



**W N Y L R C**

Western New York  
**Library Resources Council**

**THE WESTERN NEW YORK  
REGIONAL DIGITIZATION PLAN**

*March 15, 2006*

*Approved by the Regional Automation Committee*

*Friday March 31, 2006*

*Approved by the WNYLRC Board of Trustees*

*Tuesday, April 11, 2006*

Western New York Library Resources Council  
4455 Genesee St.  
PO Box 400  
Buffalo, NY 14225-0400  
Tel: (716) 633-0705  
[www.WNYLRC.org](http://www.WNYLRC.org)

## TABLE OF CONTENTS

Preface From The Committee Chair .....	3
Introduction.....	4
The Digitization Plan: Plan Purpose.....	5
Western New York Regional Digital Environmental Analysis .....	6
The Western New York Regional Digitization Plan.....	11
<i>Guidelines &amp; Standards</i> .....	12
Introduction To The Guidelines & Standards.....	13
Broad View Of The Digitization Process .....	14
Guidelines For Selection Criteria For Digitizing Collections.....	15
Descriptive Metadata Guidelines.....	23
Digital Imaging Best Practices .....	34
Resource List - Software .....	50
Resource List - Vendors .....	52
Glossary.....	54
<i>Appendix</i> .....	64
Digitization Survey Results.....	65
Digitization Plan’s Complete Goals, Objectives, And Activities.....	80
Regional Digital Heritage Advisory Subcommittee Members .....	85

*The development of this plan was funded or supported in part through the Federal Library Services and Technology Act Funds (LSTA), awarded to the New York State Library by the Federal Institute of Museum and Library Services.*

*This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.*

## **PREFACE FROM THE COMMITTEE CHAIR**

*Pamela Jones*  
*Roswell Park Cancer Institute*  
*Dr. Edwin A. Mirand Library*

### **Why digitize?**

The Internet has become the vehicle by which librarians and archivists are making collections and information about collections widely available. Creating digital assets of physical materials and making these assets available via the Internet has given patrons the ability to evaluate research materials remotely. Librarians are now challenged to create web sites with eye appealing informational product, i.e. marketable information resources, preferably with a Google-like box to provide searching capability across many formats of data. Libraries are digitizing collections of historical books, postcards, posters, photographs, art work, oral histories and more for researchers to view. Along with this plethora of digital assets comes the need to attach meaningful metadata to provide access to items in their collections.

Thoughtful consideration must be given to what should be digitized. Is it best to scan each page of a book or would it be better to create something more akin to an online exhibition which will give the patron an overview of many items in a collection without digitizing each page of each item? For librarians with a deeply ingrained methodology of cataloging each item in a collection with a bibliographic record, the impulse to digitize everything will be hard to ignore. Digital collections need not be an exact reproduction of a library's physical collection, but can be the electronic version of some interesting highlights from the physical collection.

Finding aids are the standard method of description for archival collections. For archives, digitization will often involve creating electronic versions of paper finding aids. Electronic finding aids are most commonly found on the web encoded as HTML or EAD (Encoded Archival Description). EAD is structured to facilitate the hierarchical arrangement of a finding aid and lends itself to the creation of databases of finding aids which can enhance searching abilities for researchers. When considering digitization, archivists will first consider creating digital finding aids. The digital finding aid is the information needed for researchers to evaluate whether or not a visit to the repository will be necessary. Of course, any finding aid with at least a few digital images included with it will be very appealing to look at!

### **Why collaborate?**

Since funding is scarce, libraries, museums, and archives can no longer operate efficiently in isolation. Cultural heritage institutions can provide richer collections of digitized resources by working collaboratively with each other. Collaborations may also increase the perceived values of collections in addition to fulfilling research and educational goals and objectives at the institutional level. Collaboration allows each partner to contribute its strengths in order to make the collaborative whole stronger.

WNYLRC is encouraging the development of collaborative digitization projects among diverse institutions with complementary materials. WNYLRC will strive to bring together diverse projects in a regional collaborative digital program. This program's long range view highlights the region's library and archival collections through a portal with a federated search product enabled to facilitate searching across library collections and archival finding aids.

## INTRODUCTION

In the last several years, the number of digitization projects has continued to increase, with several major players becoming involved (e.g., Google). While the word “digitization” has become commonplace in the library literature, an understanding of all the processes involved in a digitization project is not as mainstream. With the exponential rise of digitized materials available on the Internet, it seems there are still many small to medium-sized institutions that have not begun to think about digitizing their materials and how doing so would increase outreach and audience. Indeed, publicly available digital assets have proven to increase an institution’s foot traffic because more people are aware of the institution’s holdings.

In 2004, the Western New York Library Resources Council (WNYLRC) was already thinking about how to involve more of its members in digitizing their own materials. To begin its efforts, WNYLRC created the Regional Digital Heritage Advisory Subcommittee (RDHA) under the auspices of the Regional Automation Committee (RAC). In 2005, under the leadership of the RDHA, WNYLRC began a two-year project to train librarians in the region on aspects of digitization, to understand current member activities with regards to digitized materials, and to develop a plan to increase the number of digitized materials available from WNYLRC members. This project is supported by Federal Library Services and Technology Act (LSTA) funds, awarded to the New York State Library by the Federal Institute of Museum and Library Services.

This is not WNYLRC’s first foray into digitization. In 1999 – 2000, WNYLRC embarked on discussions which yielded the “Guidelines for Selection Criteria for Digitizing Collections” published by WNYLRC in September 2000. The guidelines were developed to help libraries and archives evaluate materials, and to determine whether or not the current electronic world provides a feasible and beneficial platform for their collections.

In Western NY, as in other regions, cultural heritage organizations have embraced the online environment to further their ability to share information with a wide audience. Digital collections of images, finding aids, historical material and ephemera are appearing on library, museum and archives web sites around the globe, making information about collections accessible to an international audience of researchers. Since 1999, several WNYLRC members have begun digitization projects/programs including, but not limited to, the University at Buffalo, Prendergast Library, Niagara University, and Canisius College.

This document is an outgrowth of both the work that WNYLRC has done since 1999 and the activities that the RDHA Subcommittee has engaged in over the last twelve months. This plan – to be implemented over the next several years – captures the committee’s vision, goals, and ideas for increasing the number of digitized assets available from WNYLRC members’ collections. It is hoped that this plan – written by WNYLRC members for WNYLRC members – will significantly and positively impact the region for years to come by:

- Creating an infrastructure that will support innovative digital services to patrons in WNY.
- Developing a regional collaborative digital program to enhance the access to and retrieval of digital information resources for member libraries, library systems, and their patrons.

## **THE DIGITIZATION PLAN: PLAN PURPOSE**

### **WNYLRC's Mission**

The Western New York Library Resources Council is dedicated to enhancing access to information, encouraging resource sharing, and promoting library interests for its members that serve the people of Western New York.

### **Plan Purpose Statement**

The Regional Digital Heritage Advisory Subcommittee's vision is to provide a standards-based but flexible framework for the encouragement of digital projects by member institutions. These projects will lead to the creation of a regional collaborative digital program to provide optimum access to content available in the region and to encourage resource sharing among member libraries and library systems in support of WNYLRC's mission and vision.

### **Audience**

Who should use this plan? This document was written primarily with libraries and archives in mind; however, any organization would benefit from the information contained within. In addition, any organization that is considering applying for RBDB funding or collaborating with/through WNYLRC on a program should refer to this document.

## WESTERN NEW YORK REGIONAL DIGITAL ENVIRONMENTAL ANALYSIS

### Overview

In the fall 2005, WNYLRC conducted an online survey of its members as well as members of the Documentary Heritage Program (DHP) and local government communities. The survey asked for input on a number of issues, including digitization efforts at their own institution, digitization-related services desired from WNYLRC and information on the materials available to be digitized in the region. 109 staff from WNYLRC member libraries and library systems responded to the survey and eleven members of the DHP and local government communities. Information culled from the survey results provided a current environmental picture of digitization in the Western New York region. The number one reason WNYLRC member libraries and library systems would digitize part or all of their collections would be to increase their patron's access to these resources. But when determining what to digitize, many of the respondents rely on grant funding to support the activities. It is also true that while interested in a possible collaboration with other libraries in the region, 72% of those who responded to the question are not aware of other collections in the region that share a similar scope or subject theme with their institution. (A complete summary of all survey results can be found in the section "Digitization Survey Results.")

### Who is digitizing?

- 29.3% of respondents have already begun digitizing items from their collections and 87.1% of this group felt that they would digitize again.
- Although every type of member library and library system was represented, more than half of the respondents were from academic institutions. The academic libraries seem to be the most capable of digitizing items from their collections and are currently the primary group digitizing in our region.
- Some public libraries and historical societies are digitizing materials.

### What is being digitized?

The most common types of materials reported by our members that would most likely be digitized are:

- Photographs (50.4%),
- Bound materials such as books and journals (37.6%)
- Unbound sheets of paper smaller than 8.5" by 11" (33.9%)
- Oversized unbound sheets of paper larger than 8.5" by 11" (32.1%) and photograph slides or negatives (32.1%).

## How can WNYLRC assist in regional digitization efforts?

In charting the direction WNYLRC should take, the responses regarding services that WNYLRC might provide were extremely important. The top five digitization-related services that WNYLRC members want are:<sup>1</sup>

- Training
- Provide networking opportunities for librarians
- Assist in identifying funding sources
- Provide information on suitable digitization vendors
- Assist in writing a grant for a digitization project

Additional information the survey provided regarding possible roles for WNYLRC:

- 54% of the respondents would consider using a regional digitization center established by WNYLRC if the cost was lower or the same as other vendors. Considering that WNYLRC is a non-for-profit organization and member driven, it is feasible that WNYLRC could provide member access to a regional digitization center on a cost recovery basis.
- 31% of the respondents indicated they would consider using the same image management software that other institutions in the region are in order to provide greater access through a regional portal. WNYLRC could provide access to a single management software either through consortial pricing or regional funding.

Although members of the DHP and local government communities do not have access to all of WNYLRC's services, the committee felt it was important to discern the needs of the broader community, since those institutions could be potential collaborators. The top four digitization-related services this group mentioned were:<sup>2</sup>

- Training
- Locate potential collaborators
- Assist in identifying funding sources
- Provide information on suitable digitization vendors

Many of the RHDA committee members represented other WNYLRC committees. They not only brought to the RHDA committee the concerns of the committees they represented, but also took back to those committees what they learned as a result of committee discussions, viewing survey results, and attending workshops.

As committee members reviewed the survey results and other input the committee received, they commented that:

- Based on the activity already occurring in the region and survey results, the region seems to be on the cusp of embarking on major digitization projects.
- In order to digitize their collections, smaller institutions need assistance with funding and staffing as well as the expertise WNYLRC may be able to provide.

---

<sup>1</sup> This is based on the total respondents who voted for "will use" or "likely to use" each specific service.

<sup>2</sup> This is also based on the total respondents who voted for "will use" or "likely to use" each specific service.

- There is a great need in the region for institutions to collaborate on digitization projects so that institutions can benefit from each other's strengths.
- Nationally the collaborative trend for digitization projects seems to be on a state-wide level. Noting the difficulties New York State has had in the past in developing state-wide initiatives, e.g. virtual reference service and the creation of union catalogs, it isn't feasible to think that New York State is considering a state-wide digital project. To that end, each geographic region through its regional library consortium<sup>3</sup> had made a foray into a region-wide digitization program or project.
- The region can and should learn from the successful collaborative digitization projects in other geographic regions in New York and the work that the other 3R councils have completed.

---

<sup>3</sup> The Reference and Research Library Resources multi-type systems in New York State are often referred to as "3R councils."

## **Strengths, Weakness, Opportunities and Threats: “SWOT” Analysis**

SWOT Analysis is a tool used by businesses and organizations to determine new markets, new revenue streams or funding initiatives. It helps uncover opportunities and reveal any challenges that an organization may not have been aware of. In developing a digitization program, it is necessary for WNYLRC to understand what the potential audience or market is and what resources WNYLRC has in place by asking key questions regarding WNYLRC's strengths, weaknesses, opportunities, and threats in relation to the goal or goals identified.

With this information in mind, the Regional Digital Heritage Advisory Subcommittee has identified the following:

### ***Regional Strengths***

- History of cooperation among libraries within the region on past projects (i.e. wnylibraries.org)
- Strong collections with local and national interest documenting, for example, Love Canal, the Niagara Movement, Erie Canal development, Pan American Exposition
- Increased requests for access to collections on the history of the region particularly from education institutions (i.e. K-12 sector)
- Variety and richness of collections
- Infrastructure to support digitization collaboration including technology, personnel, and collections.

### ***Regional Weaknesses***

- Lack of institutional funding for digitizing collections
- Missed opportunity with Institute of Museum and Library Services (IMLS) seed project funding
- Lack of central expertise agent for consultation and training
- Lack of personnel at local institutions to select, organize and digitize materials
- Lack of high-quality digitizing equipment available to member libraries and library systems – i.e. color book scanners, scanners for oversized items
- Only one-third of member libraries and library systems have online catalogs available through WNYlibraries.org, the regional virtual union catalog
- Lack of knowledge among members and member libraries on aspects of digitization
- Unlike the large-scale collaborative efforts that are occurring in other states, this region is affected by the lack of a state-wide initiatives to fund large-scale digital collaborative projects

### ***Regional Opportunities***

- WNYLRC can take a lead role in developing a regional collaborative digital program and digitization center that provides a quality and cost effective way for members to digitize, is focused on libraries and archives, and that could provide a cost-recovery revenue stream for WNYLRC
- WNYLRC could become the place to turn for expertise in digitization
- State education curriculum mandates the use of primary source materials in the classroom

- A regional collaborative digital program would provide greater access to unique historical materials available in the region and provide materials essential to distance learning initiatives
- WNYLRC's capability in encouraging collaboration among small and large institutions and members and non-members
- Development of regional standards for implementing digitizing projects
- Access to a large amount of materials available in the region that no one else is digitizing AND that should have increased access

### ***Regional Threats***

- New York State flat funding that affects WNYLRC and member libraries and library systems and their institutions to embark on digitization projects
- Difficulty in long-term preservation of digital assets due to changing digital standards and formats

Based on this environmental picture, the RHDA subcommittee was able to develop goals, objectives and activities that would provide a clear direction for WNYLRC and its member libraries and library systems in pursuing a collaborative regional digital program.

## THE WESTERN NEW YORK REGIONAL DIGITIZATION PLAN

The plan will support digitization efforts in Western New York, resulting in a significant expansion of local history materials available through the World Wide Web, searchable through a central online portal.

### How We Will Achieve this Outcome

#### 1. Create an infrastructure that will support innovative digital services to patrons in WNY.

- *By providing training and information needed by member libraries and library systems to explore digitizing their own collections.*
- *By re-positioning and training WNYLRC staff to support digitization projects.*

#### 2. Develop a regional collaborative digital program to enhance the access to and retrieval of digital information resources for member libraries, library systems, and their patrons.

- *By appointing a permanent Regional Digital Heritage Advisory Group knowledgeable in digitization planning, under the auspices of the Regional Automation Committee, to implement the Regional Digitization Plan.*
- *By offering presentations from other successful digitization projects that exemplify “best practices” and that encourage member libraries, library systems and other organizations to utilize recommended standards for imaging, metadata, and selection criteria.*
- *By implementing a digitizing laboratory that includes basic and advanced scanners, computers with software, work space, and WNYLRC staff to maintain it.*
- *By increasing the ability of WNYLRC and its member institutions to support a regional collaborative digital program through exploration of funding and support services (grants, cost shares, sharing hardware, fee-based services).*
- *By creating a small pilot database of content from WNYLRC member institutions that will serve as the model for the expansion of the regional collaborative digital program.*
- *By developing WNYinfo.org as the open and easily accessible single interface portal for the regional collaborative digital program.*
- *By identifying the long term preservation needs of digital assets for both WNYLRC and member libraries and library systems.*

## **GUIDELINES & STANDARDS**

## INTRODUCTION TO THE GUIDELINES & STANDARDS

As the area of digitization has progressed, many guidelines and standards have been developed to address specific situations. As the committee discussed the digitization activities that it hopes will occur in the region, it realized that it needed to look at the published guidelines and standards, and recommend those that fit into the direction that WNYLRC wants its members to take. The committee also gathered other resources that it knew would be useful for inclusion in this plan.

Therefore, the committee members and its consultant worked to gather the following resources for your use:

- Broad View of the Digitization Process
- Descriptive Metadata Guidelines
- Digital Imaging Best Practices
- Guidelines for Selection Criteria for Digitizing
- Resource List - Software
- Resource List - Vendors
- Glossary

Some of the sections have an emphasis on imaging. The committee recognizes that the technology allows audio, video and multimedia to be digitized, but feels that many projects begin with unbound and bound papers, photographs, and slides. Therefore, the committee has placed its emphasis on these materials in this document. This, however, does not preclude an institution in the region from digitizing other types of media.

## BROAD VIEW OF THE DIGITIZATION PROCESS

The questions below will help the organization to gain a quick overview of the entire process and its responsibilities. These questions can be used as discussion points as the organization contemplates starting a digitization program. An organization should also read and respond to the statements in section “Guidelines for Selection Criteria for Digitizing Collections” in order to assess its readiness.

1. Does digitization fit the organization’s mission?
2. Is there a known potential audience for the materials that are planned to be digitized?
3. Will digitization increase access, functionality or intellectual control?
4. Will digitizing these materials fill a need that is currently unmet?
5. Are the materials in the public domain or can proper rights be secured?
6. Is funding in place for the digitization program?
7. Does the staff have the correct knowledge and skills to work on this program?
8. Does the organization have access to the proper equipment (hardware and software)?
9. Will a digital asset management system or library management system be put in place so that users can locate and access the digital assets quickly and easily?
10. Has the collection been adequately processed and described?
11. Will metadata be created for the digital assets?
12. Is staff or funding available to maintain and preserve the digital assets once they are created?

