Report of the

Task Force on Developing Digital Collections for the MU Libraries

Members: Anne Barker Felicity Dykas Kelli Hansen Michael Holland, Co-chair Anselm Huelsbergen Roger Knouff Wayne Sanders Geoff Swindells, Co-chair

June 22, 2006

Contents

Introduction

- Opening Statement
- Charge and Purpose
- Definition and Scope

Executive summary

Recommendations

- Administration and Sustainability
- Collection Development and Selection
- Digitization Initiatives
- Organization and Discovery
- Preservation

Appendices

- A. Checklist for Digital Resource Selection
- B. Prioritizing Matrix for Resource Selection
- C. Digitization Standards and Considerations
- D. Digital Rights Resources
- E. Guidelines for Discovery Tool Selection for MU Digital Collection Resources
- F. Metadata Schemes and Standards for Cataloging and Description
- G. Digital Integration Matrix
- H. Possible Grant Funding Agencies
- I. MU Libraries Digitization Projects in Process and Completed
- J. Report by the University of Missouri Libraries Task Force on Imaging, September 23, 1999

Introduction

Supplying informational support for scholarship and research is becoming increasingly complex with the wide adoption of and dependence on a range of digital technologies. Purchasing a physical object regulated under the traditional copyright provisions of first sale and fair use is much less nuanced than acquiring access to digital resources or creating and integrating digital objects into the collection. No longer can an academic library support the intellectual and research needs of its constituency with only print and CD-ROM based resources. New technologies require that new and fresh approaches to resource creation, selection, acquisition, and delivery be supported by a traditional foundation of professional values and ethics.

With the variety of sources and delivery methods available for digital information, there is wide-spread confusion regarding responsibility for adding new digital materials to holdings in academic and research libraries. The Task Force on Developing Digital Collections for the MU Libraries was primarily created to suggest to the Director of the MU Libraries an organizational foundation that would make the building of digital collections from a number of sources uniform, integrated, and sustainable. We thus have offered recommendations for organizational decision-making that may advance the MU Libraries to integrate the acquisition of digital access with library-produced and University-created digital content, including the creation of an institutional repository. We have also devoted space to recommending a mechanism for advancing and regulating in-house digital projects so that the products become an integral part of MU Libraries' digital resources.

The MU Libraries' Mission (draft, April 2006) states, "In support of the instruction, research, service and economic development programs of the University, the MU Libraries acquire scholarly resources, develop innovative services, and apply new information technologies to achieve a single overarching purpose: to address the needs of our users." The expectation of our users for immediate access to digital resources, whether created locally or purchased from others, requires that the Libraries give highest priority to developing and maintaining the organizational, technological, financial, and intellectual bases needed for agile response to evolving demands.

Task Force Charge and Purpose:

- 1. Help the MU Libraries establish an organizational foundation and decision-making process for its digital collections.
- 2. Outline a process for implementation of cooperative projects with entities outside of the MU Libraries and beyond the Columbia campus.
- 3. Encourage and support the formation of interest and project groups as well as expertise within the MU Libraries with regard to developing digital collections.
- 4. Make recommendations to the Director of Libraries for appropriate next steps.

MU Libraries Digital Collection - Definition and Scope:

For the purposes of this report, the "MU Libraries Digital Collection" consists of all digital objects managed by the Libraries as part of its mission of supporting the scholarly information needs of the students, faculty and staff of the University of Missouri-Columbia. This includes networked resources, as well as tangible electronic media, and digital materials in all formats, such as text, graphics, spreadsheets, sound files, etc.

Within the MU Libraries Digital Collection, the Task Force finds it useful to distinguish among five broad classes of digital objects:

- 1. Purchased or licensed resources.
- 2. Resources that have been reformatted (digitized) by, or on behalf of, the MU Libraries.
- 3. Resources harvested by, or deposited with, University Archives or Special Collections, as part of their respective missions.
- 4. All other resources for which the MU Libraries have accepted responsibility for long-term management, and continuing access.
- 5. Freely available Internet resources represented in the MU Libraries' catalogs, databases, and web sites.

Executive Summary

Over the past few years, the acquisition of electronic information and the creation of local digital resources have made a significant impact on the way information is delivered to our users. However, to date, these efforts have not been sufficiently attentive to the kinds of organizational and financial commitments necessary for the Libraries to fulfill its proper role as a trusted repository of quality scholarly resources in digital form.

Four areas, in particular, require our attention: the need for leadership and organizational accountability for digital preservation; making sure that the costs and capacities necessary for digital preservation and metadata creation are considered as part of the collection development process; making digitization initiatives an integral part of our collection-building activities; and ensuring that Libraries have the knowledge and skills needed to manage digital collections in a rapidly changing technological environment.

We need a more holistic approach to building and managing our digital collections, one that commits adequate institutional resources to ensure that these collections will be sustainable and accessible over time. We need clear lines of authority, explicit rules, accountability mechanisms, and transparent work flows. Finally, we need to make sure that the development of our digital collections takes place in the context of a comprehensive collection management plan, a plan focused on the present and future needs of students and scholars.

The development of digital collections must be guided by the MU Libraries' general mission to support the teaching, research, outreach, and service activities of the University. In serving these goals and in attention to quality, usefulness, and significance, building digital collections is similar to building any collection and should not be isolated from more general collection activities. However, the digital aspect introduces some distinctive characteristics of availability, functionality, and responsibility that must be given special consideration.

The MU Libraries' unique, rare and valuable tangible collections are what set us apart from other libraries and cultural institutions. They are our legacy and among our most important assets. Selective digitization of these collections is an important part of the Libraries future. Most obviously, digital reformatting can enhance the value of existing collections by making them more accessible, better integrated with the Libraries' other digital collections, and more likely to be used. They can also help support current or emerging research or instruction in academic units across campus; promote collaboration between internal and external partners; and help preserve rare and endangered materials.

Moreover, the Task Force is convinced that well thought out and executed digitization initiatives can contribute to the national reputation of the Libraries and will provide benefits to the Libraries that far outweigh the costs involved. To that end, we believe that the Libraries need to adopt a much more aggressive, coordinated and forward-looking approach to the digital reformatting of its collections. That said, there are costs, and building viable, sustainable digital collections will require the infusion of scarce funds. This means that in addition to creating a culture of innovation in the Libraries, we need a mechanism to ensure that the Libraries are able to maintain digital projects over time.

The value of a digital collection is realized when the resources are organized in such a way that users can discover and access them. The four user tasks defined in the Functional Requirements for Bibliographic Records: Final Report (1998) -- find, identify, select, and obtain -- are accomplished through organization and discovery. Organization and discovery will be achieved through resource description (creation and maintenance of bibliographic records and related metadata) and the use of systems that support desired

levels of indexing, searching, retrieval, and display. The use of multiple discovery systems and metadata schemes will ensure that the unique elements of various types of resources will be targeted and that the needs of diverse users will be met. Recognized international standards must be adopted to ensure permanence and portability of the data, and compatibility among different metadata schemes and systems. Ongoing commitment to the allocation of resources (staff, equipment, and training) will be critical.

The MU Libraries must support digital preservation (i.e., the maintenance and continued accessibility) of its valued digital objects so that the original form, content, and function of each object can be recreated. The extent of recreation of form, content, and function will depend on the actual preservation means, technological and fiscal, available for each digital object. The methods employed by the Libraries to provide maintenance of and continued accessibility to those valued digital objects under its purview must take into account continuing changes in technology.

