnetic audio and videotapes, as well as still, aerial, motion, and micro photographic films, are becoming scarce due to a "market-driven" migration to computer-based digital technology. In response, NARA is adopting digitization as a preservation tool and accelerating the on-going transition of preservation reformatting processes to digital equivalents.

The Managed Environment

Archives and cultural heritage institutions have well-established methods governing the acquisition and management of physical holdings and collections. What are the analogous methods for digital versions of original records? Historically, most digitization has been done by institutions to achieve limited goals, usually enhanced access and reduced handling, with little regard for where and how the digital resources will be managed and preserved for the long-term. In many cases, digital repositories have been created as an afterthought and independently of the digitization processes used to create and document the digital resources. Perceived differences in custodial responsibility (Information Technology vs. Archival), as well as very real differences in funding sources and availability, have often contributed to this condition. What are the work processes, information technology (IT) infrastructure, and policies needed to enable institutions to achieve higher-level, long-term objectives beyond the short-term goal of completing projects?

For many institutions, we feel the "managed environment" needs to be extended beyond the digital repository and forward in time to include the entire digitization process. In this way, the complete "life cycle" of the digital resources being created can be managed more effectively, and institutions can ensure the usability and viability of both digital objects and related metadata in digital repositories. As institutions adopt digital reformatting workflows to replace traditional reformatting, the benefit of a managed work environment will allow institutions to maximize the benefits of digitization. Content created for specific purposes, like an exhibit, can have very legitimate uses for other purposes, such as preservation. The lack of a managed environment to deal with the digitization and the resulting resources can limit the potential benefits of re-purposing digital assets.

What do these work processes and supporting information technology infrastructures look like? What are conceptual approaches that may provide cost-effective answers to this

question? The managed environment should include: the collection of information about the intellectual and physical organization of the holdings to be digitized; the definition of approaches to digitization and management of the process; the creation of additional metadata to document the originals, the digital objects, the process of creating the digital objects, and any changes to the digital objects; the validation and/or authentication of the digital copies in comparison to the originals; the ability to package digital objects and related metadata for use in access systems and for submission to digital repositories; etc.

Many cultural institutions seem to be managing digitization programs independently of their digital repositories, assuming they are fortunate enough to have one. Repository policies tend to be prescriptive regarding acceptable digital object types, data formats, and metadata. When it comes to managing digitization initiatives, we suggest borrowing a couple habits from Stephen R. Covey's *The 7 Habits of Highly Effective People* and strive to "Be Proactive" and to "Begin with the End in Mind." Rather than being prescriptive, arguably it is better to be proactive in regard to the creation of digital objects and metadata to be managed within the repository. Rather than managing digitization independently of the digital repository, it is better to work from the beginning of the digitization effort towards the end goal of long-term preservation of digital objects and related metadata in the digital repository.

Digitization work processes and supporting information technology infrastructure should reflect these perspectives. Conceptually, the scope of the digital repository can be pushed forward in time to the point of creation of the digital objects and metadata. Digitization activities should be geared towards integrating appropriate functionality across all phases and activities of digitization to ensure the relatively automated submission of appropriate and high-quality digital objects and related metadata to digital repositories. We have talked to a number of institutions and experts in the field about this concept, and there is general agreement this approach makes sense, but at this time no one seems to be managing digitization or designing IT infrastructure to function in this manner.

Digitization Activities

Early in 2008, NARA began conducting an agency-wide, six-month Business Process Reengineering (BPR) effort to define and/or rethink work processes relating to digitization. Activities were already underway within the agency on these efforts, and build on years of experience dating back to NARA's Optical Digital Information Storage System (ODISS) technology evaluation project of the mid-1980s, digitization experience dating to the early 1990s in the Special Media Preservation Division laboratories within Preservation Programs, and the pilot Electronic Access Project (EAP) from the late 1990s. Other informative efforts include development of a Digitization Capture Architecture (DCA) in the Federal Records Centers to support digitization of other agency records.

Also, since the early 2000s, NARA has been working on defining, designing, and building the Electronic Records Archives (ERA) infrastructure to manage, preserve, and make available electronic records produced by Federal agencies. ERA will serve as NARA's institutional digital repository for the long-term management and preservation of digital copies of physical records. Finally, the Special Media Preservation Division digitization labs have local tools (digitization equipment for all types of records, local digital storage to support production, supporting software for lab activities, etc.) that will inform the scope and scale of tools for a centralized, institution-wide IT infrastructure.