If an organization answers “no” to any of the questions above, then it should take time for further evaluation before proceeding. In some circumstances, the organization may want to do more to prepare for the digitization effort.

## GUIDELINES FOR SELECTION CRITERIA FOR DIGITIZING COLLECTIONS

The Regional Digital Heritage Advisory Committee's *Guidelines for Selection Criteria for Digitizing Collections* is developed to help libraries and archives determine whether or not digitization provides a feasible and beneficial platform for their collections. While individual collections necessarily will face specific problems when digitizing, these *Guidelines* can help any organization recognize and address universal issues that must be considered when converting holdings into electronic form for the Internet.

The criteria are broken down into "Administrative" and "Operational" concerns, and are further formulated to address important questions regarding organizational mission, information value and appraisal, audience and access, over-all feasibility of goals, the scanning process<sup>4</sup>, and post-scanning and maintenance issues. Within these categories, there is always an emphasis on general feasibility, organizational support, costs, staffing issues, and the quality and content of the records themselves.

Because the digitizing process necessitates the involvement of many experts within a given organization, *it is essential* that the *Guidelines* be examined and filled out by such a diverse group as archivists, reference librarians, collection development and information technology staff, and administration, for example -- people whose areas of expertise would allow for multiple perspectives on the variables involved in deciding the particular costs and benefits of digitizing.

These *Guidelines* can serve as a more traditional selection criteria for institutions that are prepared, financially and structurally, to digitize, but have no particular collection in mind. The various points to consider presented in the *Guidelines* ought to give an organization an idea as to which collection(s) might be the most viable in today's electronic universe.

It is also advisable to utilize these *Guidelines* both before initiating and after completing a digital project. In this way, an institution might better clarify some changes that can serve to produce the next digital project more effectively.

Conversion into electronic format is neither particularly easy nor inexpensive. While not meant to dissuade institutions from digitizing, the *Guidelines*, and perhaps the differing answers received by the many people involved in this step of project planning, will underscore particular areas deserving of attention.

**There are no "correct" answers, but a preponderance of "disagrees" or marked conflict in responses to any particular criterion indicates that discussion and investigation must ensue in order to guarantee that the organization possesses the resources to adequately digitize, exhibit, and maintain its online collection, thus benefiting the institution itself as well as the public at large.**

---

<sup>4</sup> Although this section specifically mentions scanning, the same/similar questions would apply to the digitization of audio, video and multimedia materials.

## I. Administrative Criteria: Mission, Acquisition, Access

### 1. Mission

Consider why the organization wishes to digitize. What are the benefits, and do they support the institutional mission? Are all stakeholders fully aware of what is being digitized and for what purpose? Will digitization support the acquisition focus? Are the financial resources available? Is there a mechanism in place (designated staff or a committee) to evaluate the progress made toward enhancing the organization’s goals and objectives through a digitization program?

Additionally, keep in mind that digitizing has shown itself effective for information access, but is not yet universally accepted as an appropriate means of preserving the original item or “preservation reformatting.” (Preservation reformatting can be defined as a new format provides a faithful rendering of a printed text from the original item.) Funding is more likely obtained for access to the digital material.

- ⇒ The project has the support of the organization's main governing body and staff.
- ⇒ Creation and maintenance of digital access are realistic activities in relation to the program's capacity.
- ⇒ The digitization project supports the organizational mission.
- ⇒ The project's goals are consistent with the mission's objectives.
- ⇒ The Internet is a feasible tool for supporting the institutional mission,
  - ⇒...for increasing institutional visibility
  - ⇒...for attracting funding sources
  - ⇒...for attracting potential donated collections
- ⇒ The materials to be digitized are the main acquisition focus of the collection or,
- ⇒ The materials to be digitized complement other holdings of the institution or,
  - ⇒ The materials to be digitized are outside the main acquisition focus, however, they are of significant research/informational value to a wider audience outside the scope of the institution’s primary audience.
- ⇒ The main project goal is associated more with information access than material preservation.

	Strongly Disagree	2	3	4	Strongly Agree
⇒ The project has the support of the organization's main governing body and staff.	1	2	3	4	5
⇒ Creation and maintenance of digital access are realistic activities in relation to the program's capacity.	1	2	3	4	5
⇒ The digitization project supports the organizational mission.	1	2	3	4	5
⇒ The project's goals are consistent with the mission's objectives.	1	2	3	4	5
⇒ The Internet is a feasible tool for supporting the institutional mission,	1	2	3	4	5
⇒...for increasing institutional visibility	1	2	3	4	5
⇒...for attracting funding sources	1	2	3	4	5
⇒...for attracting potential donated collections	1	2	3	4	5
⇒ The materials to be digitized are the main acquisition focus of the collection or,	1	2	3	4	5
⇒ The materials to be digitized complement other holdings of the institution or,	1	2	3	4	5
⇒ The materials to be digitized are outside the main acquisition focus, however, they are of significant research/informational value to a wider audience outside the scope of the institution’s primary audience.	1	2	3	4	5
⇒ The main project goal is associated more with information access than material preservation.	1	2	3	4	5

**Comments:**



**3. Audience and Access**

Consider the target audience of the digital product. Audience will dictate the site content, intellectual level of documentation, modes of access, employable search features and web browsers, scanning practices, etc. Consider that a site with material intended primarily for school children will be necessarily different from that targeted to scholars, which will also vary from a site that is developed for a general or multi-purpose audiences. Both intellectual content and format as well as end-user technology must be considered.

If the goal of digitizing is archival preservation, all material in the collection ought to be present. Collections of pre-selected or “sample” materials will be less costly and perhaps more user-friendly, but this selection must be done by considering what the institution wishes to convey to the audience. If access is the goal, it is especially essential to consider the process in terms of the real and projected target audiences.

- ⇒ The organization has reviewed the records to be digitized and made certain that there are no legal restrictions on access and use by end-users.
- ⇒ There exist appropriate written policies concerning duplication, exhibition, lending, publishing, and other uses of digital images from the collection.
- ⇒ The institution can provide expertise for the level of arrangement and description deemed adequate for access on the site where the digitized images will be available.
- ⇒ Digital images fit the (perceived) needs of users and will better serve them than other physical organization or other access methods.
- ⇒ Digital images and accompanying metadata are more feasible to create and helpful to the user than digital finding aids for the collections the materials represent.
- ⇒ There is enough -- and a consistent quality of -- material to support a “publication” of the materials.

	Strongly Disagree	2	3	4	Strongly Agree
⇒ The organization has reviewed the records to be digitized and made certain that there are no legal restrictions on access and use by end-users.	1	2	3	4	5
⇒ There exist appropriate written policies concerning duplication, exhibition, lending, publishing, and other uses of digital images from the collection.	1	2	3	4	5
⇒ The institution can provide expertise for the level of arrangement and description deemed adequate for access on the site where the digitized images will be available.	1	2	3	4	5
⇒ Digital images fit the (perceived) needs of users and will better serve them than other physical organization or other access methods.	1	2	3	4	5
⇒ Digital images and accompanying metadata are more feasible to create and helpful to the user than digital finding aids for the collections the materials represent.	1	2	3	4	5
⇒ There is enough -- and a consistent quality of -- material to support a “publication” of the materials.	1	2	3	4	5

**Comments:**

**II. Operational Criteria: Funding, Staffing, and Technological Concerns**

**1. Feasibility of goals and institutional capability**

Consider the funding sources, costs, expertise, staff, equipment, technology, time required to create and maintain a viable digital project. Development of a digital product must be considered on a number of levels, and institutions must determine if they have the staff and other support required for each. Consider whether or not the funding institution can dictate site content or other specifics. Oversized or fragile materials may need to be outsourced and/or preserved. Consider the attendant costs and time delays associated with these processes.

*Digitizing as reformatting:* a technological issue not unlike microfilming.

- ⇒ A comprehensive reformatting assessment has been completed and approved by the organization.
- ⇒ Digitization is determined to be the best reformatting alternative for these materials.

*Digitization as selection:* as most digitization-for-access involves only a small portion of the entire universe of existing material on a subject, the traditional library selection process is required; “acquisitions” money, though, is spent on technology and staff, rather than for acquiring physical items.

- ⇒ Digitization is determined to be the best means of accessing the materials and the information in the materials.

*Digitization as publication:* Web sites as surrogates for books and other forms of publications. Think of the creation of a digital collection as the production of a book. Can the institution afford the time and cost of investing in research, authorial, and editing skills requisite to create a focused, coherent, well-indexed, properly source-identifying, *quality* digital product? Too, what is the relative importance of appearance vs. content?

- ⇒ Digitization is determined to be the best possible way to produce a publication based on the materials in the collection.

There exists adequate staff support and knowledge in the following areas:

- ⇒ Material selection (appraisal)
- ⇒ Preparation (arrangement and description)
- ⇒ Preservation (if necessary)
- ⇒ Image capture
- ⇒ Quality control
- ⇒ Technical support
  
- ⇒ Funding is available or assured to complete the project and to maintain it indefinitely.

Strongly Disagree	2	3	4	Strongly Agree
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

**Comments:**



**Comments:**

3. **Post-digitization/Maintenance**

Consider that the cost of future maintenance (tracking the migration and updating of files, for instance, due to technological changes/advancements) only is projected to be 100% of the original project price tag every ten years. Consider internal staff support as well as external support available locally.

⇒ Resources exist for the ongoing maintenance, updating, and security requirements for master files and derivative files.

Resources exist to develop and implement collection access through:

- ⇒...Post-digitization manipulation,...
- ⇒...Web site development,...
- ⇒...Metadata tags and indexing/cataloging/EAD

⇒ The project has the technological capabilities as well as the staff to allow and incorporate user feedback regarding ease of use, clarity and content, speed of retrieval, etc., of the web site.

⇒ Provisions are in order to respond to increased needs for reference and related services.

⇒ Provisions are in order to enhance the digital product, through the inclusion of additional materials, should significant related collections be acquired.

⇒ Regular updating of information and links is possible and can be scheduled.

⇒ A self-assessment instrument can be developed to provide ongoing feedback on cost-effectiveness, user reactions, and other aspects of the program.

	Strongly Disagree	2	3	4	Strongly Agree
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	
1	2	3	4	5	

**Comments:**

**Sources Consulted:**

- Council on Library and Information Resources, Washington, D.C., “Why Digitize” by Abby Smith, 1999, <http://www.clir.org>.
- Columbia University Libraries, “Selection Criteria for Digital Imaging” at: <http://www.columbia.edu/cu/libraries/digital/criteria.html>.
- New York Public Library, “Planning Digital Projects for Historical Collections” at: <http://digital.nypl.org/brochure/planning.htm>.
- New York State Office of the Chief Information Officer, “New York State Information Technology Strategic Plan” at: <http://www.oft.state.ny.us/policy/P04-004/stratplan.htm#toc>.
- New York University, “NINCH Guide to Good Practice” at: <http://www.nyu.edu/its/humanities/ninchguide/III/>.
- North East Document Conservation Center, “Handbook for Digital Projects: A Management Tool for Preservation and Access,” by Diane Vogt-O’Connor at: [www.nedcc.org](http://www.nedcc.org).
- Ohio Memory Project, “How to Participate in the Ohio Memory Project” at: <http://www.ohiomemory.org/om/participate2.html>.

## DESCRIPTIVE METADATA GUIDELINES

This section contains the best practices for descriptive metadata as recommended by WNYLRC for implementation in regional digital projects and programs. Any application of descriptive metadata at the item level for digital collections must be preceded by thorough processing and analysis of the item's respective collection. Every collection has its own environment of information and context. It is only by examining the whole of the collection and its context that strong descriptive metadata can be developed. This metadata must be developed with awareness of how online patrons may wish to access and use the materials.

It is strongly recommended that all institutions participating in digitization projects and programs will develop and use a set of core elements in the description of their materials to facilitate sharing of electronic resources across the region. Flexibility of description has also been addressed to meet the needs of individual collections while maintaining standards at the collaborative network level.

The recommendations were compiled by thorough examination of current best practices used by other cooperative digital collections. The "Recommended Resources and Readings" on page 30 contains the electronic addresses of these sites.

This document has been prepared in recognition of the need to create clear, accurate descriptions of digitized materials in a manner that provides **consistency and flexibility** appropriate to individual collections searchable within a shared network of resources. The goal of combining these two approaches is to provide the user with data that is accurate, predictable in its approach and application, comprehensive, and in language that is natural to the user as possible.

**Consistency** shall be applied to create:

- a standard approach to the creation of descriptive metadata
- a standard set of fields
- a standardized field structure
- a standardized formatting of data entry for information such as dates and names
- a standardized list of controlled vocabularies to be used as descriptors
- descriptive records that allow for data harvesting and migration

**Flexibility** shall be applied to allow for the creation of:

- descriptive records that meet specific institutional needs as well as the standards of the collaborative network by being as rich or minimal as need be
- descriptive records that can handle multiple layers of description

It has been decided that in order to meet these requirements:

- Dublin Core (in its qualified form) shall be used as the metadata backbone of the regional database. All local field names shall be mapped to Dublin Core elements.
- Field names used for specific collections shall be created from the context of the collection content.

Ex. A photograph collection shall use the field name Photographer. This local field name will be mapped to the Dublin Core element Creator.

An excellent example of this approach can be seen by looking at the data dictionaries used for the digital collections at the University of Washington.

<http://www.lib.washington.edu/msd/mig/datadicts/default.html>

The dictionaries are accompanied by several other documents that provide further background.

<http://www.lib.washington.edu/msd/mig/advice/default.html>

Metadata can be used to record information about several aspects of an image or object. That information can be divided in a variety of ways but can basically be said to be describing either the item itself (and/or its digital surrogate), the content of the item, or the collection in which the item is found.

For example:

**Collection level** description might include:

- Collection title
- Collection identifier
- Repository
- Collection creator

**Item level** description might include:

- Item identifier
- Title
- Creator (of original)
- Subject descriptor

- Material type descriptor
- Notes (or description)
- Rights
- File type (digital surrogate)
- File size (digital surrogate)
- File name (digital surrogate)
- Dublin Core Metadata InitiativeType,  
<http://dublincore.org/documents/dcmi-type-vocabulary/>

It is important to note that neither list is exhaustive.

Consistency and flexibility will be gained by adopting a field structure that includes both mandatory and optional fields. The list of optional fields is expandable beyond the examples included below.

<b>Mandatory Fields</b>	<b>Optional Fields (Samples)</b>
Collection Title	Creator
Collection Identifier	Date
Departmental Affiliation	Alternate Titles
Repository	Shelf Location
Item ID	Medium
Title	Dimensions
Subject Term	Notes
Material Type	Description
Rights	Style
DCMI Type	Location Depicted
File Size (system generated)	Holding institution (of original)
File name	Source of original
	Collection creator
	Technique
	Relation

Further consistency will be provided by using controlled vocabularies for appropriate fields. Sample vocabularies, as well as mappings to Dublin Core, are noted in the tables of mandatory and optional fields on pages 32 – 33.

## Project Metadata Workflow

### ***A. Evaluate and establish key metadata fields:***

1. Collection title: Establish a concise title for the entire collection.
2. Collection creator name: Use the same form of the name consistently, preferably an AACR2rev. form of name.
3. Departmental affiliation (may not be necessary for all institutions): Establish a form of name for the department within the holding institution that is responsible for the collection. It is preferable that the name be in AACR2rev. form.
4. Institution: Apply the name of the institution consistently, preferably in AACR2rev. form.
5. Collection ID: Establish an identifier for the online collection (as opposed to a shelving number or collection number for the physical collection). This number will be created by a system administrator in cases of collaborative projects. The ID should be applied consistently to each item record in the collection.
6. File naming convention: If a convention is not already in place, it is recommended best practice to use sequential numbering with a constant number of digits, e.g., for a collection of 900 items use 001, 002, --900, etc.). There are different schemes for file naming. Selection of a scheme will depend upon local practice and whether a collaborative project is involved.
7. Item ID: If a naming convention is not already in place, it is recommended best practice to follow sequential numbering as described above.
8. Rights statement: Establish a stable Web address at which the appropriate rights statement is available, or enter a rights statement in each item record.
9. Display rights: Does the collection require any restrictions and/or requirements for use of watermarking, branding, or banding? *These features must be applied before images are loaded.*

***B. Analyze the contents and nature of the collection to build consistency in description:***

## 1. Creator(s):

- What creative roles can be identified? Is there a single role, such as photographer, or multiple roles, such as architect, contractor, and photographer? Decide how many discrete roles need to be described and the names of the roles for field identification.
- Is there a set of creators that can be identified? For example, can the collection be surveyed to create a list of photographers, artists, etc. for use as a controlled vocabulary in the database? Creating such a list before data entry is begun may speed the process. Also, authority control can be applied more efficiently.

## 2. Material type, work type, medium, physical description:

- Survey the collection in order to discover types represented. If possible, create a list of controlled vocabulary that can be used in the database. Pre-existing lists of terms should be available in most cases.
- Do the types need additional description such as dimensions, medium, color, original-reprint, etc.? If so, create appropriate fields for the database and decide upon language restrictions.
- Will multiple material types be required to describe a single item?
- Is it appropriate to describe the material of the original object: glass, stone, wood, etc.? If so, create a controlled vocabulary.
- Establish which Dublin Core Metadata Initiative Types will be applied and whether they are constant throughout the collection or need to be applied item by item.

## 3. Subject headings:

- Decide upon a controlled vocabulary that is appropriate to the content of the collection – LCSH, MESH, AAT, TGMI, etc.
- Survey the collection in order to discover range of subject headings required. Identify appropriate headings from controlled vocabulary lists and create list for use in database.
- Will multiple subject headings be required to describe a single item? If so, multiple fields will be required in the database.

## 4. Dating, cultural placement:

- Identify need for dating: are specific dates available or are chronological periods and/or styles more appropriate. In some cases, both approaches may be needed.
- Decide on format of specific dates.
- Decide on controlled vocabulary for periods and/or styles.
- Is it necessary to identify nationalities and/or cultures? If so, create controlled vocabularies.