This document provides specific recommendations which address administration and sustainability, collection development and selection of resources, local digitization initiatives, organization and discovery of resources, and preservation. Appendices provide proposals for guidelines and references to additional resources in these areas.

Task Force Recommendations

Administration and Sustainability:

Over the past few years, the acquisition of electronic information and the creation of local digital resources have made a significant impact on the way information is delivered to our users. However, to date, these efforts have not been sufficiently attentive to the kinds of organizational and financial commitments necessary for the Libraries to fulfill its proper role as a trusted repository of quality scholarly resources in digital form.

Four areas, in particular, require our attention: the need for leadership and organizational accountability for the development and preservation of our digital collections; making sure that the costs and capacities necessary for digital preservation and metadata creation are considered as part of the collection development process; making digitization initiatives an integral part of our collection-building activities; and ensuring that Libraries have the knowledge and skills needed to manage digital collections in a rapidly changing technological environment.

We need a more holistic approach to building and managing our digital collections, one that commits adequate institutional resources to ensure that these collections will be sustainable and accessible over time. We need clear lines of authority, explicit rules, accountability mechanisms, and transparent work flows. Finally, we need to make sure that the development of our digital collections takes place in the context of a comprehensive collection management plan, a plan focused on the present and future needs of students and scholars.

Recommendations:

- 1. The Director of Libraries should assign responsibility for the implementation of long-term strategies for the preservation of digital content to the Collection Development Librarian for subscribed or purchased materials.
- 2. The Director of Libraries should assign responsibility for the implementation of long-term strategies for the preservation of digital content to the Head of SCARaB for materials which have been locally digitized.
- 3. The Director of Libraries should include, as part of the charge for the new Collection Development Librarian, the responsibility for administrative oversight, fiscal management, and overall coordination of the Libraries digital collection management efforts. This coordination will include establishing mechanisms to ensure that metadata and preservation considerations are addressed as part of the collection development process.
- 4. The Libraries should devote 10% of the Libraries collections budget to local and cooperative digitization efforts. These monies will provide seed-money, supplement grant and gift funding, and help provide for the long-term sustainability of digital projects.
- 5. The Director of Libraries should commission the Collection Development Librarian to identify, in consultation with appropriate stakeholders, the knowledge and skills needed to carry out the recommendations outlined in this report, assess the Libraries' current capabilities, support training and continuing education activities to help meet unmet needs, and where appropriate, secure funding for any new positions that might be necessary to achieve these goals.

Collection Development and Selection:

The development of digital collections is guided by the Libraries' general mission to support the teaching, research, outreach, and service activities of the University. In serving these goals and in attention to quality, usefulness, and significance, building digital collections is similar to building any collection and should not be isolated from more general collection activities. However, the digital aspect introduces some distinctive characteristics of availability, functionality, and responsibility that must be given special consideration.

- 1. Availability.
 - a. The digital format erases the distinction between local and global. By digitizing local holdings, we can make them available to a global audience. By pointing to digital holdings elsewhere, we bring global resources home to our local users. As the range of materials from which to select expands, we have the potential to be producers as well as mediators or suppliers.
 - b. The digital format can unlock information otherwise unavailable by providing full-text indexing, enhancing text with additional features, and allowing manipulation of materials otherwise too fragile for handling. Materials previously neglected may become more interesting or significant in the digital environment.
- 2. Functionality.
 - a. In addition to evaluating content, selection criteria must include evaluation of the functionality of digital objects, including accessibility, security, and sustainability.
 - b. Building usable digital collections involves selecting not only content, but technologies that combine disparate parts into meaningful collections.
- 3. Responsibility.
 - a. "Collecting" may range from pointing temporarily at a remote resource to digitally preserving and permanently maintaining materials. Selection decisions must include designation of the appropriate level of responsibility/commitment. See the "digital collecting levels" at <u>http://sunsite.berkeley.edu/Admin/collection.html</u>.
 - b. The various levels of digital collecting involve the commitment of funds and personnel, which must be taken into consideration as selection criteria.

Recommendations:

- 1. Digital collection decisions should not be made in isolation from other collection management considerations. Therefore there must be direct linkage and reciprocal representation/communication between the Collection Development Librarian and the Committee on Digitization Initiatives as indicated in the "Administration and Sustainability" section of the report.
- 2. Digitizing materials requires a high level of expertise in evaluating potential projects and completing their execution. Therefore, the Director of Libraries should appoint a Committee on Digitization Initiatives with responsibility for prioritizing and planning local digitization projects and for guiding the MU Libraries' participation in joint projects with other institutions.
- 3. Digitizing local materials puts us on a global stage. This requires the balancing of local needs and the national/international mission. It may open entrepreneurial possibilities or enhance the reputation of the University. We recommend that the Committee on Digitization Initiatives

evaluate local holdings using the OCLC collection analysis software or equivalent to identify unique materials and identify national/international projects that best leverage our local resources.

- 4. The selection of digital materials or materials to digitize involves the weighing of many factors. We recommend that the Committee on Digitization Initiatives and individual selectors use the **Checklist for Digital Resource Selection** (**Appendix A**) as a mechanism for guiding and documenting this process.
- 5. Many potential projects/collections will be competing for limited funds and personnel resources. We recommend that project proposals be documented, evaluated, and prioritized on a long-term basis, possibly 5 years, so that a cohesive collection can be built over time. In creating collection development policy statements and goals, selectors and the Collection Development Committee should identify useful digital resources, both existing and to be developed. Tools to aid the planning and evaluation process are included in **Appendix G Digital Integration Matrix**
- 6. The establishment of digital collections involves the efforts of almost every department in the library. We recommend that the Collection Development Committee create a documented, standardized procedure for recommending and implementing projects to insure that all necessary information is communicated to all parties involved in a timely fashion.
- 7. The development of digital collections involves not only the selection of materials, but also the selection or development of technologies that gather and organize materials. The Libraries are well positioned to test and evaluate technologies in development. We recommend that the MU Libraries Information and Technology Committee maintain communication and partner with other University units that are developing such technologies.

Digitization Initiatives:

The MU Libraries' unique, rare and valuable tangible collections are what set us apart from other libraries and cultural institutions. They are our legacy and among our most important assets. Selective digitization of these collections is an important part of the Libraries' future. Most obviously, digital reformatting can enhance the value of existing collections by making them more accessible, better integrated with the Libraries' other digital collections, and more likely to be used. They can also help support current or emerging research or instruction in academic units across campus; promote collaboration between internal and external partners; and help preserve rare and endangered materials.

Moreover, the Task Force is convinced that thoughtfully executed digitization initiatives can contribute to the national reputation of the Libraries and will provide benefits to the Libraries that far outweigh the costs involved. To that end, we believe that the Libraries need to adopt a much more aggressive, coordinated and forward-looking approach to the digital reformatting of its collections. That said, there are costs, and building viable, sustainable digital collections will require the infusion of scarce funds. This means that in addition to creating a culture of innovation in the Libraries, we need a mechanism to ensure that the Libraries are able to maintain digital projects over time.