Information collected and the concepts developed from the BPR will be used to define IT infrastructure to support managing digitization of records and, for the short-term, of the digital objects and related metadata created during these efforts, until all work is completed and can be submitted to ERA.

We believe digitization can be broken out into a sequence of project phases. In all phases, activities can be grouped into: *Management, Operational, and Program Assessment activities*. The major phases of activities follow a general sequence of steps, and can be grouped into:

1) Project planning activities
2) Pre-digitization activities
3) Digitization activities
4) Post-digitization activities

The processes necessary to manage and conduct digitization projects appropriately include the following high-level activities:

i. Selection, Assessment, and Prioritization
ii. Project Management and Tracking
iii. Copy Status and Records Management
iv. Records Preparation
v. Metadata
vi. Digitization
vii. Data Collection and Management
vii. Assessment and Evaluation

Activities such as project management and tracking, as well as metadata collection (of all types) are ongoing processes that will continue throughout an 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 such as 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:

a) general record type, media type, and physical copies available for the records to be digitized
b) condition and usability of the records and/or copies to be digitized
c) nature of the digitization effort (improve access to records; preservation reformatting; exhibits, publications, and web use; and fee and reference requests)
e) state and extent of archival processing for the records being digitized
f) nature and extent of descriptive information available in hardcopy and/or electronic form

Many of these 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. At NARA for example, digitization relating to exhibition, publication, web use, fee requests, and reference requests often start directly with digitization of the records. In these cases, the other necessary activities, such as metadata creation, will happen after digitization.

The creation of integrated IT infrastructures to coordinate and support all phases of digitization and related activities, up to and including submission of digital objects and metadata to the digital repository, will serve to create the managed environment we describe above.

Digital Objects

The more refined the definition of the digital objects to be created, the easier it will be to define infrastructure to automate and/or facilitate handling of the objects throughout the entire process. NARA has been working on defining concepts for our purposes, ranging from the identification of the components of a digital object, through the makeup of a submission information packages (SIPs) to be brought into a digital repository.

NARA's concept of a digital object is similar to definitions found in the digital library community: one or more content files (which may include derivatives and alternate versions) with associated metadata, and possibly other associated files (such as targets, color profiles, etc.), packaged together physically and logically (by means of a wrapper or bundling format, or expressed by relationships via associated metadata).

Digital objects should be defined prior to creation and may be characterized in several ways: by level of aggregation, by digital object type, by digital object class, by file type, and by essential characteristics linked to each of the above. A simple digital object (single content file and associated metadata) represents a single record or a single item. A compound or complex digital object (multiple content files and associated metadata) represents a multi-page or multi-part record or item. Compound digital objects may be packaged in a data format to hold multiple digital objects as a single file (such as multi-page Tiff, PDF, etc.) or they may be wrapped in a packaging format such as MXF (Material Exchange Format). At a higher generic level, multiple digital objects may represent an entire file unit, series, or other aggregation of records; these objects are likely to be largely undefined except through directory structure organization or file naming conventions. We anticipate multiple objects would be accompanied by locally-defined XML files in the directories, which would contain all types of associated metadata.

We are working to establish approaches to creating digital objects based on record type and media type. To help define the criteria for creation and eventual management of digital objects produced at NARA, we have drafted a four-level decision-making hierarchy: Record Type, Digital Object Type, Copy Type and File Type, and Data/File Format(s).

The first or top level is original record type (i.e., static, dynamic, digital), which informs the next level down, the digital object type level. Essential characteristics from an archival perspective will likely derive from an analysis of the important features of specific content or general record and media types that should be carried forward over time. These considerations inform decisions relating to the type of digital object to be created when digitizing original records.

Digital objects are created according to the suitability of the object type to the record or media type of the original. NARA is working on defining essential or significant characteristics of the original records, digital objects and related data/file formats; defining the metadata that will be used to describe these characteristics; and determining whether they should be maintained over time. Technical essential characteristics are a part of the digital object and data format, and are determined primarily at time of creation of the digital object, depending on information capture levels and amount and quality of metadata. Currently, NARA does not have formal descriptions for all the types of digital objects to be created. We feel there are at least four groups aligned with the original records – raster images (created from textual, cartographic, still and aerial photography, etc.), audio recordings, video recordings, and motion pictures.