## 5. Sources:

- Is it necessary to include bibliographic citations, vendor information, or name of a holding institution (of the original work, **not** the holding institution for the collection being described)? If so, create a controlled vocabulary.

## 6. Item titles:

- Establish criteria for assigning titles, including use or omission of initial articles. Since this field may control the order of item display online, it is important to consider issues such as clustering, etc. Titles need not be unique, but unique titles may make future identification simpler.
- Decide what secondary, alternate, parallel titles are required.

## 7. Other information:

- What other information needs to be recorded in order to make the database as useful as possible for the end user? In all cases look for opportunities to apply controlled vocabularies.

***C. Represent of the collection in online catalogs:***

1. Create a cataloging record for the online, digital collection. This record may closely resemble the collection-level bibliographic record for the physical collection but should be specific to the electronic collection. In most cases this bibliographic record for the online collection should be passed into both the institution's local online catalog as well as a national database, either RLIN or OCLC. The record must also be passed into WNYLRC's WNYlibraries.org virtual, regional catalog. The bibliographic record will, according to current cataloging standards, contain a link to the online collection so a patron who locates the collection in the online catalog can proceed directly to the online collection.
2. Decide if it is necessary or desirable to pass the metadata records for the items within a collection into OCLC and/or a local online catalog. If so, this is a possibility within the

ContentDM software. Institutions using other software should investigate this possibility within their own software.

***D. Create data storage:***

1. Establish a procedure for storage of digital surrogates of collection contents.
2. Establish a procedure for storage of metadata, preferably in XML (eXtensible Markup Language) format. This may also require establishing a procedure for harvesting the metadata and transforming it to XML.

***E. Perform end testing:***

1. Establish procedures for proofreading metadata and checking other collection standards.
2. Establish procedures for maintaining collections: updating system software, entering corrections, adding new material, etc.
3. Establish procedures for measuring use of the online collections.
4. Establish policies for supplying reference for the online collections and dealing with requests for duplicates of materials from the collections. These policies should ideally be developed prior to mounting any digital collection online.

## Recommended Resources and Readings

### Best Practice Documents:

- Collaborative Digitization Program. Dublin Core Metadata Best Practices. Version 2.1.  
<http://www.cdphheritage.org/cdp/documents/CDPDCMBP.pdf>
- University of Oregon Libraries. Access to Digital Collections: Best Practices  
[http://libweb.uoregon.edu/diglib/AccessstoDigitalCollections\\_Rev.html](http://libweb.uoregon.edu/diglib/AccessstoDigitalCollections_Rev.html)
- RLG. Descriptive Metadata Guidelines for RLG Cultural Materials  
[http://www.rlg.org/en/pdfs/RLG\\_desc\\_metadata.pdf](http://www.rlg.org/en/pdfs/RLG_desc_metadata.pdf)

### Metadata:

- Introduction to Metadata: Pathways to Digital Information, Online Ed. Version 2.1 (Getty)  
[http://www.getty.edu/research/conducting\\_research/standards/intrometadata/index.html](http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html)
- Guidelines for Working the DRL to Create Image Collections (Univ. of Pittsburgh)  
[http://digital.library.pitt.edu/drl/drl\\_image\\_guidelines.pdf](http://digital.library.pitt.edu/drl/drl_image_guidelines.pdf)
- A Framework of Guidance for Building Good Digital Collections (IMLS)  
<http://www.niso.org/framework/Framework2.html>
- Understanding Metadata (NISO)  
<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>
- *Cataloging & Classification Quarterly*, v. 40, Issue 4/5, Sept. 2005.  
Entire issue is devoted to metadata issues, including METS and EAD.
- University of Washington Metadata Implementation Group.  
<http://www.lib.washington.edu/msd/mig/>  
The site includes information about how to implement metadata, data dictionaries used in University of Washington projects along with crosswalks to Dublin Core, and links to controlled vocabularies  
(<http://www.lib.washington.edu/msd/mig/datadicts/default.html#refs>)

### Describing image collections:

- Zinkham, Helena. Subject Indexing for Pictures: An Overview  
<http://www.loc.gov/rr/print/tp/SubjectIndexingforPictures.pdf>
- Zinkham, Helena. Common and Useful Information Elements for Cataloging Pictorial Materials  
<http://www.loc.gov/rr/print/tp/CommonInformationElements.pdf>
- Library of Congress. Introduction to TGM II (Thesaurus of Graphic Materials).  
<http://www.loc.gov/rr/print/tgm2/>  
Contains the downloadable contents of the Thesaurus and information about how to apply the thesaurus terms in cataloging.

## Glossary

- AAT      Art and Architecture Thesaurus  
The AAT is a structured vocabulary of more than 133,000 terms, descriptions, bibliographic citations, and other information relating to fine art, architecture, decorative arts, archival materials, and material culture.  
[http://www.getty.edu/research/conducting\\_research/vocabularies/aat/](http://www.getty.edu/research/conducting_research/vocabularies/aat/)
- LCSH      Library of Congress Subject Headings
- LC NAF      Library of Congress Name Authority File
- NLM      National Library of Medicine maintains MESH (Medical Subject Headings)  
<http://www.nlm.nih.gov/mesh/introduction2006.html>
- TGMII      Thesaurus of Graphic Materials II: Genre and Physical Characteristic Terms  
A thesaurus of more than 600 terms developed by the Library of Congress Prints and Photographs Division. <http://www.loc.gov/rr/print/tgm2/>

## Core (Mandatory) Fields

Local Field Names (Recommended)	Dublin Core (see <a href="http://dublincore.org/documents/dcmi-terms/">http://dublincore.org/documents/dcmi-terms/</a> )	Required	Repeatable	Authority Control	Notes
Collection ID	Identifier	Yes	No	Local	Unique assigned alpha-numeric code
Collection title	Title	Yes	No	Local	35 character max.
Departmental Affiliation	Publisher	Yes	Yes	Local	Name of the institutional department in which item is located
Institution	Publisher	Yes	No	Local	Holding institution
Item ID	Identifier	Yes	No	Local	Number used to identify the original item within its collection. Schema will be assigned by Application Administrator if none exists.
Title	Title	Yes		None	Title of item. Must be present in every CDM record
Material Type	Type	Yes	Yes	TGMII	
Subject heading(s)	Subject	Yes	Yes	LCSH, NLM, etc.	Content or subject matter of the object. Can contain multiple entries in single field. <b>Minimum of one.</b>
File Name	Format	Yes	No	None	name of the digital file with file extension
File Size	Format	Yes	No	None	May be system-generated.
Rights Management	Rights	Yes	No	N/A	A statement of who owns the copyright for this item and what limits apply to its use
DCMI Type	Type	Yes	No	<a href="http://dublincore.org/documents/dcmi-type-vocabulary/">http://dublincore.org/documents/dcmi-type-vocabulary/</a>	The DCMI Type Vocabulary provides a general, cross-domain list of approved terms that may be used as values for the Resource Type element to identify the genre of a resource.

<b>Optional Fields</b>					
<b>Local Field Names (Recommended)</b>	<b>Dublin Core (see <a href="http://dublincore.org/documents/dc-mi-terms/">http://dublincore.org/documents/dc-mi-terms/</a>)</b>	<b>Required</b>	<b>Repeatable</b>	<b>Authority Control</b>	<b>Notes</b>
Date/Period	Date	No	Yes	AAT for Western Objects	Identify original creation date of the work.
Contributor	Contributor	No	Yes	Drop-down, LC NAF	Name of the person or department who supplied and is responsible for the collection
Shelf Location	Identifier	No	No	Local	Physical location of item within its collection
Alternate titles	Title	No	Yes	None	Alternate titles of similar types can be repeated in a single field. Alternate titles of other, specific types (parallel, uniform, etc.) should be entered in separate fields.
Name	Creator	No	Yes	LC NAF	Name of responsible agent for creation of original work. Roles will be negotiated with App. Admin.
Medium	Format	No	Yes	AAT	Used to identify the physical make-up of original object
Dimensions	Format	No		None	measurements of source object
Notes	Description	No		AAT, LC Descriptive Terms for Graphic Materials	Additional physical description. Similar to a MARC notes field.
Description	Description	No	Yes	N/A	Natural language, narrative description of source document
Nationality/Culture	Coverage	No	Yes	AAT, LCSH	The name of the culture from which a work originates or the name of the nationality with which the work has been associated.
Style	Coverage	No	Yes	AAT for Western Objects	Term used to identify the style of the work
Location Depicted	Coverage	No	Yes	None	If authority is used, it should be put in a Subject Heading instead of Location Depicted field.
Holding Institution	Publisher	No		Local	Name of institution holding the source documents/underlying work
Source of Original	Source	No		LC NAF, bibliographic citation	Name (institutional or personal) source of the surrogate file
Collector	Creator	No		LC NAF	Name of the personal or institutional collector who owns or supplied the original source documents
Technique	Format	No			
Relation	Relation	No			A reference to a related resource.

## DIGITAL IMAGING BEST PRACTICES

### Introduction

The purpose of this document is to provide basic digital imaging recommendations to WNYLRC members that are planning for or are involved in digitization projects. This document does not attempt to answer every question related to digitization, and also acknowledges that available technology continues to change and improve. Therefore, this is a guide for image capture, presentation, storage, and preservation. For additional information, readers are urged to consult some or all of the resources listed at the end of this section.

The recommendations in this document are purposely broad enough to apply to the variety of WNYLRC institutions and collections and attempt to synthesize different recommendations previously made for specific institutions or projects around the country. Three specific resources were used extensively in the preparation of this section:

Cornell University Library. Research Department. (2000-2003). *Moving theory into practice: Digital imaging tutorial*. Ithaca, NY: Cornell University Library. Retrieved February 17, 2006, from <http://www.library.cornell.edu/preservation/tutorial/contents.html>

Sitts, M. K. (Ed.). (2000). *Handbook for digital projects: A Management tool for preservation and access*. Andover, MA: Northeast Document Conservation Center. Retrieved February 17, 2006, from <http://www.nedcc.org/digital/dighome.htm>

Western States Digital Standards Group, Digital Imaging Working Group. (2003, January). *Western states digital imaging best practices*. Denver, CO: Western States Digital Standards Group. Retrieved February 17, 2006, from [http://www.cdpheritage.org/digital/scanning/documents/WSDIBP\\_v1.pdf](http://www.cdpheritage.org/digital/scanning/documents/WSDIBP_v1.pdf)

### Scope

This document addresses:

- Scanning and file format recommendations for graphic materials including but not limited to printed text documents, manuscripts, photographs, maps, and other physical material
- Suggested workspace and equipment (hardware and software) for digital imaging
- Quality control of digital images
- File management of digital images
- Storage and preservation of digital images

Because it is anticipated that regional institutions will concentrate primarily on digitizing images from rare books, photographs, and documents during the first couple of years of this plan, this document does not address digitization of microform, audio, video, moving image, 3-D objects or “born-digital” items. Resources for these projects will be added to WNYLRC’s digitization Web site in the future.

Please note that the information in this document is meant to provide libraries with starting-point suggestions. Institutions that already have guidelines or mandates should continue their practices. Libraries that are affiliated with governmental agencies should follow New York State Archives standards when digitizing governmental records or business records of the library.

### **General Digital Imaging Principles**

The Western States Digital Standards Group, Digital Imaging Working Group (2003) provided succinct guidelines for digitization in their document on page six. Their recommendations are shown below, with some modification for the WNYLRC region:

1. Scan at the highest resolution appropriate to the nature of the source material.
2. Scan at an appropriate level of quality to avoid rescanning and re-handling of the originals in the future--**scan once**.
3. Create and store a master image file that can be used to produce derivative image files, such as files optimized for the Web, and serve a variety of current and future user needs.
4. Use equipment, software, and file formats that are standard and supported, rather than proprietary and specific to certain software. Proprietary formats often involve licenses and special plug-ins for viewing. Examples of proprietary formats include Photoshop (.psd) or Kodak PhotoCD (.pcd).
5. Use image file formats and compression techniques that conform to standards within the WNYLRC community.
6. Create backup copies of all files on a stable medium.
7. Create meaningful metadata for image files or collections.
8. Store media in an appropriate environment.
9. Monitor and recopy data as necessary; pay special attention to master image files and backups so that the media storing them does not become unreadable or obsolete.
10. Document a migration strategy for transferring data across generations of technology.
11. Anticipate and plan for future technological developments.
12. Scan an original or first generation (i.e., negative rather than print) of the source material to achieve the best quality image possible.

### **Scanning & Image Creation**

These guidelines provide the minimum qualities necessary for achieving an acceptable level of image quality.

Collections differ in the ways they are used and accessed and institutions have differing purposes and clientele, which will likely have an impact on how and for what purposes and reasons collections are digitized. The recommendations in this document are not hard and fast standards for every collection and every institution. As a rule, the key to quality scanning is not to scan at the highest resolution possible but to scan at a level that matches the informational content of the original. Scanning a document in marginal condition at a higher resolution actually can magnify imperfections.

Decisions on image quality and resolution should be based on the needs of users, how the images will be used, and the nature of the materials you are scanning (dimensions, color, tonal range, format, material type, etc.). The quality and condition of the original (such as the quality of the

shooting or processing technique in the case of photographs) impacts on the resolution at which you scan and the resulting quality of the digital image.

### ***Digital Master Image File***

According to Cornell University's *Moving Theory into Practice* tutorial (2005, p. 32-33), there are access, economic, and preservation reasons for creating a rich digital master image file, or archival image, in which all significant information contained in the source document is represented. As they point out, "the master image should be the highest quality you can afford; it should not be edited or processed for any specific output; and it should be uncompressed. Intensive quality control should be applied in creating master image files." This is the file that you retain for the future to use for production of additional digital or print format files. Archiving a digital master image file eliminates the need to re-scan items.

***Access:*** A digital master should be capable of supporting a range of users' needs through the creation of derivatives for printing, display, and image processing. The richer the digital master, the better the derivatives in terms of quality and "processibility." User expectations will likely be more demanding over time—the digital master should be rich enough to accommodate future applications. Rich masters will support the development of cultural heritage resources that are comparable and interoperable across disciplines, users, and institutions.

***Economic:*** Creating a high quality digital image may cost more initially, but will be less expensive than creating a lower quality image that fails to meet long-term requirements and results in the need to re-scan. Labor costs associated with identifying, preparing, inspecting, indexing, and managing digital information far exceed the costs of the scan itself.

The key to image quality is not to capture at the highest resolution or bit depth possible, but to match the conversion process to the informational content of the original, and to scan at that level--no more, no less. In doing so, one creates a master file that can be used over time. Long-term value should be defined by the intellectual content and utility of the image file, not limited by technical decisions made at the point of conversion.

***Preservation:*** Creating a rich digital master can contribute to preservation in at least three ways:

1. *Protecting vulnerable originals.* The image surrogate must be rich enough to reduce or eliminate the user's need to view the original.
2. *Replacing originals.* Under certain circumstances, digital images can be created to replace originals or used to produce paper copies or Computer Output Microfilm. The digital replacement must satisfy all research, legal, and fiscal requirements.
3. *Preserving digital files.* It is easier to preserve digital files when they are captured consistently and well documented. The expense of doing so is more justifiable if the files offer continuing value and functionality.

### ***Derivative Image Files***

Derivative files are created from the master digital image, and are used in place of it, usually for general Internet or network access. Derivative files typically include an **access image**, which is sized to fit within the screen of an average monitor or other delivery mechanism and a **thumbnail image**, which is small enough to load quickly and linked to the larger access image.

With the proper image editing software it is not necessary to subject source materials to multiple scans as derivative files can be created from the high quality master images.

WNYLRC recommends that three versions of an image be created, as described in the following table, according to information presented by the Western States Digital Standards Group (2003, p. 26).

<b>Master Image</b>	<b>Access Image</b>	<b>Thumbnail Image</b>
<ul style="list-style-type: none"> <li>• Represents as closely as possible the information contained in the original</li> <li>• Uncompressed</li> <li>• Unedited</li> <li>• Serves as long term source for derivative files</li> <li>• Can serve as surrogate for the original</li> <li>• High quality</li> <li>• Very large file size</li> <li>• Used for creating high quality print reproductions</li> <li>• Usually stored in the TIFF file format</li> </ul>	<ul style="list-style-type: none"> <li>• Used in place of master image for general web access</li> <li>• Generally fits within viewing area of average monitor</li> <li>• Reasonable file size for fast download time; does not require a fast network connection</li> <li>• Acceptable quality for general research Compressed for speed of access</li> <li>• Usually stored in JPEG file format</li> </ul>	<ul style="list-style-type: none"> <li>• A very small image usually presented with the bibliographic record</li> <li>• Designed to display quickly online; allows user to determine whether they want to view access image</li> <li>• Usually stored in GIF or JPEG file formats</li> <li>• Not always suitable for images consisting primarily of text, musical scores, etc.; user cannot tell what content is at so small a scale</li> </ul>

GIF, JPEG and TIFF are standard file formats for storing images. GIF, which stands for “Graphics Interchange Format,” was the original graphics format used on the Web. GIF format can display only up to 256 colors with lower resolution, and therefore is not suitable for most digital images other than thumbnails. JPG, or JPEG, stands for “Joint Photographic Experts Group. The JPG format is best suited for Web display of photographs and complex illustrations. TIFF, which stands for “Tagged Image File Format,” is a standard file format for scanning and storage. TIFF files are excellent source files that retain the scanned information for later use; these files can be opened and saved multiple times without losing quality. TIFF images are generally too large to use on a Web site and must be converted into GIF or JPG files for use on the Web.

The Grainger Engineering Library at the University of Illinois at Urbana-Champaign has produced a useful *Image Quality Calculator*, which is available on the Internet at <http://images.library.uiuc.edu/projects/calculator>. This interactive tool allows the user to input the dimensions of the object to be scanned, answer questions regarding the desired quality, and then obtain information on the recommended scan resolution and resulting file size.