Recommendations:

1. The MU Director of Libraries should charge the Head of Special Collections, Archives & Rare Books (SCARaB) with actively pursuing collaborative digitization projects, and with

responsibility for the implementation of long-term strategies for the preservation of these projects.

- 2. The MU Director of Libraries should establish a standing Committee on Digitization Initiatives that will be responsible for operational matters and coordination related to digitization, including the evaluation and approval of all digitization proposals. This committee should be chaired by the Head of Special Collections, Archives and Rare Books (SCARaB) and should include representatives from the Catalog Department, Library Technology Services, the Collection Development Librarian, and when invited by the Committee, the director of the digital library program of the UM Library Systems Office in an *ex officio* capacity. Coordination will include establishing mechanisms to ensure that metadata and preservation considerations are addressed as part of the collection development process.
- 3. The MU Libraries should subscribe to the OCLC WorldCat Collection Analysis service, or another collection analysis service provider, and use that tool to identify unique holdings that would benefit from future digitization efforts.
- 4. The MU Libraries should follow appropriate standards and best practices for the creation, description, and preservation of digital masters and use-copies.
- The MU Libraries' Catalog Department should register all digitally reformatted materials in the DLF/OCLC Registry of Digital Masters (<u>http://www.oclc.org/digitalpreservation/why/digitalregistry/</u>) and other appropriate registries, such as the Registry of U.S. Government Publication Digitization Projects (<u>http://www.gpoaccess.gov/legacy/registry/</u>).
- 6. The MU Libraries should become active contributors to the UM Digital Library and active participants in the Missouri Digitization Planning Project (http://www.virtuallymissouri.org/about.aspx).
- 7. The MU Libraries should become an active, contributing member of the Open Content Alliance (<u>http://www.opencontentalliance.org/</u>).

Organization and Discovery:

The value of a digital collection is only realized when the resources are organized in such a way that users can discover and access them. The four user tasks defined in the *Functional Requirements for Bibliographic Records: Final Report* (1998) -- find, identify, select, and obtain -- will be accomplished through organization and discovery.

Organization and discovery will be achieved through resource description (creation and maintenance of bibliographic records and related metadata) and the use of systems that support desired levels of indexing, searching, retrieval, and display. The use of multiple discovery systems and metadata schemes will ensure that the unique elements of various types of resources will be targeted and that the needs of diverse users will be met.

Recognized international standards will be adopted to ensure permanence and portability of the data, and compatibility among different metadata schemes and systems. Ongoing commitment to the allocation of resources (staff, equipment, and training) will be critical. Work in this area will be informed by the work of several groups, including but not limited to the Web Advisory Group, the Federated Search Engine

Task Force, the Link Resolver Task Force, the Content Management System Task Force, and the MERLIN Quality Control Committee (MQCC).

Recommendations:

1. A variety of distinct and overlapping delivery tools are currently in place at the MU Libraries. These include the MERLIN Library Catalog, Internet finding aids in Special Collections, Archives and Rare Books, an A-to-Z list for electronic journals, a database of online abstract and index databases maintained by Ellis Reference, and a database of online resources maintained by the Health Sciences Library.

Coordinating efforts to determine, develop, and maintain discovery tools will serve our users and improve efficiencies in the work of the MU Libraries. We recommend the appointment of a task force (Organization of Library Resources Task Force) which will be charged with determining the optimum delivery system(s) for various types of materials and for determining the staff or library units best suited for the deployment, and ongoing oversight and maintenance of these systems. The task force membership should include members but not be limited to members of the of MU Libraries Catalog Department, Health Sciences Library (HSL) Technical Services, Ellis Reference, HSL and/or branch libraries public services, MQCC, and MERLIN Collection Development Reference Services Committee (MCDRSC).

See Guidelines for Discovery Tool Selection for MU Digital Collection Resources, (Appendix E), for descriptions of possible discovery tools.

2. It is metadata that allows resources to be found, collocated, identified, and located using discovery tools. We recommend that the MU Libraries use descriptive, structural, and administrative metadata and encoding schemes that are internationally recognized and appropriate to the type of material being described. This will ensure interoperability, portability, and permanence of data.

We recommend that the Organization of Library Resources Task Force be charged with developing guidelines for the choice of metadata schemes and elements to be included for various types of digital objects and used in different discovery tools. The guidelines should balance the desire for rich, detailed metadata with current and future needs and resources, should include criteria for determining and applying appropriate levels of description and granularity to best meet user needs, and should address the need for controlled vocabulary which will enhance findability and precision in retrieval and should follow internationally recognized thesauri.

3. Locally digitized material will require specialized treatment, including the creation of embedded metadata to enhance findability and support the ability of other institutions to harvest information about the objects, and the creation of more-detailed administrative and structural metadata than other digital resources.

We recommend that the Committee on Digitization Initiatives, as conceived in the "Administration and Sustainability" section of the report, follow guidelines developed by the Organization of Library Resources Task Force and address the specific requirements of locally digitized materials.

4. The MU Libraries' website should be the gateway to our digital collections. It should be designed for the convenience of the user, should integrate information about resources and services for all MU Libraries' sites, and should be designed for efficient and distributed

development and maintenance. To realize this integration, we recommend that a library-oriented content management system be purchased.

- 5. To ensure continued usefulness and success, we recommend that resources be allocated to ensure ongoing maintenance and development of metadata and discovery tools as recommended by the Organization of Library Resources Task Force.
- 6. We urge ongoing evaluation of metadata and tools, recommended by the Organization of Library Resources Task Force, be conducted to ensure their continued usefulness and adequacy.
- 7. Our expertise in the development, maintenance, and use of discovery tools and metadata gives us much to contribute to an MU institutional repository. We recommend that resources be allocated to ensure that the MU Libraries have a role in such a venture.

Preservation:

The MU Libraries supports digital preservation (i.e., the maintenance and continued accessibility) of its valued digital objects so that the original form, content, and function of each object can be recreated. The extent of recreation of form, content, and function will depend on the actual preservation means, technological and fiscal, available for each digital object. Regardless of the distinction between digital objects controlled and not controlled by the MU Libraries, consideration shall be given to digital preservation (e.g., giving preference to license agreements for explicit preservation strategies and continued access). It should be noted, however, that the methods employed by the MU Libraries to provide maintenance of and continued accessibility to those valued digital objects under its purview should continue to take into account changes in technology.

In carrying out responsibilities for preservation, the Collection Development Librarian and the Head of Special Collections, Archives & Rare Books (SCARaB) should base preservation efforts on an assessment of the digital object's value.

Recommendations:

- 1. **[NOTE**: the following priority levels are taken from the Cornell University Library's 2004 Digital Preservation Policy Framework (<u>http://commondepository.library.cornell.edu/docs/cul-dp-framework-0405_main.pdf</u>)]
 - a. Priority 1: born digital materials Rigorous effort will be made to ensure preservation in perpetuity of material selected for preservation, both library resources and institutional records.
 - b. Priority 2: digitized materials (no available analog) Every reasonable step will be taken to preserve materials without a print analog, when redigitizing is not possible or no analog versions are located elsewhere. Also included are digitized materials that have annotations or other value-added features making them difficult or impossible to recreate.
 - c. Priority 3: digitized materials (available analog) Reasonable measures will be taken to extend the life of the digital objects with a readily available print analog. However, the cost of redigitizing will be weighed against the cost of preserving the existing digital objects.
 - d. Priority 4: items and other materials No preservation steps will be taken for ephemeral materials such as materials scanned for E-reserve and Document Delivery, odds and ends

of collections, portions of text, and content that is deemed nonessential to the comprehensiveness of collections.