Next, digital objects to be created should belong to a class of copy. Digital objects within each class of copy share similar attributes, and classes of copies are tied to a specific record status, retention period, level of effort to manage, and approaches in creating the digital copies from physical records. Classes of digital copies not only determine how a digital object will be managed and stored in a digital repository environment (such as level of service or effort to preserve) and accessed in an online environment, but also to what quality or informational capture level a digital copy is initially created during digitization. Management requirements include type of storage, any actions that are applied to the files, and the policies and procedures needed to ensure data integrity, authenticity, and disaster recovery for an appropriate length of time.

Currently, NARA has defined three classes of digital copies and each class contains one or more file types:

1. Digital Record Copies comprised of
- *a. Archival Master Files*

2. Master Copies comprised of
- *a. Preservation Master Files*
- *b. Production Master Files*
- *c. Reference Files*

3. Access Copies comprised of
- *a. Derivative Access Files*
- *b. Access Files*

The digital file types are created according to the technical parameters followed during the creation of digital files, the reason for reformatting and the quality level desired, the completeness of metadata produced, and the risk level of the originals. File types are therefore directly linked to a digitization specification or guideline that provides a rationale or approach to the creation of consistent digital objects so they can be managed in the same manner over time. In this case, reference files are linked to a

minimum specification as outlined in the NARA's *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) and production masters and preservation masters are created to recommended or higher specification as outlined in the NARA 2004 guidelines. The level of information capture is linked to the level of effort to maintain these objects over time.

To summarize, digital record copies are master copies that have been designated as having record copy status (that is, digital versions of physical records identified as being the official copy). The archival master files replace the original physical records. For master copies, it is presumed the original physical records still exist. All types of master files can stand in for the original records, if necessary. Access copies may be produced from any type of master file or directly from the original physical records to an information capture level that is less than NARA's minimum. These copies are produced for specific purposes, including use in access systems, or to fulfill a reference or fee request.

Finally, these classes of digital objects and related file types are in turn linked to preferred data or file formats at the lowest level of the hierarchy. The data/file formats are selected according to the suitability for the record/media type of the original, for the digital object type, and for the copy and file type in regard to the maintenance of essential characteristics and the sustainability, usability, and authenticity of both the digital representation and the specific data or file formats.

Metadata

NARA collects records-related metadata, including records management information (accession information, rights and restrictions information) and descriptive information about records and the agencies that created them (record group, series, file unit, and item level information). This descriptive information has been managed in both central and local systems at NARA. For many years, NARA has followed a descriptive framework reflecting the lifecycle of archival materials, the *Lifecycle Data Requirements Guide,* which serves as a guide for descriptive metadata to be entered into NARA's Archival Research Catalog (ARC), the online catalog of NARA's holdings.

Recently, NARA has been working on evaluating and formalizing a complement of metadata for digitization projects that includes additional types of metadata beyond descriptive, such as technical, administrative, structural, and preservation. These types of metadata describe the digital object, the digitization process, changes made to the resource over time, and the content itself.

Digital object-related metadata, currently collected only in the local digitization lab infrastructure, consists of a complement of the following types:

- *Records management metadata indicating the record status of the digital copy (permanent or non-record)*
- *Technical metadata describing attributes of the digital copy, the conversion process, file format information, specifications followed, validation and verification procedures, and assessments of image quality*
- *Structural metadata describing relationships between multiple components of a digital resource. Replicates the original by relating digital objects to the structure and sequence of the record. May also describe multiple versions of the digital object.*
- *Administrative metadata describing rights, restrictions and permissions; the workflow and digitization process; archival verification of essential characteristics; quality control information; and preservation-related metadata, such as audit trails, checksums, change history, and digital provenance information*

We assume metadata will be collected at different points in the digitization workflow, and metadata will reside in different places and different systems to varying degrees of completeness. Metadata will be mapped to or exchanged between various systems; therefore, we expect a degree of redundancy in the metadata in these systems. We are in the process of defining a rationale for metadata storage to identify acceptable redundancy, and defining the authoritative systems for each type of metadata. To this end, NARA is exploring where and in what systems different types or complements of metadata are held, and how they relate to different leveling of metadata at the project, record, and digital object levels. We envision a hierarchy or model of where metadata is held, including different types of metadata associated with different levels in the hierarchy.