The Western States Digital Standards Group (2003, pp. 31-33) offers additional image quality guidelines in the following tables:

	<b>Master</b>	<b>Access</b>	<b>Thumbnail</b>
<b>File Format</b>	TIFF	JPEG	JPEG or GIF
<b>Bit Depth</b>	1 bit bitonal 8 bit grayscale 24 bit color	1 bit bitonal 8 bit grayscale 24 bit color	1 bit bitonal 8 bit grayscale 8 bit indexed color (GIF) 24 bit color
<b>Spatial Resolution</b>	600 ppi	150 dpi	72 dpi
<b>Spatial Dimensions</b>	100% of original	600 pixels across the long dimension	150-200 pixels across the long dimension

Table 1. Text

	<b>Master</b>	<b>Access</b>	<b>Thumbnail</b>
<b>File Format</b>	TIFF	JPEG	JPEG or GIF
<b>Bit Depth</b>	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color	8 bit grayscale 8 bit indexed color (GIF) 24 bit color
<b>Spatial Resolution</b>	3000 to 5000 pixels across the long dimension	150 dpi	72 dpi
<b>Spatial Dimensions</b>	100% of original	600 pixels across the long dimension	150-200 pixels across the long dimension

Table 2. Photographs

	<b>Master</b>	<b>Access</b>	<b>Thumbnail</b>
<b>File Format</b>	TIFF	JPEG	JPEG or GIF
<b>Bit Depth</b>	8 bit grayscale 24 bit color	8 bit grayscale 24 bit color	8 bit grayscale 8 bit indexed color (GIF) 24 bit color
<b>Spatial Resolution</b>	3000 pixels across the long dimension	150 dpi	72 dpi
<b>Spatial Dimensions</b>	100% of original	600 pixels across the long dimension	150-200 pixels across the long dimension

Table 3. Maps and Graphic Materials

Portable Document Format, or PDF files, are also commonly used to deliver digitized content electronically. PDF is often a good choice for text documents and forms that will be put on a Web site. PDF format is truly portable across different platforms, and works equally as well on PC and Apple computers. The format retains all of the formatting and detail present in the original, and can be compressed into smaller file sizes for the Web. The industry standard software package for creating PDF files is Adobe Acrobat Standard or Professional; please note that the free Adobe Acrobat Reader cannot create PDF files, but only provides a mechanism for reading them. Some higher-end scanners also come with software that will create PDF output of scanned documents.

### ***File Sizes***

Cornell University's *Moving Theory into Practice* tutorial (2005, p. 14) provides useful information for estimating file sizes of digitized items. According to their tutorial:

File size is calculated by multiplying the surface area of a document (height x width) to be scanned by the bit depth and the dpi<sup>2</sup>. Because image file size is represented in bytes, which are made up of 8 bits, divide this figure by 8.

<p><i>Formula 1 for File Size</i>  <math display="block">\text{File Size} = (\text{height} \times \text{width} \times \text{bit depth} \times \text{dpi}^2) / 8</math></p>
--

If the pixel dimensions are given, multiply them by each other and the bit depth to determine the number of bits in an image file. For instance, if a 24-bit image is captured with a digital camera with pixel dimensions of 2,048 x 3,072, then the file size equals  $(2048 \times 3072 \times 24) / 8$ , or 18,187,368 bytes.

<p><i>Formula 2 for File Size</i>  <math display="block">\text{File Size} = (\text{pixel dimensions} \times \text{bit depth}) / 8</math></p>
--

Because digital images often result in very large files, the number of bytes is usually represented in increments of 2<sup>10</sup> (1,024) or more:

- 1 Kilobyte (KB) = 1,024 bytes
- 1 Megabyte (MB) = 1,024 KB
- 1 Gigabyte (GB) = 1,024 MB
- 1 Terabyte (TB) = 1,024 GB

The Northeast Document Conservation Center, in its *Handbook for Digital Projects* (2006, p. 96) identifies the following three important measures of every static digital image:

1. Resolution. The number of dots, or pixels (picture elements), used to represent an image. This is always given as a measure of linear or area density (e.g., 300 dots/inch).
2. Pixel Bit Depth. This measure defines the number of shades that can actually be represented by the amount of information saved for each pixel. These can range from 1 bit/pixel for binary (fax type) images to 24 bits per pixel in high quality color images.
3. Color. There are many ways to represent, compress, and distribute color images. Suffice it to say that the smaller the image file size, the less accurately it renders the original image.

### **Quality Control**

Quality control is a necessary part of any project, and is especially important in digital imaging projects, where errors can produce lower quality visual information, as well as the possibility of having to handle material, which could be brittle, again for a re-scan.

The following points to consider when developing a quality control program were adapted from Cornell University's *Moving theory into Practice* tutorial (2003, p. 45-48).

### 1. Identify Your Products

The first step is to clearly identify the products to be evaluated. These might include master and derivative images, printouts, image databases, and accompanying metadata, including converted text and marked-up files.

### 2. Develop a Consistent Approach

To measure quality and judge whether the products are satisfactory, clearly define baseline characteristics for "acceptable" and "unacceptable" digital products.

### 3. Determine a Reference Point

What are you judging the images against? Answering this question is not always straightforward. For example, if conversion is based on an intermediate file, the digital image is two "generations" away from the original. It has been copied to film (first generation), which is then scanned (second generation). What should be the reference point in assessing such an image, the original document or the transparency? Will the master or derivative (or both) be the focus of image quality inspection?

### 4. Define the Scope and Methods

Determine the scope of your quality review. Will you inspect all the images, or only on a sampled subset (e.g., 20%)?

Describe your methodology and define how quality judgments will be made. For example, will you visually evaluate the images at 100% (1:1) magnification onscreen and compare them to the original documents? Or, will your evaluation be based only on a subjective assessment of images onscreen, without reference to the originals?

### 5. Control the Quality Control Environment

The impact of image-display conditions on perceived quality is often underestimated. Given an improper environment, even a high-quality image may come across as unsatisfactory. For example, a 24-bit color image might look heavily "posterized" when viewed using an improperly configured computer that cannot provide a full palette of colors.

- *Hardware Configuration*

It is difficult to prescribe the ideal hardware configuration. The rule of thumb is to assemble a system that supports your requirements for speed, memory, storage, and display quality. What kinds of images are being created? How many? For what purposes? What level of on-screen review is needed? You will need a fast and reliable computer with ample processing power and memory to be able to retrieve and manipulate the large files you are creating, especially when creating color images.

- *Image Retrieval Software*

Use retrieval software appropriate to your images. For example, if you are evaluating images created and stored in Kodak ImagePac format, retrieve them using one of the viewing freeware and shareware products available on the Web that support the format and color space.

- *Viewing Conditions*  
Control your viewing environment. Understand that the monitor and the source document require distinct viewing conditions. The original is best viewed in a bright surrounding, and the monitor works best in a low-light environment. However, a low-light environment does not equate to a dark room. Viewed in the dark, an on-screen image would appear deficient in contrast. Remember that these suggestions are for quality control viewing conditions and not for general viewing when the public version is available on the Web.

## 6. Evaluate System Performance

Whether conversion takes place in-house or is outsourced, system performance should be evaluated to ensure consistency throughout the conversion process. Among characteristics to evaluate are resolution, linearity, flare, scanner noise, color reproduction, and various artifacts.

## 7. Codify Your Inspection Procedures

Quality control data has long-term value, from supporting different stages of quality inspection to facilitating future manipulation and migration. For in-house components of quality control, we recommend detailing the inspection procedures in a short manual (or in a series of workforms) to be used in training and to facilitate workflow. Issues that need to be addressed include: quality control procedures; staff involved and skills needed; instruments, hardware, and software needs; and rejecting and replacing unacceptable products.

## Equipment

### *Workspace*

The Western States Digital Standards Group (2003, p. 17) gives the following tips for ensuring an efficient workspace for digital imaging projects:

Providing a comfortable, safe, and secure workspace for a digitization project can increase productivity and quality of images by reducing operator fatigue and potential damage to collections. Proper climate control and security are important if collections will stay in the lab for extended periods of time. A workspace for digitization should offer a controlled lighting source to maintain consistency and quality of images. Changes in room lighting can affect how images are represented on computer monitors and may introduce challenges to accurate calibration.

### *Hardware*

#### 1. Computers

Recommendations regarding hardware become out-of-date so quickly that it is practical to give only general guidelines in a planning document of this type. Trade publications, such as *PC Magazine* (<http://www.pcmagazine.com>), *PC World* (<http://www.pcworld.com>) or *Computer Shopper* (<http://www.computershopper.com>), are useful sources for up-to-date hardware information.

It is possible to use lower-end computers for some of the digitization work if you will be using multiple workstations. For example, a lower-end, or older, machine could be used to

run the scanner and acquire the images. The images could then be transferred to a higher-end computer for final processing.

If you will be digitizing an extensive collection, it is best to purchase computer equipment that will be dedicated to that task. When purchasing a dedicated computer, keep these guidelines in mind:

- Choose the most advanced processor available. For example, a person choosing a new Windows-based computer in 2006 would choose a Pentium processor over a Celeron processor.
- Purchase as much memory (RAM, or Random Access Memory) as you can afford; additional increments of memory are often quite affordable. Ample memory allows a computer to work faster and process image data more quickly without having to cache, or write back and forth to the computer's hard drive.
- Ensure that the computer includes a high-quality graphics card. The card should be capable of displaying high resolutions. Lower-end graphics cards often cause slow processing or uneven display.
- A large hard drive would obviously be useful for storage. However, additional external storage can also be added, such one or more external high-capacity hard drives.
- Check what connections are available on the computer and ensure that they are the fastest technology available. For example, in 2006, you would look for a computer with a USB 2.0 and/or a Firewire connection. These components are normally clearly listed in the detailed specs for a computer, or you can ask the sales representative.
- Computers that are advertised as "media-ready" or "media center ready" often have the necessary components. However, it is still necessary to double-check the amount of memory included with the machine.

## 2. Displays

Information regarding displays is also best described in general terms in this planning document, since advances are continually being made. It is important to note that high quality monitors necessary to see and manipulate details in digitized material cost more than standard monitors that are suited more for standard word processing and Internet work. If the materials that you will be digitizing are standard text materials, you may be fine with a less expensive monitor. However, if you will be working with more complex materials, including colors and images, you will want a higher-end monitor.

- Purchase a monitor that is at least 19-21 inches and capable of displaying high resolutions (at least 1280-x1024).
- Make sure that the monitor is capable of calibration to insure that the colors displayed remain true. Monitors that are used for image processing should be calibrated every couple of months, as monitor displays can change with age and use. Calibration software often comes with image processing software, and Apple computers generally have a built-in display calibration package.

### 3. Types of Scanners

Scanners have different features and speeds. When considering the speed of a scanner, check to see how many pages or images a minute it can scan; this is known as the rated capacity or throughput. However, in actual use, the number of pages scanned per minute will be lower and is often approximately 50 percent of the rated capacity, due to the time needed to place and adjust documents on the scanner bed, software manipulation, and worker rest breaks.

Information in this section was adapted from Western States Digital Standards Group (2003, p. 14).

- ***Flatbed Scanners***

Flatbed scanners are one of the most popular types scanners used in libraries and archives and are suitable for scanning papers, flat photographs, and other printed materials. An important consideration when selecting a flatbed scanner is the size of the scan area. Most consumer models are limited to a scan area of 8.5" x 11" but professional grade models are available with a larger scan area. Some models of flatbed scanners are available with accessories such as transparency adapters (for slides and negatives) and automatic document feeders. The quality of scans from transparency adapters varies greatly and projects with large numbers of slides or negatives should consider a dedicated slide and film scanner. Automatic feeders are not recommended for original materials because of the danger of damage but can increase efficiency of scanning contemporary documents such as archival finding aids.

- ***Slide Scanners***

Slide and film scanners are specifically designed to digitize transparent materials such as slides and 35mm film. Transparency scanners generally produce higher quality scans over flatbed transparency adapters due to higher dynamic tonal ranges and optical resolutions. Projects with a large number of slides should consider the advantages of slide scanners and accessories such as automatic slide feeders.

- ***Drum Scanners***

Drum scanners are most frequently used by pre-press and graphic design professionals working with contemporary materials. Because materials are affixed to a drum rotating at high speed around a sensor, drum scanners are not recommended for cultural heritage materials, particularly materials that are fragile or brittle. Drum scanners do produce high resolution scans with high color fidelity and dynamic ranges and are suitable for scanning surrogate negatives and transparencies. The cost of drum scanners will be a limiting factor for most projects.

- ***Wide-format Scanners***

Wide-format scanners were developed to digitize large format materials such as engineering drawings and architectural blueprints and are frequently found in municipal engineering departments or local blueprint shops. Materials are drawn over the scanning sensor through a pair of drums. Due to the danger of mechanical damage (ripping, tearing) these types of scanners are not recommended for cultural heritage materials.

#### 4. Digital cameras

While scanners are generally the equipment of choice for digitization projects, there are instances when a digital camera is a better choice. Some items, such as maps, posters, 3-dimensional objects and artwork, may be too large even for a large format scanner bed, or color scanning may be desired, and some large-bed scanners can scan only in black and white or grayscale. However, many consumer and low-end professional digital cameras do not have sufficient resolution for archival capture of these materials for digital imaging. In addition to resolution difficulties, the lens in these cameras may introduce distortions to flat materials.

According to the Western States Digital Standards Group (2003, p. 15), “for the creation of high resolution, high quality digital images using a camera many cultural heritage institutions employ a “digital scan back” camera. A digital scan back consists of a scanning array that attaches in place of a film holder to a 4” x 5” view camera body. Digital scan back cameras are ideal for original items that cannot be put onto a more traditional scanner—three dimensional objects, photos or artwork bound in albums, large photographs, artwork or maps. Projects considering digital scan-backs should consult with manufacturers to select the correct lenses and lighting needed for the types of material being scanned. At present, the cost of digital scan backs are beyond modest sized projects and best practice has been to scan negatives of oversize materials created using traditional photography.”

### *Software*

#### 1. Scanner Software

In order to properly operate a scanner and take advantage of all of its features, it is necessary to install scanning software, or drivers, on the computer that will be used with the scanner. Most scanners include software that allows the operator to manually adjust resolution and other scan settings, but consumer model scanners often do not include the software necessary for additional adjustments, which can result in lower quality images.

Scanner software should also be able to output image files in different file formats. Lower end scanner software frequently limits file format choices and may not produce files adequate for high quality master images.

#### 2. Image Editing Software

Some scanners also include image manipulation software, which can be of varying quality and effectiveness. It is usually best for institutions that plan on digitizing large numbers of items to plan on purchasing professional image editing software. An example of such software is Adobe’s Photoshop or the lower end, but still quite capable Photoshop Elements. These programs can assist in fixing minor quality problems, as well as convert image files and produce thumbnails for a Web site. In fact, Photoshop and Photoshop Elements both include a Web photo gallery feature, which automates the production of thumbnails and navigable index pages for all image files in a given directory or folder.

According to the Western States Digital Standards Group (2003, p. 16), the following features should be considered when selecting image editing software:

- Ability to work directly with scanner software through TWAIN or other plug-ins
- Support for common non-proprietary file formats. TWAIN refers to the connection between the scanner and the software, and actually stands for “Technology Without Any Interesting Name.”
- Tools for controllable image optimization (color adjustment, tonal adjustments, color spaces)
- Features for the optimization of images for web delivery and automatic creation of HTML templates
- Ability to convert color spaces (RGB (Red-Green-Blue color model of monitors) to CMYK (Cyan-Magenta-Yellow-Black color model) for print output)
- Usable documentation and reliable technical support
- Ability to extend functionality through custom plug-ins
- Ability to create action sets or macros for frequently applied functions.
- Ability to process images in automatic batches
- Other project specific needs and goals

In addition, the Western States Digital Standards Group (2003, p. 17) advises that projects should consider the costs of implementing the software beyond its initial costs. For example, “Does your computer hardware exceed the minimum requirements of the software? Do you have the staff with skills to use the software or the funds to provide training? Can you afford future upgrades to the software? Does the software feature automated processes that can increase efficiency and reduce staffing costs?”

## **File Management**

### ***File Naming***

Systematic file naming is important for identifying the file later, as well as for system compatibility, interoperability, and to demonstrate ownership of the digital asset. Although modern personal computers are no longer restricted to file names consisting of eight characters or less, it is still a good idea to pay careful attention to file names. Because the same file name will often be used if the image is placed on the Web, it is a good rule of thumb to ensure that file names contain no spaces, because some Web servers have difficulty with filenames that contain spaces, or they automatically rename them by putting characters such as “%20” into the file name in place of the space. The underscore (\_) character can be used in file names to take the place of a space. File names should utilize a three-character extension to identify the type of file (e.g. filename.jpg). In addition, avoid using any characters reserved for system use (i.e. \ / ? \* |, etc.).

## **Storage & Preservation**

Digitization and the creation of digital collections is labor intensive, sometimes costly, and therefore merits careful planning for the storage and long-term preservation. The Western States Digital Standards Group (2003, pp. 18-21) has identified the following storage options.

## ***Storage Options***

### **1. Optical Media Storage**

Optical media include CD-ROM (Compact Disc – Read Only Memory), CD-R (Compact Disc – Recordable), and DVD-ROM (Digital Versatile Disc – Read Only Memory). All employ a laser to read data from a metallic coating over the disc, with a clear acrylic coating covering the metallic layer for protection. Optical discs are a popular medium for storing digital masters, transporting images, and as a publication medium.

Issues of cost, convenience, speed of retrieval and security factor into decisions regarding optical discs as an image storage and retrieval medium. It is recommended that the digital assets be stored on CD-ROMs that conform to ISO 9660, the 1988 standard for volume and file structure of CD-ROMs for information interchange. Audio CD, and DVD Video and Audio are not recommended at this time as a reliable storage medium.