- 2. The following procedures will be applied relative to the above priority levels:
 - a. periodic migration of digital objects as warranted by changes or obsolescence of the objects' underlying technological format (e.g., JPEG), software (e.g., SPSS), or hardware (e.g., 5 ¹/₂" floppy drives).
 - b. migration includes
 - i. refreshing (copying a digital object from storage medium to storage medium)
 - ii. forward migration (moving from one file specification standard or software application to another)
 - iii. choosing standard formats (e.g., TIFF, XML) for those digital objects that come under the preservation responsibility of the MU Libraries. The intent of this is to reduce the number of formats and platforms that need attention and support.
 Format standards should also determine forward migration directions of digital objects whose original format or underlying application has become obsolete.
 - c. applying redundancy to the storage of digital objects through backups (e.g., on-line, nearline, off-line, on-site, off-site) or other formalized redundancy schemes (e.g., Lots of Copies Keep Stuff Safe [LOCKSS], Portico).
- 3. The MU Libraries' Collection Development Committee in acquiring digital resources of which the creation and maintenance are not within its purview, should give preference to:
 - a. license agreements that provide for
 - i. continuing access to material previously subscribed to after termination of the license
 - ii. long-term preservation and access
 - b. public domain and open access content that includes an explicit preservation strategy most in line with that of the MU Libraries.
- 4. The MU Director of Libraries should play an integral part in establishing and managing an institutional repository for the University of Missouri's Columbia campus.
 - a. establishing an institutional repository or using third-party resources to those ends needs to include application of the Open Archival Information System Reference Model as well as subsequent standards based on the OAIS Reference Model such as the Producer-Archive Interface Methodology Abstract Standard (2004).

Task Force Conclusion

New technologies require new and sophisticated approaches to creating, selecting, acquiring, preserving, delivering, and managing sufficient informational resources to support the University academic and research community. The members of this Task Force hope that this document will assist in directing the MU Libraries to the next step in developing viable, important, and integrated digital collections.

Report Appendices:

Appendix A. Checklist for Digital Resource Selection

1. Project Criteria

a. Does the resource meet the following basic criteria?

- The digital resource should be easy to use, requiring minimum training.
- Documentation must be clear.
- The digital resource should provide convenient output to printers and/or users' files.
- The digital resource should be accessible to people with disabilities.
- The authority and reliability of the resource must be certified in the digital environment.
- The digital format must be stable and durable.

The resource should operate on equipment and operating systems either currently or expected-to-be available.

b. If the resource is to be purchased or licensed:

- Access to the digital product should be appropriately controlled.
- The integrity of the digital collection should be secure.
- The privacy and security of patron/data activity should be guaranteed.

c. If the resource is to be digitized, MU must have the legal right to make and disseminate copies. What is the copyright status of the materials to be digitized? What problems, if any, do you foresee?

	*
	Þ
d. Comments:	
4	Þ

2. Purpose

a. Access and Usability

1. Will the digital resource enhance access or usability of the materials?



3. Will the resource attract new users to the libraries?

4. Does the resource enhance strengths of the libraries' collections? Will it supplement existing digital collections, either at MU or elsewhere?



b. Preservation

1. Does the resource preserve rare materials? Could it reduce the cost of maintaining or providing physical access to heavily used materials?



2. If materials from MU's collections are to be digitized, what will be the effect of digitization on the physical object?



3. Labor, Infrastructure, and Maintenance

a. What are the specific technological requirements of this resource (operating systems, hardware, maintenance, staff resources for programming, user interface design, search engine development, etc.)?



b. Do the MU Libraries have sufficient expertise and technological resources to assure that the resource can fulfill the functions for which it is intended?



c. Will the resource attract or promote further acquisitions of digital materials (through digitization or purchase)? Could it contribute to collaborative or consortial efforts or to partnership with a commercial provider?

4

d. Will the resource increase demand on public services, security, or access services?

e. How is continuing access to be provided for this resource?

	<u>*</u>
4	Þ

4. Organization and Discovery Methods

What staff and equipment resources will be needed to provide access to this resource? Refer to Appendix E, Guidelines for Discovery Tool Selection for MU Digital Collection Resources

	-
4	Þ

5. Funding

a. Consider the total one-time and ongoing costs of digitization, metadata, cataloging, preservation/archiving, potential migration to developing platforms, disposition of originals. What are the specific funding requirements for this resource?

-

b. Could the resource attract funding or generate income through marketing or development? Can external funding be secured to support the proposed resource?

	b.
c. Could an acceptable product be created at a lower cost?	
	-
	-
	E.

Definitions of terms and questions used in Appendix A - Checklist for Digital Resource Selection

Documentation

• What is meant by documentation?

Authority and Reliability

• What is meant by authority and reliability?

Integrity

• What is meant by integrity?

Copyright Status

- In the case of materials to be digitized, MU must have a legal right to make and disseminate copies, and projects must comply with the <u>UM System's copyright and fair</u> <u>use policy</u>. At least one of the three conditions below must be satisfied.
 - The original materials or collection must be in the <u>public domain</u>
 - The purpose of the project must fall under <u>fair use</u>
 - <u>Permission</u> to make and disseminate copies must be secured from the copyright holder

Enhanced Access and Usability

- Would the resource reinstate material into circulation that up to now has been difficult to access?
- Would the resource make items of historical or intellectual value more widely accessible to university patrons?
- With appropriate metadata and systems, would the resource make materials easier to locate and access?
- Would digital surrogates add functionality to the way the materials are used?

High Priority Activities

- Are the materials in high or medium demand?
- Will the resource increase availability of material with direct relevance to teaching?
- Will the resource satisfy existing requests for digital surrogates from users?
- Is the project in keeping with policies and goals at an institutional level?

Attracting New Users

- Would the resource make items of historical or intellectual value more widely accessible to patrons outside the university?
- Would the resource increase demand for or interest in materials in the collections which up to now have been relatively ignored?

Enhancing Strengths

- Is the subject area a goal of the current collection development policy?
- Has this subject area been ignored in previous digital projects?

Preserving Rare Materials

- Would the resource allow access to materials which have previously been restricted due to preservation concerns?
- Would the resource provide a surrogate of high enough quality that requests to handle, retrieve, or photocopy the physical materials will be significantly decreased?

Effects of Digitization

- Can digitization be carried out without damage to the materials?
- Are the materials at risk of disintegration (acetate, newspaper, etc.)?
- In the case of rare materials, digitization from a photographic intermediary should be considered to avoid damage to the original object.
- Determine whether the physical object will need extra security after digitization. Should it be sent to remote storage, Special Collections, or designated for library use only?

Expertise and Technological Resources

- Are hardware, software, and storage appropriate to the digitization project or digital resource available either in-house or through a third party?
- Is there sufficient hardware and software in place to allow access to the digitized collection?

Increased Demand

- Would the resource generate new interest in collections that have previously been ignored?
- Would the resource create new interest in materials that are vulnerable to theft or damage?
- Would requests to see materials "in person" tax staff resources in Access Services, Public Services, Special Collections, or Archives?