For example, at the lowest level of the hierarchy, metadata may be embedded in the file name itself or represented in a directory structure. This provides basic identification of the file and relationships to other files. At the next level up, metadata is embedded within the data file itself, in the file header or stored and output in XML, depending on the data format. At the next level up, a full complement of metadata is packaged as part of the digital object or is contained within a wrapper or packaging format (such as MXF for audio and video objects). This approach most likely contains metadata that will be required for a SIP, such as technical, descriptive, verification and validation information, and information about the object's authenticity, produced according to a local XML schema. This approach contains the most complete information about the digital object, and may contain pointers to systems with more complete metadata of other types.

At the next level in the hierarchy, all metadata associated with the digital files is stored in an XML file alongside the content files in a directory. The metadata in the XML file will apply to all files in the directory, and can be easily parsed out into the appropriate systems for that type of metadata. At the next level up, metadata would be created within systems with specific functionality for that type of metadata. For example, the workflow software and the digital asset management system in the digitization labs will contain the most complete technical information, information about quality control, workflow processes, and technical specifications followed, but are less likely to contain complete archival descriptive information. At the highest level, metadata resides within an authoritative or official system for that type of metadata (for example, ARC for the most complete descriptive metadata, ERA for the most complete preservation and records management information).

On a local level, the digitization labs have been working on establishing a data model for metadata collected during the digitization process that can be implemented in the indexing or cataloging module of the workflow software currently used in the labs. The software provides enough flexibility that for any project, we can define where metadata will be held (file name, headers, other systems) and where and what types of metadata will be held redundantly (in other systems or locations). We will be able to export any level of the metadata hierarchy to any external system at NARA, in any format or mapped to any XML schema that is required.

Generally, in the digitization labs' implementation, metadata will be separated into different catalogs depending on its function and at what level it applies. Job or project tracking information would be applied at a global or job/project level. Project metadata, or information that is common to all digital objects within a project, would also be applied at a global level. This might include records management information, copy status, technical and administrative metadata, and rights and restrictions information. Descriptive metadata, pulled from existing descriptions, finding aids, and other sources, would reside in project-specific catalogs or tables.

Other specific project-related metadata would reside in workflow tables in the database, as well as specifications particular to a project, such as number and sizes of derivatives produced. These tables would include metadata at the project, record, and digital object levels. Specific digital object metadata might include checksums, file properties, etc. All tables are linked via project or digital object identifiers.

IT Systems and Digitization Lab Tools – Existing, Under Development, and Under Consideration

The size, scope, and complexity of NARA have been previously described. It is an organization that is geographically distributed, and houses vast and varied holdings of records that range in terms of value, classification, and legal status. While digitization promises to help make this complexity invisible to researchers and the general public by providing access to electronic versions of records regardless the physical location of the original record or the requesting researcher, the IT infrastructure being implemented to enable this and support a repository is necessarily complex. Currently, no single system will provide the ability to collect, create, manage, store, and provide access to all of the data that provides intellectual and physical control of both the physical holdings and the digital copies that represent them.

In the past, the limitations inherent in wide area networking technologies, computer hardware and software, and application development have unfortunately resulted in "stove-piped" silos of information, each developed to address a specific requirement, but unable to integrate with other applications in a manner that would support the development of a digital repository. Work is underway to address these problems by developing integrated applications across the enterprise that promise to support and manage wide-scale digitization with a managed repository.

Central to these plans are two systems that will support the repository concept: the Electronic Records Archive (ERA) and the Holdings Management System (HMS). ERA is currently being tested and implemented across NARA and will allow federal agencies to transfer electronic and digital copies of records directly to NARA for permanent storage. It will ultimately provide the network, storage and preservation infrastructure that will support a managed repository.

HMS will subsume over one hundred applications, databases, spreadsheets and paper ledgers currently being used to manage the storage and preservation of the holdings. It will serve as a central repository for all information relating to the history of the records while they are in NARA's custody and will allow NARA to track their use, make informed decisions, and take action to minimize the risk of information loss due to handling, theft, material obsolescence, and environmental factors such as temperature and humidity. HMS will manage circulation of the records and will subsume the Master Location Register (MLR) that was first developed in the 1990's to support the movement of records into the College Park, MD, storage facility. As its name implies, the MLR provides for the physical control of records and is used solely by NARA staff to determine the stack and shelf location of the boxes that house the records.