Some institutions may wish to use optical media for storage media due to cost considerations. When acquired in bulk, 650 MB CD-R discs can cost as little as 35 cents each, or .0005 cents per megabyte of storage. For example, an uncompressed 32 MB TIFF file (800 x 600 pixel black and white textual documents scanned at 600 dpi, 8 bit grayscale) will therefore cost approximately 1.6 cents for storage. Thus, approximately 20 32 MB TIFF files can be accommodated on the disc.

Optical discs can provide cost-efficient storage, but they incur staffing costs when accessing and managing the digital assets stored on them. The main drawbacks are slow recording and reading speeds, for even with a fast CD burner, creating CDs is a slow process, and a CD player needs to be available for access. While the CD burner and players are widely available, having to locate the disc on which the image is stored, loading the disc, and then locating the image adds a layer to the data access process.

Digital collections stored on CD ROM work well for small collections. Using this media as a long-term solution however presents major challenges. Storage and retrieval costs escalate as the CD ROM collections become larger and more challenging to manage.

CD-ROMs have a limited physical life span and the images stored on them are vulnerable due to physical deterioration, mishandling, improper storage and obsolescence. Proper conditions can prolong the life of the discs, and optimal environmental levels are 72° Fahrenheit and a Relative Humidity between 20 and 50 percent.

Both adhesive labels and permanent ink markers can cause early failure of CD-ROMs through chemical interaction with the CD's foil. Best practice is to not write or label CDs directly on the body of the CD. Some projects have placed small identification numbers in the central plastic "hub" of the CD.

Projects considering CD ROM for storage of master images should pay attention to the burning speed of the media they purchase and the maximum speed of the CD burner. Like the hardware used to create them, CD-Rs have a maximum burn speed that should not be exceeded even if the hardware is capable of higher speeds

## 2. Online Storage

Storing scanned images on “live” servers is an option for those wishing to archive high resolution TIFF images on a stable platform that offers sustainability and easy storage and retrieval. To prevent the loss of data projects need to properly configure hardware and software, develop responsible backup and disaster recovery policies and procedures, and create realistic plans to deal with technological obsolescence.

## 3. RAID Arrays

A Redundant Array of Inexpensive (or Independent) Disks (RAID) is a collection of disk drives that, if configured properly, can act as a single storage system. These configurations are designed to enable a system to operate when an individual drive fails and prevent the loss of data. Currently, there are about 10 types of RAID configurations. Each configuration has its own unique strong and weak points. Some configurations are best suited for rebuild speed while others have are designed to maximize disk capacity and others are well suited for fault tolerance.

## 4. Network Attached Storage (NAS)

Network attached storage devices are servers that are optimized for file sharing rather than running applications. A NAS can offer flexible storage solution to institutions that already have a local area network and server for user authentication. As storage needs increase additional NAS devices can be added to your network for access and storage of digital images. NAS devices used for digitization initiatives should use a RAID configuration to protect against data corruption or loss.

## 5. Outsourced Storage

Projects that have created images for local and web access may wish to place their large, high- resolution masters into a “dark archive” provided by an outside vendor. Several vendors are developing service plans for the preservation and archiving of digital resources.

## Preservation

According to Cornell University’s *Moving Theory into Practice* tutorial (2005, p. 109), issues to be addressed in digital preservation include:

1. Retaining the physical reliability of the image files, accompanying metadata, scripts, and programs (e.g., make sure that the storage medium is reliable with back-ups, maintain the necessary hardware and software infrastructure to store and provide access to the collection)
2. Ensuring continued usability of the digital image collection (e.g., maintain an up-to-date user interface, enable users to retrieve and manipulate information to meet their information needs)
3. Maintaining collection security (e.g., implement strategies to control unauthorized alteration to the collection, develop and maintain a rights management program for fee-based services)

## Recommended Resources and Readings

*D-Lib Magazine*. Retrieved February 17, 2006, from <http://www.dlib.org>

A quality, free, online magazine covering digital library research and development, including new technologies.

*Collaborative Digitization Program*. (2006). Retrieved February 17, 2006, from

<http://www.cdpheritage.org>

In addition to best practices, this site also provides access to collections and lesson plans.

*Copyright issues for libraries when digitizing materials for the Web*. (2005-2006). Retrieved February 17, 2006,

from Michigan Library Consortium Web site: [http://www.mlc.lib.mi.us/cms/sitem.cfm/library\\_tools/copyright/copyrightdigitize/](http://www.mlc.lib.mi.us/cms/sitem.cfm/library_tools/copyright/copyrightdigitize/)

A helpful listing to assist library staff in determining copyright issues when digitizing. Includes questions to ask about published and unpublished materials.

Cornell University Library. (2003). *Digital preservation management: Implementing short-term strategies for long-term problems*. Ithaca, NY: Author. Retrieved February 17, 2006, from

<http://www.library.cornell.edu/iris/tutorial/dpm/contents.html>

This work looks at digitization for preservation, and includes information on avoiding obsolescence, digital repositories, and advice.

Cornell University Library. Research Department. (2000-2003). *Moving theory into practice: Digital imaging tutorial*. Ithaca, NY: Cornell University Library. Retrieved February 17, 2006, from

<http://www.library.cornell.edu/preservation/tutorial/contents.html>

As its title suggests, this work concentrates on digital imaging, and covers selection of materials to be digitized, quality control, metadata, digital preservation, and management.

DIGLIB electronic mailing list, archived at <http://www.ifla.org/II/lists/diglib.htm>

A “discussion list for digital libraries researchers and librarians”

Hurst Associates, Ltd. (2006). *Digitization 101*. Retrieved February 17, 2006, from

<http://hurstassociates.blogspot.com>

An active blog covering “issues, topics, and lessons learned surrounding the creation, management, marketing and preservation of digital assets.”

NISO Framework Advisory Group. (2004). *A Framework of guidance for building good digital collections* (2nd ed.). Bethesda, MD: National Information Standards Organization. Retrieved February 17, 2006, from <http://www.niso.org/framework/framework2.html>

A helpful book that includes guidance on selecting collections, managing projects, and generating metadata.

Research Libraries Group (Ed.). (n.d.). *RLG DigiNews*. Retrieved February 17, 2006, from

[http://www.rlg.org/en/page.php?Page\\_ID=12081](http://www.rlg.org/en/page.php?Page_ID=12081)

A bi-monthly electronic newsletter that focuses on digitization and digital preservation.

Research Libraries Group. (2002, May). *Trusted digital repositories: Attributes and responsibilities*. Mountain View, CA: Author. Retrieved February 17, 2006, from <http://www.rlg.org/legacy/longterm/repositories.pdf>

Useful for institutions that need additional information on digital repositories or who wish to adopt practices used by such repositories.

Sitts, M. K. (Ed.). (2000). *Handbook for digital projects: A Management tool for preservation and access*. Andover, MA: Northeast Document Conservation Center. Retrieved February 17, 2006, from <http://www.nedcc.org/digital/dighome.htm>

Comprehensive guide covering all aspects of image digitization, including copyright.

University of Illinois at Urbana-Champaign. Grainger Engineering Library. Digital Services and Development Unit. (2000). *Image Quality Calculator*. Retrieved February 17, 2006, from <http://images.library.uiuc.edu/projects/calculator>

This interactive tool allows users to input the size of an image, the size of the “smallest significant character or visual element,” and the desired quality and scan type. The calculator then determines the recommended resolution and approximate file size.

Washington State Library. (2005). *Digital best practices*. Retrieved February 17, 2006, from <http://digitalwa.statelib.wa.gov/newsite/best.htm>

Includes best practices information on the techniques of digitizing, project management, technology, and possible sources of funding.

Western States Digital Standards Group, Digital Imaging Working Group. (2003, January). *Western states digital imaging best practices*. Denver, CO: Western States Digital Standards Group. Retrieved February 17, 2006, from [http://www.cdpheritage.org/digital/scanning/documents/WSDIBP\\_v1.pdf](http://www.cdpheritage.org/digital/scanning/documents/WSDIBP_v1.pdf)

Extensive coverage of digital imaging, including project planning, quality control, and intellectual property concerns.

## RESOURCE LIST - SOFTWARE

The need to purchase software is a requirement for any digitization project or program. Exactly what is purchased is dependent upon the specifications of the program itself. The types of software packages to be considered and WNYLRC's recommendations are as follows:

**Audio Editing** – If analog audio is being converted into a digital format, it may be desirable to edit a copy of the file in order to provide smaller sections online or to create a different file format. WNYLRC recommends Adobe Audition® for this task. Audition is professional audio editing software that offers advanced audio mixing and editing capabilities. Additional information is available at <http://www.adobe.com>.

**Digital Asset Management** – Many software programs are available to help manage digital assets. Some are geared towards specific environments (e.g., PastPerfect for museums) and some are open source offerings (e.g., Greenstone). Many projects in New York State are using CONTENTdm from DiMeMa and OCLC, which is endorsed by WNYLRC. Additional information on CONTENTdm is available at <http://www.contentdm.com>.

**EAD Creation** – Currently, the Society of American Archivists (SAA) is recommending two different text editors for creating Encoded Archival Description (EAD): NoteTab Pro (<http://www.notetab.com/>) and XMetal Author from Blast Radius (<http://www.xmetal.com>).

There are other software available that can be used including XMLSpy by Altova (<http://www.altova.com>). The XMLSpy® 2006 Home Edition can be downloaded at [http://www.altova.com/support\\_freexmlspyhome.asp](http://www.altova.com/support_freexmlspyhome.asp) (free).

**Image Editing** – Image editing software is used to create and edit a copy of the original image. For example, a digital image may need to be edited in order to create a different file format or to display a section of the original image, rather than its entirety. Adobe Photoshop® is widely used for this and is endorsed by WNYLRC.

It should be noted that “CONTENTdm also provides a range of options for working with high resolution JPEG or TIFF files, from automatically creating lower resolution display copies that enable modifications to image quality and legibility, to using custom display copies you create.”<sup>5</sup> However, CONTENTdm may not provide the flexibility found in Adobe Photoshop.

**Optical Character Recognition (OCR)** – Two respected products are OmniPage Professional (<http://www.nuance.com>) and ABBYY FineReader (<http://www.abbyy.com>). OmniPage's web site call the product a “Powerful new OCR technology, advanced layout analysis and intuitive editing tools allow you to quickly turn paper and PDF files into more than 30 different editable electronic file formats that look just like the original – complete with text, tables and graphics.”<sup>6</sup> The ‘buzz’ on ABBYY FineReader is that it is one of the

---

<sup>5</sup> From <http://www.contentdm.com/products/features.html>, accessed Dec. 19, 2005.

<sup>6</sup> From <http://www.nuance.com/omnipage/standard/>, accessed Dec. 19, 2005.

best out there, particularly for cultural heritage projects. It will recognize a wide range of typefaces and fonts, and can handle poor quality print.

It should be noted that CONTENTdm does have an add-on product that provides OCR capabilities. “The feature uses ABBYY's award-winning FineReader® OCR software to capture text for addition to searchable metadata fields within CONTENTdm collections.”<sup>7</sup>

**Scanning** – In general, the software needed to manipulate the scanner and create the digital images of hardcopy materials will come with the scanner. That software should be able to service the digitization program's requirements. However, some high-end professional digitization operations will make very specific decisions on what software to use with their scanner hardware in order to achieve a different level of functionality. If a digitization program needs added scanning functionality, it should contact the hardware manufacturer and WNYLRC for recommendations.

**Video Editing** – Digitized video or film may need to be edited in order to create smaller sections to display online or to create a different file format. Adobe Premiere® Pro provides frame-accurate editing of digital video and audio in real time, and is endorsed by WNYLRC.

**Why does WNYLRC endorse the products above?** The Council recognizes the wide acceptance of these products within the library community. In addition, since these software packages are widely used, there is very good product support available both from the software companies and from other users.

**Will WNYLRC's software recommendations change?** Yes. As new software is developed, WNYLRC members should understand that the Council's recommendations may change. For up-to-date recommendations, please contact WNYLRC or check its web site.

---

<sup>7</sup> From <http://www.contentdm.com/products/ocr-pricing.html>, accessed Dec. 19, 2005.

## RESOURCE LIST - VENDORS

When embarking on a digitization project, an institution must decide if it wants to do the entire project in-house or if outsourcing is an option. If the institution decides to outsource portions of the project, there are many things that must be considered when thinking about working with a vendor, e.g., cost, digitization quality, and customer service. For information on working with vendors, consult the following resources:

- The **Northeast Document Conservation Center** provides a guide for working with digitization vendors as part of its *Handbook for Digital Projects: A Management Tool for Preservation and Access* (Section VIII, "Vendor Relations," <http://www.nedcc.org/digital/dman.pdf> or <http://www.nedcc.org/digital/viii.htm>). The section covers many of the basic areas that must be addressed including Request for Information, Request for Proposal, and contracts.
- The **California Digital Library** web site contains information on outsourcing, [http://www.cdlib.org/inside/diglib/resources/onres\\_dig\\_digitization.html#4](http://www.cdlib.org/inside/diglib/resources/onres_dig_digitization.html#4).
- The **Library of Congress** has published their Requests for Proposals (RFPs). These are real world examples that demonstrate how requirements are specified. <http://memory.loc.gov/ammem/techdocs/conversion.html>

There are many vendors who provide services to digitization projects and several lists of vendors have been developed. Below are links to lists of vendors. ***No recommendation or endorsement of any vendor is implied.*** In addition, do not assume that the institution that created any list is also endorsing the vendors listed. No list should be considered comprehensive, since new vendors may appear at any time. Please contact the vendors directly for additional information, references, and to discuss specific project requirements. In addition, talk to organizations that have completed digitization projects for suggestions and recommendations.

### ***New York State Archives – Managing Government Records /Imaging Consultants***

[http://www.archives.nysed.gov/a/nysaservices/ns\\_mgr\\_cons\\_vend.shtml](http://www.archives.nysed.gov/a/nysaservices/ns_mgr_cons_vend.shtml)

This site is maintained by the New York State Archives and contains lists of consultants and vendors in several areas including imaging. Specific information on a vendor's services may not be provided, so please contact the vendor to verify its area of expertise before requesting a quote or proposal.

### ***CDLC Digitization Plan w/list of vendors (pp. 26-42)***

<http://www.cdlic.org/pdfs/CDLCDigiPlan.pdf>

This list is primarily comprised of vendors located in New York State. The list was developed and verified during the summer of 2004.

### ***Digital Imaging and Data Conversion Vendors***

<http://gort.ucsd.edu/dlpwg/vendors.html>

This is a list of vendors with whom the University of California at San Diego's Digital Library Program has experience.

***Digitization Providers and Equipment Vendor***

<http://www.njdigitalhighway.org/documents/digitization-providers-and-vendors-2004-02-26.doc>

Maintained by the New Jersey Digital Highway. This is a list of digitization and equipment vendors, as well as resources (documentation) for digitization.

***Digital Conversion Service Bureaus***

[http://www.rlg.org/en/page.php?Page\\_ID=222](http://www.rlg.org/en/page.php?Page_ID=222)

Maintained by RLG who reports that, "RLG does not endorse these service providers, but has received positive reports from those who have used them."

***Vendors***

[http://www.cdpheritage.org/resource/scanning/rsrc\\_vendors.html](http://www.cdpheritage.org/resource/scanning/rsrc_vendors.html)

Maintained by the Colorado Digitization Program.

***Document Imaging Service Bureau Listings***

[http://www.imagepub.net/sb\\_list.html](http://www.imagepub.net/sb_list.html)

Extensive list of imaging services vendors provided by Imaging Publishing. Many specialize in digital imaging required for records management by businesses and healthcare institutions but some also do the kind of work required for digitization projects by cultural heritage organizations. This is not a comprehensive list, but contains those that register for inclusion.

**Preservation Suppliers and Services from the Northeast Document Conservation Center**

Reformatting - Digital Imaging

<http://www.nedcc.org/suppliers/supdig.htm>

Reformatting - Microfilm

<http://www.nedcc.org/suppliers/supmic.htm>

Reformatting - Video, Film and Other AV Reformatting Providers

<http://www.nedcc.org/suppliers/supvid.htm>

## GLOSSARY

### ***Authentication***

Authentication is the validation of a user, a computer, or a digital object to ensure that it is what it claims to be.

### ***Authority Control***

“The process of verifying and authorizing the choice of unique access points, such as names, subjects, and forms and assuring that the access points are consistently applied and maintained in an information retrieval system.”

(<http://www.ncecho.org/Guide/glossary.htm>)

### ***AIFF (Audio Interchange File Format)***

This is an audio format developed by Apple Computer for storing high-quality sampled sound.

### ***AVI (Audio Video Interleave)***

AVI is a common sound and video file format used with the Windows operating systems.

### ***BMP (Bitmap)***

BMP is a proprietary Microsoft Windows image format. BMP files are large and are not used on the Internet due to their sizes.

### ***Browser***

A general-purpose user interface, used with the Internet. Popular browsers include Netscape Navigator, Internet Explorer and Firefox.

### ***Copyright***

“Copyright is a form of protection provided by the laws of the United States (title 17, U.S. Code) to the authors of ‘original works of authorship,’ including literary, dramatic, musical, artistic, and certain other intellectual works. This protection is available to both published and unpublished works.” (<http://www.copyright.gov/circs/circ1.html#wci>)

### ***Crosswalk***

A crosswalk is the mapping of one metadata framework to another metadata framework. Crosswalks help promote interoperability.

### ***Cultural heritage organizations***

Cultural heritage organizations may include museums, historical societies and associations, archives, genealogical societies, religious institutions and others that contain materials related to human work and thought.

### ***Data aggregation***

Data aggregation is a method of massing together or clustering of independent but similar information. Well known aggregators in the library community are ProQuest, LexisNexis

and Dialog. In the context of a digitization program, a data aggregator might cluster – or point to – similar digitized materials.

***DCMI (Dublin Core Metadata Initiative)***

DCMI is the maintenance agency for the Dublin Core metadata schema.

***Descriptive content standards (e.g., AACR2)***

These content standards are sets of rules that describe the content that is contained in library catalog records.