Continuing Access

- In the case of licensed resources, the license should allow for fair use and for the usual library functions provided for in copyright law, e.g. interlibrary loan, preservation/maintaining, reserve use, and on-site use by guests. For digital objects not under the Libraries' control these needs can be met in a number of ways:
 - By preferring license agreements (similar to <u>GWLA's model license agreement</u>) that provide for
 - continuing access to material previously subscribed to after termination of the license
 - long-term preservation and archival access
 - By preferring public domain and open access content that includes an explicit preservation strategy, such as that provided by the <u>Open Content Alliance.</u>
 - For digital objects that are under the Libraries' control there are a <u>number of options</u> the Libraries should consider, among them participation in the <u>LOCKSS</u> <u>project</u>, where cooperating publishers give permission for libraries to collect and preserve digital content.
- For digitized materials, continuing access will need to be planned for in terms of maintenance, hardware and software upgrades, staffing, etc.

Appendix B. Prioritizing Matrix for Resource Selection

Using the information provided in the Checklist for Digital Resources, each project is scored in the areas of purpose, labor/infrastructure/maintenance, and cost/funding, using a scale of 1 to 10. This generates a ratio of purpose/value to labor/cost that may be useful in comparing and prioritizing projects.

Project	Purpose/Value	Labor/Infrastructure	Cost/Funding	Score
Audio-Prose	8	5	5	8:10=80%
Library				
Early American	8	4	8	8:12=67%
Newspapers				

In the first example above, a proposal to digitize the American Audio Prose Library from the cassette tapes held by the Libraries has a high score for purpose/value because it enhances access to materials held by the Libraries, has the potential to attract users, and provides an opportunity for collaboration with several entities at MU, in the UM system, and statewide. It has a medium score for labor because staff time will be involved, but it may be possible to collaborate by using the services of the Marr Sound Archive at UMKC. The score for cost is medium because there may be some cost involved in negotiating the copyright and in purchasing equipment, but this may be offset by grant funding. This results in a ratio of 8 to 10, or 80% correlation between benefit and cost.

In the second example, a proposed purchase of the Early American Newspapers online has a high score for value because it enhances access to materials that are of interest to users across many disciplines. It also has the potential of attracting users to the Libraries. Since this is a purchased database, the labor/infrastructure score is lower, but the cost/funding score is higher, although this is mitigated somewhat by the potential to attract a donor. (The immediate presence of a willing donor might lower this score even more.) The resulting score is a ratio of 8 to 12, or 67%.

Final decisions about the priority of one project over another need not be based solely on this ratio, but this may provide a method for moderating discussion of the relative merits of various proposals.

This scoring matrix is based in part on the section "Putting It All Together: How to Score and Rank Collections" from The Handbook for Digital Projects. See <u>http://www.nedcc.org/digital/iv.htm#score</u>

Appendix C. Digitization Standards and Considerations

Digital Library Federation:

- Benchmark for Faithful Digital Reproductions of Monographs and Serials.
 - Version 1 December 2002
 - The Digital Library Federation Benchmark Working Group (2001-2002)
 - o HTML http://www.diglib.org/standards/bmarkfin.htm
 - PDF <u>http://www.diglib.org/standards/bmarkfin.pdf</u>

California Digital Library:

- Digital Library Building Blocks Best Practices and Standards (<u>http://www.cdlib.org/inside/diglib/</u>)
- CDL guidelines for digital objects, version 2.0 11-2005 (<u>http://www.cdlib.org/inside/diglib/guidelines/cdl_gdo_v2_draft.pdf</u>)
- CDL guidelines for digital images, version 2.0 11-2005 (<u>http://www.cdlib.org/inside/diglib/guidelines/bpgimages/cdl_gdi_v2.pdf</u>)

Library of Congress:

• Digital Formats for Content Reproductions (<u>http://memory.loc.gov/ammem/formats.html</u>)

Council on Library and Information Resources:

- Publication 137 Capturing Analog Sound for Digital Preservation: Report of a Roundtable Discussion of Best Practices for Transferring Analog Discs and Tapes. March 2006. (http://www.clir.org/pubs/reports/pub137/pub137.pdf)
- Publication 112 Copyright Issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment. January 2003. (<u>http://www.clir.org/pubs/reports/pub112/pub112.pdf</u>)

National Archives and Records Administration:

- Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files - Raster Images (http://www.archives.gov/research/arc/preservation/digitizing-archival-materials.pdf) or
- (http://www.archives.gov/research/arc/digitizing-archival-materials.html)

International Association of Sound and Audiovisual Archives - Technical Committee:

• IASA-TC04 -- Guidelines on the production and preservation of digital audio objects: standards, recommended practices, and strategies. 2004

Appendix D. Digital Rights Resources

- Digital Rights, Copyright, Trademark, and Patent Laws http://www.firstgov.gov/webcontent/reqs_bestpractices/laws_regs/copyright.shtml
- Digital Millennium Copyright Act Archive (Electronic Frontier Foundation) <u>http://www.eff.org/IP/DMCA/</u>
- Copyright (U.S.): Law and Policy http://www.copyright.gov/laws/
- U.S. Code collection: Title 17—Copyrights (Cornell Law School) http://www4.law.cornell.edu/uscode/html/uscode17/usc_sup_01_17.html
- Information Policy: Copyright and Intellectual Property (IFLA) http://www.ifla.org/II/cpyright.htm
- Copyright & intellectual property (American Library Association) <u>http://www.arl.org/arl/programs.html#copyright</u>
- U.S. Code collection: Title 17—Copyrights (Cornell Law School) http://www4.law.cornell.edu/uscode/html/uscode17/usc_sup_01_17.html

Council on Library and Information Resources Publication 134 - Acquiring Copyright Permission to Digitize and Provide Open Access to Books. October 2005. (http://www.clir.org/pubs/reports/pub134/pub134col.pdf)

Copyright and fair use (Stanford University Libraries) <u>http://fairuse.stanford.edu/</u>

Locating U.S. copyright holders (Harry Ransom Humanities Research Center at the University of Texas at Austin). Watch – Writers, Artists, and Their Copyright Holders) http://tyler.hrc.utexas.edu/us.cfm

Copyright crash course (University of Texas) http://www.utsystem.edu/ogc/intellectualproperty/cprtindx.htm#top

Copyright issues and intellectual property rights (i-DLR – Interactive Digital Library Resources Information System, School of Library and Information Science, University of Missouri—Columbia) <u>http://www.coe.missouri.edu/~DL/iDLR/index.php?cid=28</u>

- Copyown: a resource on copyright ownership for the higher education community <u>http://www.nethics.umd.edu/copyown/</u>
- WIPO Intellectual Property Handbook: Policy, Law and Use (World Intellectual Property Organization) <u>http://www.wipo.int/about-ip/en/iprm/index.htm</u>

Open Access Bibliography: Liberating Scholarly Literature with E-Prints and Open Access Journals. Charles W. Bailey, Jr.