The descriptive information that provides NARA's intellectual control over the holdings is housed in the Archival Research Catalog (ARC). ARC serves as the main finding aid at NARA and allows staff and researchers (via a web interface available at http://www.archives.gov) to search for records based on the federal record hierarchy, subject, or keywords. While it does contain some electronic copies of records, it lacks the ability to appropriately manage digital files or to maintain and display the complex relationships that exist between records and their various paper, film, and electronic copies. It is anticipated that ARC will eventually be subsumed into ERA, but the design and structure of this application is as yet undefined.

Several applications will allow NARA's digitization labs to manage digitization workflow, file creation, and metadata collection at a local level. Two applications currently being implemented are the Image Workflow Solution (IWS) from Stokes Software Inc. for image processing, workflow management, quality control, optical character recognition and indexing or metadata collection; and Telescope, a digital asset management application from North Plains Systems for locating completed work and managing project level metadata. These applications will allow staff to name files in a standardized manner, associate metadata from applications such as HMS, ARC, and ERA, and embed appropriate descriptive, structural, technical, and preservation metadata to provide a core context for each file so that they can be used independently of any particular application or technology environment. Other applications the digitization labs have or are considering using to support collection and management of technical and preservation metadata, validation and authentication of files, and the creation of SIPs are: JHOVE (JSTOR/Harvard Object Validation Environment) or the Library

of Congress' CIV implementation of JHOVE, Harvard's Batch Builder, and an XML tool suite (currently Altova MissionKit).

Integration of the major applications for the automated submission of digital objects and metadata to the digital asset management application is being investigated – this will represent a direct test of the concepts we are forwarding in this paper. We will investigate the integration of the other applications for validation of digital objects and metadata, creation of submission information packages, the automated creation and exchange of XML formatted data, and the potential auto-generation of static browse web-pages for projects that have minimal item level descriptive metadata.

Prior to the full operating capability for the Electronic Records Archives, NARA will have to bridge current and planned systems with interim IT infrastructure for managing and making available digital copies of physical records. As records are digitized, we anticipate the Holdings Management System will be updated to indicate new digital copies exist. The interim infrastructure will need to have the following functionality:

Access- Access to both digital files and metadata will be needed by both the public (online) and by internal NARA 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.
- Move versions 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
- Move metadata into other systems for access, display, presentation, etc.

Managed Storage- The infrastructure will need to store, manage, and provide access to digital copies until ERA is able to assume those responsibilities:

- Provide storage for all types of digital copies and file types (all types of master and access files), metadata, and address data migration issues
- Ensure data integrity, disaster recovery, and authenticity of the digital resources created
- Provide minimal bit preservation activity
    - Accept packages/aggregations 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
- Move objects into one or more long-term destinations
- 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, move digital resources and metadata into ERA

NARA has committed to digitization as a tool to achieve its strategic goals of preserving its holdings in perpetuity while providing access to them to the public. All of the previously described applications (and many others as yet unplanned) will be needed in order to access, index, point to, or at least indicate the existence of digital files that NARA and our partners create and to preserve them over time. We anticipate the BPR will result in a recommendation to integrate current and planned IT systems by implementing a digitization management infrastructure that bridges the Holdings Management System and the Electronic Records Archives. The digitization management infrastructure will reflect the concepts and models we have described in this paper; and will provide centralized tools for all staff to complete necessary work including collection of existing metadata, the submission and update of metadata generated and modified during digitization, management of the digitization process, the creation of submission information packages, and the automated submission of digital objects and metadata to the digital repository. If we define appropriate concepts and models, and build integrated infrastructure that reflects the concepts and models, we believe we will have created an appropriate managed environment – from creation to the digital repository.

Author Biography

Steven T. Puglia has been a Preservation and Imaging Specialist at NARA for 20 years. He has been involved with preservation reformatting for over 20 years and digital imaging for over 15 years. He is a co-author of NARA's 2004 Technical Guidelines for Digitizing Archival Materials.

Kevin De Vorsey has been involved in preservation and access programs for over 15 years. Formerly of NARA, he is now a Preservation Analyst in the National Digital Heritage Archive, National Library of New Zealand, Wellington, NZ.

Erin Rhodes has been a Digital Imaging Specialist at NARA for 6 years. She is a co-author of NARA's 2004 Technical Guidelines for Digitizing Archival Materials.