***Digital Asset Management (DAM)***

Digital asset management is “a system that creates a centralized repository for digital files that allows the content to be archived, searched and retrieved.”

(<http://drmwatch.webopedia.com/TERM/D/DAM.html>)

***Digital collection***

To WNYLRC, a digital collection is a collection of objects (e.g., print, audio, artifacts) that were digitized from a variety of other formats. A digital collection may be part of a digital library, but in itself is not a digital library.

***Digital library (also known as an electronic library or virtual library)***

There is no one correct definition for a digital library. One might define it as collection of documents in organized electronic form, available on the Internet. A digital library can include magazine articles, books, papers, images, sound files, and videos.

To WNYLRC, a digital library provides electronic access to a variety of e-resources. A digital library is more than just digitized images.

***Digital object***

A digital object is an item as stored in a digital collection. It consists of the data itself (e.g., a digitized photo), metadata, and an identifier.

***Digital preservation***

“Refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary.” (<http://www.dpconline.org/text/intro/definitions.html>)

***“Bit-level” preservation***

Bit-level preservation is when a file is preserved in its original format only.

***Full preservation***

According to the Florida Center for Library Automation (FCLA) Digital Archive:

Full preservation includes bit-level preservation of the originally submitted files, as well as services intended to ensure that the information content of the files will remain usable into the indefinite future. These services vary according to the file type but may include the creation of normalized forms of the file and/or the reformatting of obsolete formats to reasonably comparable successor formats. It is not guaranteed, however, that normalized or migrated versions of any file will

be identical in functionality or in “look and feel” to the original file. Note also that if a logical object is comprised of individual files in both supported and unsupported formats, there is no guarantee that the logical object will remain usable as intended.

([www.fcla.edu/digitalArchive/pdfs/DigitalArchivePolicyGuide1\\_1.pdf](http://www.fcla.edu/digitalArchive/pdfs/DigitalArchivePolicyGuide1_1.pdf))

### ***Digital rights***

Digital rights are those intellectual property rights that an institution wants to grant to its users or retain for itself.

### ***Digital rights management (DRM)***

Digital rights management is a system for protecting the copyrights of data circulated via the Internet or distribution channels. Digital rights management can include techniques such as password protection, data encryption, and digital watermarks.

### ***Digitization***

Digitization is creating electronic copies of non-electronic materials. For example, when an item such as a book, photograph or an object is digitized, an electronic copy is created that can be stored and viewed on a computer. Digitization is done to increase access to materials by creating digital surrogates. Digitization does not preserve the original artifact.

### ***Document imaging***

Document imaging is often used synonymously for the word “digitization.”

### ***DTD (Document Type Definition)***

A document type definition is a set of rules for an XML document that specifies which elements (the markup tags) and attributes (values associated with specific tags) are allowed in the documents.

### ***DPI (Dots Per Inch)***

DPI is a term used in reprographics (printing) that describes the number of dots per inch used to create an image. DPI is often used interchangeably with PPI (pixels per inch) although the two are not the same.

### ***Dublin Core***

Dublin Core is a type of metadata structure to provide card catalog-like definitions for describing electronic resources. It was developed in association with OCLC and is named after Dublin, OH.

“The Dublin Core Metadata Element Set is a set of 15 descriptive semantic definitions. It represents a core set of elements likely to be useful across a broad range of vertical industries and disciplines of study.” (<http://www.dublincore.org/resources/faq/>)

More information on Dublin Core can be found at <http://www.dublincore.org/>

### ***Qualified Dublin Core***

“Qualified Dublin Core’ employs additional qualifiers to further refine the meaning of a resource. One use for such qualifiers are to indicate if a metadata value is a compound or structured value, rather than just a string.

“Qualifiers allow applications to increase the specificity or precision of the metadata. They may also introduce complexity that could impair the metadata's compatibility with other Dublin Core software applications. With this in mind, designers should only select from the set of approved Dublin Core qualifiers that were developed by the Dublin Core community process.

“Unfortunately, qualifiers often introduce additional complexity that can make metadata less interoperable unless approved DC Qualifiers developed within the DCMI are used with such interoperability considerations in mind.”

(<http://dublincore.org/resources/faq/#whatisthedifference>)

### ***Simple Dublin Core (Unqualified Dublin Core)***

“Simple Dublin Core’ is Dublin Core metadata that uses no qualifiers; only the main 15 elements of the Dublin Core Metadata Element Set are expressed as simple attribute-value pairs without any "qualifiers" (such as encoding schemes, enumerated lists of values, or other processing clues) to provide more detailed information about a resource.” (<http://dublincore.org/resources/faq/#whatisthedifference>)

### ***EAD (Encoded Archival Description)***

EAD is a type of metadata structure used to encode electronic versions of finding aids for archival materials. EAD can be used to represent complete archival structures, including hierarchies and associations. More information on EAD is available at <http://www.loc.gov/ead/>

### ***Emulation***

One method currently employed for preserving access to digital objects is emulation. With emulation, the old computer system (either hardware or software) is replicated in a new computing environment so that the information created with the old system can continue to be used.

### ***Encoding***

The activity of inserting specific computer codes (or “tags”), based on a specific tag set, in data files to designate various structural components of the text. This process is also referred to as “marking-up” the data files.

(<http://www.masshist.org/DIGITALADAMS/AEA/about/glossary.html>)

***Fair Use***

The doctrine of Fair Use is defined in the Copyright Act (title 17, U.S. Code). Section 107 contains examples of what is considered Fair Use as well as how to determine if a specific use would be deemed “fair.” The four factors used to determine Fair Use are:

1. the purpose and character of the use, including whether such use is of commercial nature or is for nonprofit educational purposes;
2. the nature of the copyrighted work;
3. amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
4. the effect of the use upon the potential market for or value of the copyrighted work.

If Fair Use cannot safely be determined, then permission should be sought before using the materials.

***Federated search (also called metasearch)***

Federated search is software that allows a user to search across content repositories (e.g., databases) using one common interface. The benefits of federated search are ease of use and obtaining one list of results. Although many resources can be searched using federated search software, not every database provider makes their content compatible with federated search software.

***Finding Aid***

A finding aid – or finding tool – is a “published or unpublished guide, inventory, index, register, calendar, list, or other system for retrieving archival primary source materials that provides more detailed description of each item than is customary in a library catalog record.” (From the *Online Dictionary for Library and Information Science*)

***GIF (Graphic Interchange Format)***

GIF image files generally have very low resolution, which means that they display quickly. They have a built-in compression. GIFs are frequently used on the Internet.

***Harvesting***

Harvesting is the gathering together of metadata from a number of distributed repositories into a combined data store.

***Identifier***

An identifier is “an unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system.”

<http://dublincore.org/documents/usageguide/glossary.shtml>

Examples of formal identification systems include Uniform Resource Locator (URL), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

***IMLS (Institute of Museum and Library Services)***

IMLS is an “independent federal grant-making agency dedicated to creating and sustaining a nation of learners by helping libraries and museums serve their communities.” IMLS often does this by providing grants to museums and libraries. It also oversees the distribution on the LSTA money from the federal government. [http://www.ims.gov/about/abt\\_mission.htm](http://www.ims.gov/about/abt_mission.htm)

***Interface***

“The interaction between the computer and the user or the control of the flow of data between a computer and its peripherals.” ([www.leprint.com/glossaries.html](http://www.leprint.com/glossaries.html))

***Interoperability***

The ability of software and hardware on multiple machines from multiple vendors to communicate with each other.

***JPEG (Joint Photographic Experts Group)***

JPEG is a common/standard image lossy compression mechanism. Because JPEG files are smaller than uncompressed files (e.g., TIFF files), the JPEG format is often used to display images on the Internet.

***Lossy***

Image files formats are described as either lossy or lossless. A lossy file format using a compression algorithm that discards some of the information or data to make the file take less storage space or less transmission bandwidth. A lossy file format will never be as good (i.e., contain as much data) as a lossless file format.

***Lossless***

A lossless file format is either uncompressed or used a compression format that retains all of the original information in the file. Generally, the high quality files made during a digitization project are lossless files and are retained (archived) long-term since they contain the most information from the digitization process.

***LSTA (Library Services and Technology Act)***

The Federal Library Services and Technology Act (LSTA) program provides federal funds that help support local libraries throughout New York State.

***MARC (MAchine-Readable Cataloging)***

MARC is a type of metadata structure. The MARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form. Most online catalogues used in libraries use the MARC formats. More information is available at <http://www.loc.gov/marc/>

***Markup language***

“A language that has codes for indicating layout and styling (such as boldface, italics, paragraphs, placement of graphics, etc.) within a text file. Widely used markup languages include SGML (Standard General Markup Language) and HTML (Hypertext Markup Language.” (<http://faculty.valencia.cc.fl.us/jdelisle/lis2004/glossary.htm>)

**Metadata**

Metadata is “data about data.” Metadata describes the content, quality, condition, and other characteristics of an item (e.g., book, photograph). MARC is a structure for capturing metadata, as is Dublin Core.

**METS (Metadata Encoding & Transmission Standard)**

“The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation.”

(<http://www.loc.gov/standards/mets/>)

**Migration**

Migration is a technique used to preserve digital content. Migration entails the replacing older file formats and internal structures with newer ones. For example, a JPEG file might be migrated to a newer version of that format. The assumption is that the older version will eventually not be supported, so it is better to migrate files to the newer, supported formats.

**MPEG (Moving Picture Experts Group)**

This is the standard for compression and storage of motion video.

**MP3**

MP3 is an acronym for MPEG-1 or MPEG-2 audio layer 3. MP3 is a standard for the compression of audio signals and is used frequently on the Internet.

**OAI (Open Archives Initiative)**

An initiative to develop and promote interoperability standards to facilitate the efficient dissemination of content (<http://www.openarchives.org/>).

**OCLC (Online Computer Library System)**

An organization that (1) provides a bibliographic utility for libraries to share catalog records, (2) sells and supports CONTENTdm for content management, (3) helped to develop the Dublin Core metadata standard.

**OCR (optical character recognition)**

OCR is software that converts scanned images containing text into editable text.

**Pathfinder**

A pathfinder – or topical guide – is “A subject bibliography designed to lead the user through the process of researching a specific topic, or any topic in a given field or discipline, usually in a systematic, step-by-step way, making use of the best finding tools the library has to offer.” (From the *Online Dictionary for Library and Information Science*)

**PCT (PICTure)**

PCT is the bitmap picture format used on Macintosh computers. It is somewhat similar to the BMP format.

***PDF (Portable Document Format)***

PDF “is a file format that has captured all the elements of a printed document as an electronic image that you can view, navigate, print, or forward to someone else. PDF files are created using Adobe Acrobat, Acrobat Capture, or similar products. To view and use the files, you need the free Acrobat Reader, which you can easily download. Once you've downloaded the Reader, it will start automatically whenever you want to look at a PDF file.” ([http://www.cesa8.k12.wi.us/media/digital\\_dictionary.htm](http://www.cesa8.k12.wi.us/media/digital_dictionary.htm))

***Pixels Per Inch (PPI)***

This is a measure of the resolution used with display monitors and other electronic devices. It is a measure of the sharpness (or the density of illuminated points). The higher the PPI, the higher (better) the image quality.

***Portal***

A portal is a web site that serves as a gateway to additional information or resources. In general, a portal may have a specific purpose or focus.

***RBDB (Regional Bibliographic Data Bases and Interlibrary Resources Sharing Program)***

RBDB stands for the Regional Bibliographic Data Bases and Interlibrary Resources Sharing Program. This grant from New York State provides funding that can be used for increasing access to materials. RBDB funds have been used for converting card catalog records into electronic form. Recently library councils have begun to use RBDB funds to assist libraries in their digitization efforts.

***Refresh***

Refreshing is a technique used to preserve digital content. When files are refreshed, exact copies are made of them on newer media. This is done because of the concern that older media may have a limited shelf life (or may have already outlived its shelf life).

***Resolution***

Resolution equates to the number of pixels in an image. The more pixels, the higher the resolution. Conversely, the higher the resolution, the better the digital image.

***Scanning***

Scanning is the process of converting a physical object – e.g., a printed page, photograph, slide, or map – into a digital format that is usable by a computer.

***TEI (Text Encoding Initiative)***

TWI is a specific DTD (or tag set) that is useful for encoding texts.

***TIFF (Tagged Image File Format)***

TIFF is an industry standard file format developed for the purpose of storing high-resolution bit-mapped, gray-scale, and color images. It is considered a lossless format.

***Trusted repository***

“...has the mission to provide reliable long-term access of managed digital resources to its designated community, now and in the future.”

([www.ffos.hr/lida/lida2005/datoteke/lida2005-smid\\_majcenovic.ppt](http://www.ffos.hr/lida/lida2005/datoteke/lida2005-smid_majcenovic.ppt))

***Watermark***

A watermark is a code that is embedded into digital material in order to establish ownership. The code may be visible (e.g., text or logo) or invisible to the user.

***WAV (Waveform Audio)***

WAV is an uncompressed audio format and is commonly used with computers running the Windows operating system.

***XML (eXtensible Markup Language)***

“Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere.” (<http://www.w3.org/XML/>)

***Z39.50***

Z39.50 actually refers to the numbers of the ISO and ANSI/NISO standards (ISO 23950 and ANSI/NISO Z39.50) that have to do with search and retrieve protocols. Z39.50 is a standard that specifies a technical protocol for searching and retrieving information from multiple remote online catalogs or databases, regardless of their native search commands.

Additional definitions are available at:

- DCMi Glossary, <http://dublincore.org/documents/usageguide/glossary.shtml>
- Digital Dictionary, [http://www.cesa8.k12.wi.us/media/digital\\_dictionary.htm](http://www.cesa8.k12.wi.us/media/digital_dictionary.htm)
- Digital Preservation Coalition, <http://www.dpconline.org/graphics/intro/definitions.html>
- EduTools Wiki, <http://www.edtechpost.ca/pmwiki/pmwiki.php/Main/GlossaryAnalysis>
- Google (<http://www.google.com>), where you can use the define command to locate definitions. For example, type in the search: define: xml
- Moving Theory Into Practice, <http://www.library.cornell.edu/preservation/tutorial/intro/intro-01.html>
- NC:ECHO Glossary, <http://www.ncecho.org/Guide/glossary.htm>
- Online Dictionary for Library and Information Science, <http://lu.com/odlis/>
- Glossary from *Digital Libraries* by William Arms, <http://www.cs.cornell.edu/wya/DigLib/MS1999/glossary.html>

## **APPENDIX**

## DIGITIZATION SURVEY RESULTS

In the fall of 2005, WNYLRC created an online survey to gather information from its members and from members of the Document Heritage Program (DHP) and local government communities. 109 members completed the survey along with 14 people from the DHP/local government areas. The surveys were reviewed by the Digital Heritage Advisory Subcommittee and used in construction of this plan. What follows here is a summary of the results. The complete results are available from WNYLRC.

### *WNYLRC Member Survey Summary*

**Demographics:** The respondents self-identified with the following primary job functions (top 5, multiple answers allowed):

# of Respondent	Primary Job Function
33	Reference
22	Administration
22	I'm a one person shop
20	Cataloguing
16	Instructor
16	Collection Development

80 of the respondents worked at institutions in located in Erie County. 14 were from Cattaraugus. The rest were scattered among the other counties serviced by WNYLRC.

In looking at the types of libraries where the respondents worked, they provided the following information:

# of Respondent	Type of Library
61	Academic library
24	School library or library system <sup>8</sup>
13	Public libraries or public library systems.
4	Museum / archives
4	Corporate
4	Other
2	Hospital

---

<sup>8</sup> Later in the survey, 20 self-identified themselves as school librarians.

**Possible WNYLRC Services:** In response to “What digitization-related services do you think you would use if offered by WNYLRC?” respondents offered no clear mandates. All of the items listed (save “Digitize collections on your behalf”) received the most votes (29 – 49) for “likely to use.” The list of possible services included (with number of votes for “Will Use” and “Likely to Use”):

Possible Service	Will Use	Likely To Use	Total
Training	23	47	70
Provide networking opportunities for librarians	17	52	69
Assist in identifying funding sources	18	47	65
Provide information on suitable digitization vendors	10	49	59
Assist in writing a grant for a digitization project	15	40	55
Provide info on hardware/software for digitizing	11	43	54
Locate potential collaborators	8	44	52
Consortial pricing for obtaining/offering digitization training	10	40	50
Consortial pricing on digitization hardware/software	10	39	49
Consulting on digitization options for your collection	6	40	46
Assist with outreach, promotion & curriculum	6	40	46
Review of your collection for potential digitization	7	37	44
Fund a single interface to access multiple collections	11	32	43
Establish a regional digitization center for your use	9	32	41
Assist with project management	2	29	31
Consortial pricing for joint storage & retrieval	9	29	31
Digitize collections on your behalf	5	22	27

**Training:** 68 respondents had attended no digitization-related training prior to completing the survey. In looking ahead to the digitization-related training being offered by WNYLRC, the majority of respondents either were not going to attend the sessions or were unsure. Only a few (generally 13 – 19%) were sure that they were going to attend a specific session. A majority (53.8%), however, said that they intended to go to the Digitization Expo in May 2006. Respondents did suggest additional workshop topics:

- Encoded Archival Description (EAD)
- Basic training (held on Saturday or in the evening)
- 'Digitization for Dummies'
- Searching and using a new digitized collection, e.g., workshop on using a digitized version of many of the historical records at the Buffalo Historical Society
- Available digitized collections (and how to use them)
- Success stories (e.g., the digitized snowflake collection at the Museum of Science)
- Integration of digital assets
- Digital asset management
- Grant funding for digitization training and/or implementation (including small collections)
- Metadata training

- Outreach and curriculum integration
- How to organize digital documents
- How to create taxonomies or thesauri on your own
- How to use existing taxonomies
- How to organize your institutions digital documents
- How best to create an internal policy/procedure from a basic template. Including this type of 'blueprint' for any project - not only digitization - is appreciated.
- I would like training in what is out there, once you have it all set up and ready to go. I don't really think that I have anything that should really be digitized at this time, but I am not sure.