http://www.escholarlypub.com/oab/oab.htm

Licensingmodels.com: model standard licenses for use by publishers, librarians and subscription agents for electronic resources (sponsored by and developed in close co-operation with four major subscription agents: EBSCO, Harrassowitz, RoweCom, and Swets Blackwell)

http://www.licensingmodels.com/

Licensing electronic resources: strategic and practical considerations for signing electronic information delivery agreements. Prepared by Patricia Brennan, Karen Hersey, Georgia Harper (Includes bibliography to additional resources)

http://www.arl.org/scomm/licensing/licbooklet.html

Liblicense: Licensing digital information: a resource for librarians (Yale University Library, Council on Library & Information Resources)

http://www.library.yale.edu/%7Ellicense/index.shtml

The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials: IV. Rights Management http://www.nyu.edu/its/humanities/ninchguide/IV/

Electronic resource management: report of the DLF ERM initiative. Timothy D. Jewell, et al. 2004 http://www.diglib.org/pubs/dlfermi0408/

Statements and Documents of the International Coalition of Library Consortia (ICOLC) <u>http://www.library.yale.edu/consortia/statementsanddocuments.html</u>

	Web list(s)	Local database(s)	Catalog	External database(s)
RESOURCE				
Purchased or freely	Selected purchased	Selected purchased	All purchased	
available or locally	Selected locally digitized	Selected locally digitized	All locally digitized	
digitized	Selected freely available	Selected freely available	Selected freely available	
Persistence of resource	Ephemeral/trial	medium – long-term	medium – long-term	
Types of resources	Materials freely available on	Subject guides	E-books	Locally digitized
(Examples)	the Web	E-journals	[E-journals]	material
	Trial subscriptions	Locally digitized materials	Locally digitized materials	
		Subscription/licensed materials	Subscription/licensed materials	
		Materials freely available on	Materials freely available on	
		the Web	the Web	
		Databases		
DIGGOUEDU				
DISCOVERY	Web list(s)	Local database(s)	Catalog	External database(s)
	Orainhannan	Ordela a series	Detailed description	Internal an etermine and iterate
Attributes	Quick access	Quick access	Detailed description	Interplanetary audience
	Limited description	Medium description	Fign precision	
	Low precision	Medium precision	Complex collocation	
	Limited collocation	Varied audience	Koutine maintenance	
	Limited maintenance	format	Provides inventory of whole	
	Drovidos focus by	Tormat	allestion	
	allection tonic or format		conection	
	conection, topic, or format			
Examples	Announcement page	Database lists	MERLIN	WorldCat
p100	Subject guides	A-to-Z list of e-journals		Virtually Missouri
	Finding aids	Finding aids		UM Digital Library
	6	Subject guide database		UNESCO/IFLA
		3 6		Directory of Digitized
				Collections

Appendix E. Guidelines for Discovery Tool Selection for MU Digital Collection Resources

Appendix F. Metadata Schemes and Standards for Cataloging and Description

Metadata encoding standards

- Digital Libraries : Metadata Resources (IFLA): <u>http://www.ifla.org/II/metadata.htm</u>
- Minnesota State Archives: <u>http://www.mnhs.org/preserve/records/metadataresources.html</u>
- Library of Congress: <u>http://www.loc.gov/standards/</u>
- Canadian Heritage Information Network: <u>http://www.chin.gc.ca/English/index.html</u>
- Metadata Standards Crosswalk (Getty Institute): <u>http://www.getty.edu/research/conducting_research/standards/intrometadata/3_crosswalk</u> <u>s/crosswalk1.html</u>
- Technical standards for electronic bibliographic data/metadata: <u>http://staff.library.mun.ca/staff/toolbox/standards.htm</u>
- Metadata Reference Guide from MIT Libraries: http://libraries.mit.edu/guides/subjects/metadata/
- Metadata Standards. Metadata standards (Princeton University): <u>http://diglib.princeton.edu/?_xq=html&_xsl=metadata.xsl</u>
- MARC Standards <u>http://www.loc.gov/marc/</u>
- OCLC Bibliographic Formats and Standards: <u>http://www.oclc.org/bibformats/</u>
- Digital Libraries: Cataloguing and Indexing of Electronic Resources / IFLA: http://www.ifla.org/II/catalog.htm

Metadata content standards

- Metadata Standards. Data content standards (Princeton University): <u>http://diglib.princeton.edu/?_xq=html&_xsl=metadata.xsl</u>
- Digital Libraries: Cataloguing and Indexing of Electronic Resources / IFLA: <u>http://www.ifla.org/II/catalog.htm</u>
- Content standards (Moving Image Collections Cataloging and Metadata Portal) http://mic.imtc.gatech.edu/catalogers_portal/cat_standrs.htm#standards

Finding aids

• Standards for Archival Description. Chapter 5, Finding aids and other descriptive formats (Non-Cataloging Structure and Content Standards): <u>http://www.archivists.org/catalog/stds99/chapter5.html</u>

Discovery tools/systems

- Z39.50
- Z39.88

Thesauri

- Thesauri / W3C: <u>http://www.w3c.rl.ac.uk/SWAD/thes_links.htm</u>
- Standards for Thesauri Construction (NLC): http://www.chin.gc.ca/English/Standards/vocabulary_thesaurus.html
- National Library of Canada: <u>http://www.collectionscanada.ca/8/4/r4-287-e.html</u>

Universal access

• World Wide Web Consortium (W3C): <u>http://www.w3.org</u>

Appendix G. Digital Integration Matrix

In Integration of Digital Library Services

(http://www.library.jhu.edu/departments/librarydean/integration.html) Sayeed Choudhury presents the matrix used to describe the digital library activities of Johns Hopkins University and its partners. This matrix may be useful for evaluating and planning MU Libraries digital collections:

	Research	Learning	Dissemination	Preservation
Collections				
Services				
Infrastructure				

"If this matrix is filled in with specific digital initiatives or activities, one notes the following matrix:

	Research	Learning	Dissemination	Preservation
Collections	Digitization Finding aids	E-reserves Virtual reality	OAI harvesting Websites Portals	Digital preservation
Services	Virtual reference Library catalog Annotation Search	Courseware E-portfolios Simulations Virtual reality	E-publishing E-portfolios Portals	Data curation Emulation Migration
Infrastructure	Data mining	Integration with service modules		Repositories

As the final step in considering integration of digital library services at Hopkins, the Sheridan Libraries' (and its partners') digital initiatives are described within this library digital programs organizational framework:

	Research	Learning	Dissemination	Preservation
Collections	Levy sheet music Roman de la Rose ETDs	E-reserves Roman de la Rose	Fowler ETDs	
Services	Virtual reference Library Catalog Gamera GIS E-science data curation	Sakai E-portfolios SCALE Usability GIS E-science data curation	E-portfolios Gamera SCALE GIS Usability CAPM E-science data curation	DiVA DiVA E-science data curation
Infrastructure			DSpace Fedora LOCKSS	DSpace Fedora Archive Ingest Handling Test

Viewing projects in this way highlights the following observations:

- The projects are part of the overall, comprehensive library digital programs.
- The library digital programs at Hopkins encompass a substantial and diverse portion of this matrix.
- Viewing projects in this manner helps foster "strategic fit" in the manner described by Michael Porter in his November-December 1996 Harvard Business Review article "What is Strategy?" Porter persuasively argues that a system of interconnected activities, rather than isolated practices, provides strategic advantage.
- Hopkins faces a "gap" in digital preservation of collections; however, it may be best for Hopkins to collaborate with others who have greater experience and expertise in this regard.
- The infrastructure "gaps" reflect an unexplored research area."