One person noted that it can be very difficult to leave work during the day and suggested an evening or Saturday workshop. The person also suggested that the Council teach a class at the Buffalo Teacher Center (e.g., 3 hours a day from 4 - 7 p.m., once a week for 5 weeks) or conduct a session at one of the Library In-services.

**Regional Digitization Center:** Respondents were asked what type of materials they might want to digitize at a regional digitization center, if one existed. The top responses were: (multiple responses allowed)

# of Respondent	Type of Materials
55	Photographs
41	Bound materials
37	Unbound sheets of paper
35	Photograph slides or negatives
35	Oversized sheets of paper
27	Microfilm/microfiche
27	Artifacts
26	Cartographic/architectural materials

59 respondents said they would use the center if the cost was lower than digitizing materials using other vendors.

In other part of the survey, one person wrote: “Although I can see a benefit to having a regional digitization center -- I am not in the position to make any of these decisions for my institution.”

**WNY Digital Heritage Portal:** 36 respondents said that if they digitized materials in the next two year (or if they have already digitized materials), they would consider using the same software as other institutions in the region so that end-users could search several repositories at once. 51 people, however, were unsure about helping to share the cost of such software. When asked if they would the federated search tool be capable of searching various vendors’ software to eliminate the requirement of all libraries having the same software, an equal number of people (34) responded both “yes” and unsure.

### Digitization Efforts:

- 25 felt that their institutions would collaborate on a digitization project with another institution in the region that had similar or complementary subject themes.
- 32 said that their institution had already digitized part of their collection.
- Of those who had not digitized materials already, it was more probable that they would not begin a digitization project in the next three years.
- If an institution had digitized materials already, 87.1% felt that they would digitize again.

Respondents mentioned the following digitized collections that are available online:

- Technical Reports Collection, Arts & Science Library, Univ. of Buffalo, [http://ublib.buffalo.edu/libraries/asl/guides/tech\\_repts.html](http://ublib.buffalo.edu/libraries/asl/guides/tech_repts.html)
- History of the City of Buffalo and Erie County, <http://www.niagara.edu/library/buffhist/eriehome.html><sup>9</sup>
- Digital Collection of Niagara Falls Guidebooks from the 19th Century, <http://www.niagara.edu/library/nfguides/gdhome.html><sup>10</sup>
- UBdigit, <http://ubdigit.buffalo.edu/>
- Extra! Extra! (newspapers), Prendergast Library, <http://www.prendergastlibrary.org/Extra!Extra!/index.htm>
- Pamphlets, Prendergast Library, <http://prendergastlibrary.org/jamestown/information.html>
- Archive collections (including photographs), Rev. J. Clayton Murray S.J. Archives, Canisius College, <http://www.canisius.edu/archives/>
- The New York to Paris Race, Buffalo and Erie County Public Library, <http://becpldigital.cdm.oclc.org/>
- Polish Royal Documents, University of Buffalo, <http://ublib.buffalo.edu/libraries/asl/guides/polish-room/documents/documents.html>
- Franciscan Engravings & Woodcuts, <http://web.sbu.edu/friedsam/scan/>

Two respondents noted:

- “Recent changes in the institutions web site have buried the items we have digitized. There was no real utility in the items we digitized either. No metadata or real means of organizing and retrieval.”
- “I think it was put on CD-ROM and is not on the web at this time.”

---

<sup>9</sup> This project received RBDB funding from WNYLRC.

<sup>10</sup> This project received RBDB funding from WNYLRC.

### Digitizing Finding Aids:

- 20 respondents said that their institution had digitized finding aids, with most (13) reporting that the finding aids were in HTML.
- 31 reported that they would like to learn more about encoded archival description (EAD).
- 26 reported that they would be interested in a regional cooperative project to create EAD encoded finding aids and make them available on the web.

**School Librarians:** The survey asked some specific questions of the school librarians. 20 school librarians completed this section of the survey.

- When asked if their school or school system was interested in using collections digitized by other libraries in the region, 7 answered “yes,” while 13 answered “unsure.” The digital collections would be used for:
  - library pathfinders faculty support – curriculum materials library marketing – 'look what we can offer' the teachers
  - classroom research
  - Primary source documents in the classroom.
  - Preserving and promoting local history to our students, especially in the social science classrooms.
  - Primary source documents for research
  - Classroom instruction and integrated into the curriculum (5-12 grades).
  - Classroom instruction and for assistance in science and social studies, etc.
  - Primary source documents to support the Social Studies curriculum.
- 15 said that they were unsure if their school or school system was interested in digitizing materials.
- School librarians would like to see the following materials digitized:
  - Buffalo Neighborhoods – history and current events Urban Planning, Local History, Civil Rights, Protests/Politics, Environmental, Technology, Community Service Opportunities, School Curriculums
  - New York State local history
  - Local history artifacts.
  - Historical documents – old papers, yearbooks, out-of-print historical documents, photographs, student artwork, etc.
  - Local history of the various small communities in the rural counties.
  - Local newspaper content.
  - Any historical collections of Buffalo history. African American history of the area. Ephemera from any era. Newspapers and organization collections.
  - West Valley Demonstration Project Data
  - Histories of the areas served
  - Erie Canal, history and people of the western NY area and how they impacted history
  - Local History, Erie Canal, Native Americans
  - Agricultural, local history topics - railroads, natural history, etc.
  - Stereoviews of WNY history – the Pan Am, Niagara Falls, building of downtown, etc.

- One person said: “Honestly, we have not given this any thought at all. No one has discussed it.”
- As for their own (school) materials, they would perhaps digitize the following:
  - Archival material of school.
  - Historical documents – old papers, yearbooks, out-of-print historical documents, photographs, student artwork, etc.
  - Photographs from the early days, old copies of student work that reflects the history of the school.
  - Old student newspapers, memorabilia about the Christian Brothers and the history of St. Joseph's Collegiate Institute.
  - School records.
  - Documents related to the history of the district.
  - Local history archived materials in individual libraries.
  - Scrapbooks.

**Themes/Collections in the Region:** In response to the question:

WNYLRC is interested in compiling a list of unique and high impact collections in this region. Please take a few minutes and tell us about your collections. What do they document or what stories do they tell? What is the format of the materials (e.g., photographs, audiotapes, books)?

Survey participants gave the following information:

Albright-Knox Art Gallery	The collections have a strong narrative about contemporary art of the modern era. There is ephemera, exhibition catalogs, annual reports from major institutions, information on the Buffalo Fine Arts Society, etc.
BECPL	Books, articles, mostly paper/photo-type, etc of specific interest in the community that I serve
Buffalo & Erie County Public Library	The B&ECPL has extensive local history and rare book collections which are too broad to go into detail here. The formats are found in answer to question 15 previously.
Buffalo and Erie County Historical Society Research Library	All formats, mostly on paper, on just about any subject concerning the history of Buffalo.
Buffalo and Erie County Historical Society Research Library	Larkin Company records (documents and photographs); Erie Canal history (documents, letters); Pan-American Exposition history (letters, scrapbooks, photographs, ephemeral materials); McKinley assassination (letters, scrapbooks, photographs, ephemeral materials); local Native American history (papers, treaties, letters); Millard Fillmore (letters, papers, agency records).
Buffalo and Erie County Library	Shaker Collection - 275 volumes of original writings and hymns. (B&ECPL Central Library - Grosvenor Rare Book Room) Bible Collection - a collection of significant editions (B&ECPL Central Library - Grosvenor Rare Book Room) English Literature Collection - 1,000 rare volumes of English literature, including the four Shakespeare folios of 1623, 1632, 1664, and 1685. (B&ECPL Central Library - Grosvenor Rare Book Room) History of Science and Medicine Collection (B&ECPL Central Library - Grosvenor Rare Book Room)
Buffalo State	Buff State has a great collection of historical documents relating to Buffalo...the Buffalo Courier, etc.

Buffalo State College	Fronczak Collection - Polonia Courier Express Library
Canisius College	Canisius College activities since college founding.
Canisius College Library	Archives - the history of Canisius College - paper based collection, yearbooks, newspapers, etc.
Christ the King Seminary	We have a small collection of books on the history of the Niagara Frontier and French Canada.
D'Youville College	Photographs, pamphlets, slides, newspapers
Erie Community College	I'm not sure exactly what we would want to digitalize if anything. It would have to be a group decision.
Erie Community College North	College Archives--various formats
Erie Community College/North	The official archives of Erie Community College including Board of Trustees minutes, photographs, news clippings, artifacts, college catalogs, yearbooks, graduation & other programs -- all materials that document the development of ECTI/ECC over 50+ years.
Health Sciences Library UB	History of Medicine has many unique items that would lend itself most readily. Exhibit material.
Honeywell	our collection is published material that is covered by copyright so I do not see anything that can be digitized at the moment; I need to learn more about it; it also is not a high priority so time spent on it will be limited
James Prendergast Library	We have a few things online such as newspaper clippings, photos, and pamphlets. May be including things such as obituaries and death notices
Lackawanna Public Library	The Lackawanna Public Library has a collection of local newspapers dating back to early 1900's as well some old photos.
Lockwood (ASL), Univ. at Buffalo	I am in charge of the Polish collection, which has mostly books and journals, but also some unique archival materials and rare published materials, some in very bad physical condition. We recently acquired some materials that I believe are not held elsewhere, materials of historical value.
MCEER Information Service	Multimedia collections pertaining to earthquakes, hurricanes and other natural hazards, including specific events such as the December 2004 tsunami, Kobe & North Ridge earthquakes, etc.
Medaille College	Medaille has some old, possibly free of copyright restrictions, books that pertain to local history. While searching through the archives, I have found some old photographs of Buffalo and the Olmstead Park area, but not a huge cache of these.
Medaille College Library	I haven't worked at Medaille long enough to answer this fully.
Medaille College Library	Some Buffalo History material, primarily books. Medaille's archive; manuscripts, papers, photos, which tell the history of the school and the Sisters of Saint Joseph.
NCCC	student newspapers
Niagara University	Niagara Falls Guidebooks from the 19th century Various histories of Niagara County written in the 19th century. Niagara University student newspaper documents the history of an important institution in Niagara county.
North Collins Historical Society	Civil War--minutes, photos, tintypes, artifacts Town Board Meetings--cassettes
Olean Public Library	Genealogy
Olean Public Library	Local history
Olean Public	Local History materials for Olean and area including maps.

Library	
Orleans/Niagara BOCES	System office, just a few professional materials. Media Services collection is only open to participating school districts, there is a fee involved.
Project FLIGHT at Buffalo State	Western New York Women's Hall of Fame materials, both text and visual. Various tools for tutoring early readers. Adult literacy tools. Children's resources.
St. Bonaventure University/Friedsam Memorial Library	1. Extensive rare book collection that covers mostly theological subjects. 2. Subject collections in university archives, which comprise a variety of formats--handwritten materials, journals, photographs, audiotapes, video, books. The subjects range from university history to specific things, such as the Thomas Merton Collection, Robert Lax Collection (international poet and friend of Thomas Merton), American Terrorist Collection (Dan Herbeck and Lou Michel), Jim Bishop Collection, Douglas Edwards Collection, Robert Golden Collection, SLA Marshall Collection, Brother Juniper (a comic strip character born in Friedsam Library), and the Lewis Songer Masthead Collection (a unique look at a half century of newspaper publishing)
SUNY Fredonia/Reed Library	local history (Chautauqua and Cattaraugus counties, primarily); Seneca Nation collections; all types of formats
UB Law Library	1. John Lord O'Brian Papers Documentation of the career of attorney, John Lord O'Brian (1874-1973) Primarily copies of typed manuscripts with some newspaper articles and photographs 2. Howard Berman Papers The personal papers of Howard R. Berman (1945-1997)include research materials on human rights in general, the sources and development of aboriginal rights of Indian nations within North America and their modern applications with regard to the Haudenosaunee (Six Nations, Iroquois Confederacy), indigenous peoples from around the world, and the international forums that address rights of minorities and indigenous populations, i.e., the International Labor Organization (ILO) and the United Nations (UN). Primarily typed manuscripts.
University at Buffalo	Archaeology slides and material artifacts from around the world, pre-1900 oversized site reports containing drawings, illustrations, and maps, rare and valuable pre-1900 monographs in brittle condition.
University at Buffalo	History of Medicine Collection James Joyce Collection Poetry Collection Zine Collection
University at Buffalo	See <a href="http://ubdigit.buffalo.edu">http://ubdigit.buffalo.edu</a> Numerous digital collections covering Pulp Fiction Cover Art, Music collections, opera photographs, archival collections, biology curriculum materials.
University at Buffalo	UB has a VERY extensive technical reports collection--millions on microfiche UB did a digitization project 5+ years ago on government technical reports. It needs to be expanded and updated Others from UB have a MUCH greater knowledge and I'll hope they respond to this survey
University at Buffalo Health Sciences Library	Historical collections of health sciences disciplines, including books, journals, realia, artifacts, and instruments. We also have videos.
University at Buffalo Health Sciences Library	UBdigit <a href="http://ubdigit.buffalo.edu/">http://ubdigit.buffalo.edu/</a> As listed on the website: 'UBdigit includes primarily collections of still images, but anticipates future inclusion of a variety of digital media formats, including audio, video, kinetic images, animation, virtual reality, interactive sequences and multi-media constructs.'
University at Buffalo Libraries	Map and Aerial Photograph Collection. See our website: <a href="http://ublib.buffalo.edu/libraries/asl/maps/map_room.html">http://ublib.buffalo.edu/libraries/asl/maps/map_room.html</a> emphasis on Buffalo and WNY region for maps, photos, other carto data.
University at Buffalo Libraries	UBdigit < <a href="http://ubdigit.buffalo.edu">http://ubdigit.buffalo.edu</a> > Interdisciplinary multi-media digital collections. UBdigit includes primarily collections of still images, but anticipates future inclusion of a variety of digital media formats, including audio, video, kinetic images, animation, virtual reality, interactive sequences and multi-media constructs.
University at	The major Music Library image collections are already represented online in UBdigit. The

Buffalo Music Library	audio collections, largely on reel to reel tape, need attention but public access to any digitized product would probably be limited due to rights. Just a note - I am not fully qualified to speak about the University Archives collections but know that what they have includes the Frank Lloyd Wright materials, Love Canal, and university history. All have potential for connections with other local collections.
University at Buffalo/Science & Engineering Library	Our microfiche collection of technical reports is one of the largest in New York State.
Villa Maria College	Glass slides of Buffalo historical settings - early 1900.
Villa Maria College Library	Not sure about other materials, but we own several volumes of 'Echo Myzyczne,' a Polish music periodical published in Chicago. According to WorldCat, the only other library with this set is the U. of Pittsburgh. 'Polonia' is strong in WNY, so perhaps digitization would be worthwhile.

**What Should be Digitized:** When asked “What topics or subjects, that are important to the region, would you like to see digitized?” respondents answered:

- Historical records of cultural institutions (Buffalo Philharmonic, Albright-Knox, Science Museum, Historical Society, Hallwalls, Shea's, etc.) in the Western New York region.
- Similar College and university archives; arts development in WNY including significant written texts and visual images.
- The UB Archives' Love Canal Collection
- Materials relating to the history of towns and counties in the region.
- Obviously anything regional.
- Anything on Western New York history – there are lots of virtual reference questions about this, many from outside the region, and having this material available digitally would be a great service.
- Buffalo and New York State History. Rare Books and/or manuscripts (U.B. has the Henry James Collection). Cartographic History of the region and state. Archaeological collections of region and state.
- Historical information, especially local industries.
- Ethnic, Niagara Movement, etc.
- History and current local government documents if they are in the public domain.
- County and local community histories. Could be a good collaborative project.
- Specific to the community
- Genealogy
- Early Buffalo history
- Genealogy records, county records
- Erie Canal papers and photographs; African-American historical materials; other minority-population materials; Sanborn maps and other Erie County atlases; Buffalo industrial history materials; Buffalo newspapers (Courier Express, Commercial Advertiser, Daily Courier, and Evening News).
- Buffalo and Erie County History Early Photographs Architecture Rare Books
- Local history of African Americans and other diverse groups in Western New York
- Complete record of the Buffalo News, and its precursors in the 19th century. UB Special Collections Genesee Valley Historical Collection (at SUNY Geneseo) Buffalo Historical Society appropriate materials Albright Knox Art Gallery appropriate materials. Buffalo Museum of Science appropriate materials.

- Historical documents.
- historic papers and artifacts of local people (from historical societies and other hist. sources)
- Historical regional info; unique to area; historical info from local trade organizations; I am historian for ASQ Buffalo (American Society for Quality) and have some materials that should be digitized and some put on the ASQ Buffalo web site. I have a list of local org's - You could also digitize history info from local MASONS, unions, churches any org that has records - a definitive statement about the area especially as many of these are downsizing or consolidating for lack of members - catch them before they throw records out.
- local history rare/fragile research collections Love Canal collection
- See above. History of the South Campus. Pan American exhibit. Uncrowned Queens - work by Peggy Bertram, Underground Railroad items, Campaign for Preservation - Tim Tielman's group - architectural buildings, The religious buildings locally - many Catholic churches, and other Houses of Worship, especially those slated to be closed.
- Steel or manufacturing industry history (especially the union histories). Photographs of industries/factories at their peek.
- The history of education in Buffalo has not been fully studied and understood, and part of the reason is the difficulty of studying all the primary documents. These would be a very good project: the documents of the Board of Education, starting from the earliest years; the public high school magazines and yearbooks, which I hope are still in the BECPL Central branch; the private school records and the records of the colleges. If the project started with everything up to the year 1920, for example, it might be manageable.
- Regional architecture Regional arts & crafts movement (e.g. Roycroft), Labor history, African-American history, incl. Niagara Movement & Underground RR, PanAmerican Exposition
- Information on local artists, galleries and other institutions.
- historical primary sources; ethnological/ethnic group materials
- Local history
- Pan American Exposition, Lilydale, Niagara Falls History

**Other Collections Worth Digitizing:** When asked what collections in the region the respondents would like to see digitized, they responded:

- AJCU - Association of Jesuit Colleges and Universities - is looking into consortial arrangements too to digitize and make accessible items having to do with the Jesuits
- Buffalo & Erie County Historical Society (local history); SUNY at Buffalo (rare books); Niagara Falls Public Library (Niagara Falls materials)
- Buffalo History Museum
- Buffalo State has the Fronczak collection, which includes archival resources, and they also have some microfilms and other materials relating to the history of Polish Americans in Buffalo. Likewise, the BECPL Central branch has many important resources dealing with immigrant communities, including the Poles & Polish Americans. The Historical Society, Canisius College (to lesser extent), and the Czytelnia Polska, now in one of the Catholic schools, have rare and perhaps unique resources on Buffalo's Polonia.
- Cattaraugus County Historical Society - County History Olean Historical and Preservation Society - Olean History
- Erie County gov't has aerial collections similar to ours of different years.
- I believe that the Historical Society also holds some of the same materials.

- In terms of tech reports, Buffalo & Erie County Public downtown and ECC North have significant collections.
- Mentioned above. We also own many Polish books that are still in boxes and may prove unique.
- Natural History Museum (archaeology, anthropology)
- Niagara County Historical Society (Melissa Dunlap, Exec. Dir) has extensive collections with photographs, artwork (in large frames) and historical books.
- No, but Hauptman's crystal collection is simply fabulous. I would love to see that digitized and cataloged.
- possibly the Amherst Museum
- Potential for complementary collections at Buffalo & Erie County Public Library, Buffalo & Erie County Historical Society, Buffalo State College, Albright Knox Art Galleries - Library & archives
- The Historical Society and the One-Room Schoolhouse
- UB - Polish materials

### ***DHP and Local Government Survey Summary***

**Demographics:** There were 14 respondents to WNYLRC’s online digitization survey for members of the DHP and local government communities. 10 respondents identified themselves as being paid staff. 8 said they worked for non-government institution, while 6 worked for a local government. 6 were from Niagara County, while the rest were scattered among the other five counties served by WNYLRC.

**Possible WNYLRC Services:** In response to “What digitization-related services do you think you would use if offered by WNYLRC?” respondents said that they would use or would likely use all of the services listed, except “fund a single interface to access multiple collections “(the majority were unsure). The list of possible services included (with number of votes for “Will Use” and “Likely to Use”):

Possible Service	Will Use	Likely To Use	Total
Training	5	6	11
Assist in identifying funding sources	5	5	10
Provide information on suitable digitization vendors	4	6	10
Locate potential collaborators	4	6	10
Consulting on digitization options for your collection	4	5	9
Consortial pricing for obtaining/offering digitization training	3	6	9
Assist with outreach, promotion & curriculum	2	7	9
Review of your collection for potential digitization	4	4	8
Assist in writing a grant for a digitization project	4	4	8
Provide info on hardware/software for digitizing	3	5	8
Establish a regional digitization center for your use	5	3	8
Digitize collections on your behalf	2	6	8
Consortial pricing on digitization hardware/software	2	6	8
Consortial pricing for joint storage & retrieval	3	5	8
Assist with project management	3	4	7
Provide networking opportunities for librarians	2	4	6
Fund a single interface to access multiple collections	0	4	4

**Training:** 9 had attended no digitization-related training prior to completing the survey. Three listed the following training sessions that they had attended:

- 1996 WNYLRC Digitization Seminar
- With ARMA and SAA and some with NYS Archives
- DHP workshop on photo scanning and copying

In looking ahead to the digitization-related training being offered by WNYLRC, the majority of respondents either were not going to attend the sessions or were unsure. Only a few (2 – 4 people) were sure that they were going to attend a specific session. 7 respondents, however, said that they intended to go to the Digitization Expo in May 2006.

**Regional Digitization Center:** Respondents were asked what type of materials they might want to digitize at a regional digitization center, if one existed. The top responses were: (multiple responses allowed)

# of Respondent	Type of Materials
10	Oversized sheets of paper (larger than 8.5 x 11")
9	Bound materials
8	Photographs
7	Unbound sheets of paper
7	Cartographic/architectural materials
7	Photograph slides or negatives
7	Microfilm/microfiche

9 respondents said they would use the center if the cost was lower than digitizing materials using other vendors.

**WNY Digital Heritage Portal:** 7 respondents said that if they digitized materials in the next two years (or if they have already digitized materials), they would consider using the same software as other institutions in the region so that end-users could search several repositories at once. 10 respondents, however, were unsure about helping to share the cost of such software. When asked if they would the federated search tool be capable of searching various vendors' software to eliminate the requirement of all libraries having the same software, 5 answered "yes," while 7 were unsure.

#### **Digitization Efforts:**

- 5 felt that their institutions would collaborate on a digitization project with another institution in the region that had similar or complementary subject themes.
- 4 said that their institution had already digitized part of their collection.
- Of those who had not digitized materials already, it was more probable that they would begin a digitization project in the next three years.
- If an institution had digitized materials already, 3 felt that they would digitize again.

#### **Digitizing Finding Aids:**

- 8 respondents said that their institution did not have had digitized finding aids. Only 2 institutions had digitized finding aids.
- 6 reported that they would like to learn more about encoded archival description (EAD).
- 6 reported being unsure if they would be interested in a regional cooperative project to create EAD encoded finding aids and make them available on the web.

**Themes/Collections in the Region:** In response to the question:

WNYLRC is interested in compiling a list of unique and high impact collections in this region. Please take a few minutes and tell us about your collections. What do they document or what stories do they tell? What is the format of the materials (e.g., photographs, audiotapes, books)?

Survey participants gave the following information:

Archives/Sisters of St. Francis of Holy Name Province, Inc.	The collection documents the history of the life, administration, and ministries of the Sisters of St. Francis, Stella Niagara, NY. The collection consists of photographic material, some audiotapes, some books, some architectural records, primarily unbound single sheets (administrative files).
Documentary Heritage Program	DHP Lending Library - maybe some items
Felician Sisters Archives/Immaculate Heart of Mary Province	Documents & materials of any form, made or received by the province since 1900, in the pursuit of its religious and legal obligations and in the transaction of its business.
Historical Assoc of Lewiston	The Lewiston Museum holds many unique images: art, photos, slides, negatives, drawings. These images (approximately 500 of them) are much in demand by residents, authors, businesses, and returnees. The images tell of early settlers and their families, buildings and businesses, and the aesthetics of a location on the Niagara River viewing Lake Ontario. Underground Railroad connections are prevalent in Lewiston, and need illustrated for future generations. Books and ledgers tell of early portages around Niagara Falls and early commerce in a village once larger than Buffalo, and a hub of transportation across the Lakes and St. Lawrence to the sea.
Old Fort Niagara	The Old Fort Niagara collection consists of over 101,000 archaeological specimens excavated from the fort and related sites; nearly 2000 historic items related to the site's history and spanning the eighteenth, nineteenth and twentieth centuries; roughly 8,000 photographs and more than 5,000 books, maps, and periodicals related to regional and fort history.
St. Bonaventure University	History of St. Bonaventure University (photographs, slides, audiotape (cassette, reel-to-reel), videotape (all formats), paper records; Journalism collections: Douglas Edwards, Jim Bishop, Roi Ottley
The D. R. Barker Historical Museum	The D. R. Barker Historical Museum has a very extensive Civil War Collection. It also has a good cemetery record collection. The Barker has records of the earliest academy of higher education in the county. The collections also contain materials on the Grange located in Fredonia, the first Grange in the country. There is also a large photograph collection including prints and negatives, most photographs are of local interest. Many objects express pioneer living including portraits and objects. Archival materials expressing early pioneers include items such as ledgers and newspapers.
Tonawanda Reservation Historical Society	Our collection documents the history of the Tonawanda Reservation and its residents. It also contains some general information on Iroquoian history and culture. It depicts daily activities, family histories, and notable events. Our collection includes photographs, scrapbooks, newsletters, audiotapes, videotapes, maps, artifacts, and ephemera.

**What Should be Digitized:** When asked “What topics or subjects, that are important to the region, would you like to see digitized?” respondents answered:

- Women, natural history, Civil War and civil rights, literature, ethnic and racial communities, etc.
- Early settler images Buildings and businesses, photos and ledgers
- Lake Ontario Ordnance Works history and archives
- Underground Railroad history and reenactments
- Nineteenth Century life
- Preserving cemetery records
- Preserving early newspapers
- Preserving Civil War diaries
- Preserving documents that describe the establishment of the Grange, WCTU, the Town of Pomfret, Village of Fredonia
- Town of Pomfret Records
- Native American culture and history

**Other Collections Worth Digitizing:**

- Elial Foote Papers, Cushing family correspondence, Albion Tourgee correspondence, Holland Land office survey notebooks (originals) all at McClurg Museum would be great. A diary of a women in Lockport at Niagara County Historical Society (with other women's collections), local sports collections would be a big draw! Also materials on local ethnic groups at places like BECPL, FSSJ, Urban Center, Irish Center, St. Paul's, Catholic Diocese, Buffalo State College's MFRHC, etc.
- Archives of other congregations of women religious, e.g., Sisters of Mercy, Sisters of St. Joseph, Franciscan Sisters of St. Joseph, Felician Sisters, etc.
- The Porter Historical Society also collects local history items, but specializes in the Village of Youngstown and Fort Niagara.
- I have attended a meeting regarding creating a computerized database and GIF mapping of the Pioneer cemetery, the other interested parties do have additional cemetery records. The Barker's Academy records correspond with the Reed Library's holdings possibly. The Barker's collections are similar to other area museums in that they express the early settlers to Chautauqua County, although the Barker's main mission is to collect those items that express the early establishment mainly of the town of Pomfret and Village of Fredonia.
- Seneca-Iroquois National Museum, Iroquois culture and history; Special Archives, Reed Library, SUNY-Fredonia, Iroquois history and genealogy.

## **DIGITIZATION PLAN'S COMPLETE GOALS, OBJECTIVES, AND ACTIVITIES**

This section outlines the two goals identified by the Regional Digital Heritage Advisory Subcommittee as essential to the development of a regional collaborative digital program for the Western New York region:

### **Goal 1: Organizational Infrastructure**

Develop the organizational and human resources within member libraries and library systems and within WNYLRC to create an infrastructure that will support innovative digital services to patrons in WNY.

### **Goal 2: Regional Collaborative Digital Program**

Develop a regional collaborative digital program to enhance the access to and retrieval of digital information resources for member libraries, library systems, and their patrons.

The objectives and activities for the achievement of each goal are presented on the following pages.

## **Goal 1: Organizational Infrastructure**

**Develop the infrastructure of organization and human resources within member libraries and library systems and within WNYLRC to create and support innovative digital services to patrons in WNY.**

**Objective 1.1:** Provide the training and information needed for member libraries and library systems to explore digitizing their own institutional collections.

**Activity:**

- ◇ Implement a training calendar of workshops and facilitated discussions.

**Objective 1.2:** Provide the training and information needed for WNYLRC staff to develop a regional expertise and to oversee and maintain a regional collaborative digital program.

**Activities:**

- ◇ Identify key skills necessary for implementation and maintenance of a WNYLRC-sponsored regional collaborative digital program.
- ◇ Evaluate staffing strengths and areas of improvement needed for implementation and maintenance of a WNYLRC-sponsored regional collaborative digital program.
- ◇ Identify current staff or hire as needed to implement and maintain a WNYLRC-sponsored regional collaborative digital program.
- ◇ Require designated staff to attend professional development and training opportunities.
- ◇ Promote WNYLRC's knowledge and expertise to membership.
- ◇ Provide consulting services to members that are considering or involved in digitization programs. WNYLRC will provide guidance on all aspects of a digitization program.

## Goal 2: Regional Collaborative Digital Program

**Develop a regional collaborative digital program to enhance the access to and retrieval of digital information resources for member libraries, library systems, and their patrons.**

**Objective 2.1:** Encourage member libraries and library systems to develop content-rich collaborative digital projects that meet minimal standards of best practices and that enhance regional digitization efforts by WNYLRC and other institutions as outlined by the regional digitization plan.

### Activities:

- ◇ Provide informational sessions on the contents of the regional digitization plan.
- ◇ Offer presentations from other successful digitization projects that exemplify “best practices” and that encourage member libraries, library systems and other organizations to utilize the recommended standards for imaging, metadata, and selection criteria outlined in the regional digitization plan.
- ◇ Offer “best of breed” presentations of both state-wide and regional collaborative efforts to encourage WNYLRC member libraries and library systems to collaborate with WNYLRC and with other WNYLRC member libraries and library systems to develop a regional digital collection.
- ◇ Fund and implement a digitizing laboratory that includes basic and advanced scanners, computers with software, work space, and WNYLRC staff to maintain it.
- ◇ Develop a training program for member libraries and library systems using the digitizing laboratory to provide basic and advanced skills in creating digitized collections that can be a part of the regional collaborative collection.

**Objective 2.2:** Increase the ability of WNYLRC and its member institutions to support a regional collaborative digital program through the appointment of an advisory group.

### Activity:

- ◇ Appoint a permanent Regional Digital Heritage Advisory Group of individuals knowledgeable in digitization and program planning under the auspices of the Regional Automation Committee to implement the Regional Digitization Plan. This Group will create guidelines for member contribution to the regional collaborative collection, as well as best practice guidelines for formats not included in this document.

## **Goal 2: Development of a Program (continued)**

### **Develop a regional collaborative digital program to enhance the access to and retrieval of digital information resources for member libraries, library systems, and their patrons.**

**Objective 2.3:** Increase the ability of WNYLRC and its member institutions to support a regional collaborative digital program through exploration of funding and support services.

#### **Activities:**

- ◇ Explore funding sources and cost-share models.
- ◇ Revise RBDB Member Grant Guidelines to allow RBDB grant funding for member collaborative digitization projects.
- ◇ Meet with regional library directors to encourage a regional collaborative digital program.
- ◇ Seek out member institutions that can provide server space for hosting digital collections from other institutions.
- ◇ Explore WNYLRC's capability of hosting a server for housing digital collections of smaller institutions.
- ◇ Provide member libraries and library systems with timely information for sources of grant funding available to support the regional collaborative digital program.
- ◇ WNYLRC staff will explore the feasibility to provide fee-based metadata services for member libraries and library systems and other organizations.

**Objective 2.4:** Create a collaborative collection of regional digital content from WNYLRC member institutions that will serve as the model for the regional collaborative digital program.

#### **Activities:**

- ◇ Confirm recommendations of the original guidelines for WNY regional access to digitized collections.
- ◇ Plan a small, ongoing digitization project focused on the collections of smaller institutions to build awareness of and confidence in a collaborative approach to digitization.
- ◇ Develop a test regional database of selected materials from member libraries and library systems with WNYLRC as the lead agent.
- ◇ Recommend the process by which other member libraries and library systems can contribute digital assets to a regional database that is the foundation of the regional collaborative digital program.
- ◇ Provide training in the digitizing laboratory for member libraries and library systems to scan materials and contribute data to the regional database.

## **Goal 2: Development of a Program (continued)**

**Develop a regional collaborative digital program to enhance the access to and retrieval of digital information resources for member libraries, library systems, and their patrons.**

**Objective 2.5:** Develop WNYinfo.org as the open and easily accessible single interface portal for the regional collaborative digital program.

### **Activities:**

- ◇ Enhance patron access to regional digital assets by creating the WNY Digital Heritage Collection via WNYinfo.org.
- ◇ Explore WNYLRC's capability of implementing federated searching software to connect diverse collections via WNYinfo.org.
- ◇ Implement federated searching software to connect diverse collections through the collaborative collection.
- ◇ Recommend appropriate access protocol for member libraries and library systems to contribute to the collaborative collection as a digital collection target through the selected federated search software.

**Objective 2.6:** Identify the long term preservation needs of digital assets for both WNYLRC and member libraries and library systems.

### **Activities:**

- ◇ Explore and identify the methods appropriate for preservation of digital assets and provide training programs on digital preservation beneficial to the region.
- ◇ Develop regional guidelines for digital asset preservation.
- ◇ Once guidelines for digital asset preservation are established, investigate the feasibility of maintaining a regional digital repository for the preservation of digital assets for member institutions participating in the regional collaborative digital program.

## REGIONAL DIGITAL HERITAGE ADVISORY SUBCOMMITTEE MEMBERS

The following people are current and past members of the Regional Digital Heritage Advisory (RDHA) Subcommittee who helped to influence this digitization plan.

<b>Name</b>	<b>Institution</b>	<b>Representing</b>
Toniann Scime*	Amherst Museum	Member at Large
Laura Schiefer*	Buffalo & Erie County Historical Society	Special Libraries
Mary Ann Tingley*	Buffalo and Erie County Public Library	Public Library
Elise Torre	Buffalo General Hospital	Regional Automation Committee
Musa Abdul Hakim	Buffalo State College	Continuing Education Committee
Lynnette Mende*	Erie Community College/North	Resource Sharing Committee
James M. Tammaro*	New York State Archives	
David Schoen*	Niagara University	Preservation Committee
Heather Glogowski*	Nichols School	School Library
Pamela Jones, Chair*	Roswell Park Cancer Institute	Regional Automation Committee
Ann Tenglund*	Saint Bonaventure University	Private Academic Library
Dean Hendrix	University at Buffalo	Continuing Education Committee
John Bewley*	University at Buffalo	University at Buffalo
Heidi Bamford*	WNYLRC	Documentary Heritage Program
Sheryl Knab*	WNYLRC	WNYLRC
Robin Tashjian	WNYLRC	WNYLRC

The RDHA Subcommittee hired a digitization consultant to work with it on this regional digitization plan. That consultant was Jill Hurst-Wahl of Hurst Associates, Ltd.

\* Committee member as of March 31, 2006.