Sayeed Choudhury. Integration of Digital Library Services

(http://www.library.jhu.edu/departments/librarydean/integration.html) Accessed May 17, 2006.

Appendix H. Possible Grant Funding Agencies

National Endowment for the Humanities:

- Grants to Preserve and Create Access to Humanities Collections <u>http://www.neh.gov/grants/guidelines/pcahc.html</u>
- Preservation and Access Reference Materials Grants
 <u>http://www.neh.gov/grants/guidelines/referencematerials.html</u>
- Preservation and Access Research and Development Projects Grants <u>http://www.neh.gov/grants/guidelines/researchdevelopment.html</u>
- Preservation Assistance Grants for Smaller Institutions <u>http://www.neh.gov/grants/guidelines/pag.html</u>
- Libraries and Archives: Implementation Grants
 <u>http://www.neh.gov/grants/guidelines/implement-libraries.html</u>
- Museums, Libraries, or Special Projects: Planning Grants
 <u>http://www.neh.gov/grants/guidelines/public-planning.html</u>

Institute of Museum and Library Services

• National Leadership Grants (http://www.imls.gov/applicants/grants/nationalLeadership.shtm)

Missouri State Library

- Library Services and Technology Act funding
 - Bring in an Expert http://sos.mo.gov/library/development/grants/ST_BEX_March2006.pdf
- Digital Imaging
 - Track I First Steps in Digitization
 <u>http://sos.mo.gov/library/development/grants/LT_DigitalImagingTrackI_Firs</u>
 <u>tSteps_Winter-Spring2006.pdf</u>
 - Track II Demonstration Projects <u>http://sos.mo.gov/library/development/grants/LT_DigitalImagingTrackII_De</u> <u>moProjects_Winter-Spring2006.pdf</u>
 - Track III Retrospective Metadata <u>http://sos.mo.gov/library/development/grants/LT_DigitalImagingTrackIII_R</u> <u>etroMetadata_Winter-Spring2006.pdf</u>
- Cooperation Grants
 <u>http://sos.mo.gov/library/development/grants/LT_Cooperation_Spring2006.pdf</u>

Additional grant-related information will be available through established grants management controls currently being prepared by Grants Task Force (2006).

Appendix I. MU Libraries Digitization Projects in Process or Completed 6-23-2006

The Savitar, 1893-2000
 Status: Complete: Savitars through the year 2000 have been digitized, OCR processed, and equipped with a search mechanism. The project will continue so that yearbooks up to current may be digitized and made accessible on the UM Digital Library. Lead: Michael Holland
 Partner/Service Provider: UM Library Systems Office http://digital.library.umsystem.edu/ebind/savitar.html

2. The Historic Newspaper Project Status: Complete: *The* Columbia *Missourian* from 1976- 1985 has been digitized, OCR processed, and prepared using Olive Software Lead: Pat Timberlake Partner/Service Provider: UM Library Systems Office <u>http://newspapers.umsystem.edu/archive/skins/missouri/navigator.asp</u>

3. Guides and Finding Aids to the Collections of MU Libraries Special Collections

Status: Complete: Seven guides produced by the Ellis Library and dating from 1949 to 2001 have been scanned are available on the UM Digital Library website. New guides will also be added as they are developed.
Lead: Michael Holland
Partner/Service Provider: UM Library Systems Office

Partner/Service Provider: UM Library Systems Office http://digital.library.umsystem.edu/cgi/t/text/text-idx?page=home;c=espe

4. **Rare Texts on Demand**

Status: Ongoing: Thus far 4 rare texts have been digitized and made available on the UM Digital Library. The scanning and image processing is paid for by the requestor and the digitized texts will be made available through links from the Merlin catalog. This service will continue on a cost recovery basis. Lead: Michael Holland, Kurt Kopp

Partner/Service Provider: UM Library Systems Office http://digital.library.umsystem.edu/cgi/t/text/text-idx?page=home;c=esm

5. Daniel Webster Speech Collection

Status: Complete: Collected speeches and orations of Daniel Webster from a collection of approximately 100 pieces (1806-1932) from the holdings of MU Special Collections, Archives, and Rare Books. All pieces are available in the UM Digital Library. Lead: Michael Holland, Kelli Hansen Partner/Service Provider: UM Library Systems Office http://digital.library.umsystem.edu/cgi/t/text/text-idx?page=home;c=webster

6. Forth of July Oration Collection

Status: Complete: Speeches given on the anniversary of the signing of the Declaration of Independence from (1791-1925) exist as a single collection within the holdings of MU Special Collections, Archives, and Rare Books. All 458 orations in the collection are available in the UM Digital Library. Lead: Michael Holland, Kelli Hansen

Partner/Service Provider: UM Library Systems Office http://digital.library.umsystem.edu/cgi/t/text/text-idx?page=home;c=jul

7. University of Missouri Alumni Magazines.

Status: In Process: The overwhelming majority of the issues of the alumni magazine from 1905-2005 have been digitized and OCR processed. We are currently working to fill the gaps in the periodical holdings of the University Archives with holdings from the MU Alumni Association Publications Office. All digitized issues are available in the UM Digital Library. New issues of MIZZOU will be added to the collection one year after publication.

Lead: Michael Holland, Gary Cox

Partner/Service Provider: UM Library Systems Office and the MU Alumni Association Publications Office

http://digital.library.umsystem.edu/cgi/t/text/text-idx?page=home;c=alum

8. British Pamphlets about Religious Descent (Howie Collection)

Status: Pending: Almost 700 pamphlets of varying age and type faces from the collection have been digitized as a pilot project. This limited part of the Howie Collection was digitized as a model to support grant applications for the digitization of the remaining 17,000 items in the collection. Development of a grant application is ongoing. All currently digitized pieces are available in the UM Digital Library. Lead: Michael Holland, Kelli Hansen Partner/Service Provider: UM Library Systems Office http://digital.library.umsystem.edu/cgi/t/text/text-idx?page=home;c=pam

9. The Digital Scriptorium: A Prototype Image and Database & Visual Union Catalog or Medieval and Renaissance Manuscripts

Status: Complete: All items from the *Fragmenta Manuscripta* Collection in the Special Collections, Archives, and Rare Books department have been scanned and mounted and 18 manuscript codices from MU Libraries Special Collections and the MU Museum of Art & Archeology have been scanned and added to the project. The high density images for this project reside on the UM Libraries Digital Library website and are accessed by the database program residing on the *sunsite* server at the University of California Berkeley.

Lead: Michael Holland/Alice Allen/Margaret Howell/Kelli Hansen Partner/Service Provider: Digitization by MU Libraries Library Computer Support unit and the UM Library Systems Office, and indexed and described by contractor Tim Spense and Lisa Hunt.

http://sunsite.berkeley.edu/Scriptorium

10. MU in Brick and Mortar: Building History of the University of Missouri-Columbia

Status: Complete: this cooperative project was LSTA Digital Imaging Grant funded.
Photographic images were selected from the University Archives and building plans and drawings from the MU Infrastructure Archives to form the content of the project.
Lead: Jim Cogswell, Andrew Skupsky and Mark Meade
Partner/Service Provider: Infrastructure Archives and University Archives
http://spam-gis03.col.missouri.edu/historic/Histpreserv.htm

11. MU College and Department Histories

Status: Complete: Selected from the holdings of the University Archives over 20 histories of University of Missouri-Columbia departments and colleges, authored by various MU faculty and staff, were digitized, OCR processed, and prepared for searching. As new

histories with Board of Curators copyright are published or public domain histories come to light they will be added to the collection on UM Digital Library. Lead: Gary Cox, Michael Holland Partner/Service Provider: UM Library Systems Office http://digital.library.umsystem.edu/cgi/t/text/text-idx?page=home;c=sch

12. Missouri School of Journalism Bulletins

Status: Complete: Thirty-six bulletins in the Missouri School of Journalism Bulletin Series, 1912-1967 were digitized, OCR processed, and prepared for searching. All pieces are available in the UM Digital Library. Lead: Pat Timberlake

Partner/Service Provider: UM Library Systems Office http://digital.library.umsystem.edu/cgi/t/text/text-idx?page=home;c=umjour

13. The Tiger Claw

Status: Complete: All issues of the University of Missouri High School yearbook, The Tiger Claw, were drawn from the holdings of the University Archives. The issues from 1923-1973 (except 1926) have been digitized, OCR processed, and equipped with a search mechanism. All volumes are available in the UM Digital Library. Lead: Gary Cox

Partner/Service Provider: UM Library Systems Office and Robert Luebbers from the University High School Alumni Association. http://digital.library.umsystem.edu/cgi/t/text/text-idx?page=home;c=tclaw

14. Missouri History, Geology Culture Digital Library

Status: Ongoing: This collection of books, pamphlets, maps and other printed materials about the state of Missouri has drawn heavily from the holding of the Ellis Library, the Special Collections, Archives and Rare Book department, and other collections to make a very rich collection of Missouriana. The collection will continue to be added to as appropriate items come to light. All pieces are available in the UM Digital Library. Lead: Kurt Kopp, Michael Holland

Partner/Service Provider: UM Library Systems Office http://digital.library.umsystem.edu/cgi/t/text/text-idx?page=home;c=umlib

15. University Archives Web Exhibits

Status: Ongoing: Half a dozen of the Internet exhibits created by the University Archives relating to the History of the University of Missouri-Columbia were mounted as image collections in the UM Digital Library. There is also a collection of sports posters included within this group of materials. New exhibits may be added as they are developed by the University Archives.

Lead: University Archives, Kurt Kopp

Partner/Service Provider: UM Library Systems Office

http://digital.library.umsystem.edu/

http://digital.library.umsystem.edu/cgi/i/image/image-idx?page=index;c=arpostic http://digital.library.umsystem.edu/cgi/i/image/image-idx?page=index;c=umcabasketic http://digital.library.umsystem.edu/cgi/i/image/image-idx?page=index;c=mizzouy19kic http://digital.library.umsystem.edu/cgi/i/image/image-idx?page=index;c=umcaairic http://digital.library.umsystem.edu/cgi/i/image/image-idx?page=index;c=umcadairyic http://digital.library.umsystem.edu/cgi/i/image/image-idx?page=index;c=umcadairyic

16. WWI and WWII Posters Collection, 1915-1948

Status: Pending: 150 WWI posters from the 1000 item collection have been digitized and indexed and print testing is in progress. Grant funding will be requested for this project if the results of the pilot study are encouraging. Lead: Michael Holland, Karen Paulik Witt, Kelli Hansen Partner/Service Provider: UM Library Systems Office, et al.

17. *Missouri Agricultural Experiment Station Publications Missouri Cooperative Service Circulars*, 1888-2005 Status: In Process: All Agricultural Experiment Station publications identified in SR333 (Bulletins, Circulars, Special Reports, and Research Bulletins) are in the process of being digitized and prepared for public use. This project was funded by a LSTA Digitization grant administered by the Missouri Secretary of State's Office through the Missouri State Library. All 2,827 titles are or will be available in the UM Digital Library. Lead: Michael Holland and Judy Maseles Partner/Service Provider: UM Library Systems Office and the LSTA through the Missouri State Library of the Missouri Secretary of State's Office. http://digital.library.umsystem.edu/cgi/t/text/text-idx?page=home;c=agext

18. Missouri County Plat Books Published by W.W. Hickson, and Co.

Status: Complete: 118 individual county plat books published in 1930 are held in the Special Collections, Archives and Rare Books Department of Ellis Library. These cartographic materials were digitized for preservation purposes as well as increasing their access to the public-at-large.

All plat books are available in the UM Digital Library. Lead: Michael Holland, Kelli Hansen Partner/Service Provider: UM Library Systems Office http://digital.library.umsystem.edu/cgi/i/image/image-idx?page=index;c=platic

19. Nitrate and Damaged Negative Collection Conversion

Status: Complete: Approximately 500 damaged and nitrate based images of campus and student events 1931-1964 were located in the University Archives and then scanned and described.

C:1/141/6 Box 6 Lead: Michael Holland, Karen Paulik Witt Partner/Service Provider: University Archives

20. Scanning of the MU Libraries 3,000,000th book

Status: Complete: The volume, *The Navigator* by Zadok Cramer, 1817, has been digitized and OCR processed and is available to the public on the UM Digital Library as a volume within the Missouri *History, Geology, and Culture* digital text collection. Lead: Jim Cogswell, John Wedman Partner/Service Provider: UM Library Systems Office and the School of Information Science and Learning Technologies

21. State of Missouri Sanborn Fire Insurance Maps 1885 – 1922

Status: In Process: This project was funded by an LSTA Digitization grant administered by the Missouri Secretary of State's Office through the Missouri State Library. All 6,400 Missouri fire insurance maps will be available in the UM Digital Library. Lead: Michael Holland, Karen Paulik Witt Partner/Service Provider: UM Library Systems Office

Potential Digitizing Projects in Special Collections, Archives, & Rare Books

- 1. *University of Missouri Bulletin: Engineering Series*; no. 1. *Some experiments in the storage of coal*, [by] E.A. Fessenden [and] J.R. Wharton. Status: Completed and in MU Departmental Histories in UM Digital Library
- 2. *Engineering Experiment Station Series Bulletins* (University of Missouri--Columbia. College of Engineering). No 1-67 (1910-1969)
- 3. *Engineering Extension Bulletin Series* (University of Missouri--Columbia. College of Engineering) no. 1-21 (1963-1972).
- 4. University of Missouri Traffic Engineering Conference Engineering Extension Series Bulletin Series (University of Missouri--Columbia. College of Engineering)
- 5. **Glass Plate Negative Collection from the School of Journalism** 150 plates dating from 1909-1915 University Archives: C:11/13/1
- 6. **Glass Plate Negative Collection from the School of Forestry** 15 plates dating from 1913-1914 University Archives: C: 4/1/4
- Glass Plate Negative Collection from the MU Alumni Association 6 plates of St. Patrick's Day Festivities at Engineering dating from 1909 University Archives: C: 0/3/8
- 8. Selected Portions and *Elements of Incunabula* in the Ellis Library: 1453-1499

Johann Gutenberg. Bible (1 leaf), 1454

Franciscus Renner, de Heilbronn. Bible, 1480

Boccaccio, Giovanni. Genealogie deorum, 1494

Diaiogus inter clericum et militem, 1498

Eusebius Caesariensis. De evangelica praepartione, 1497

Ficinus, Marsilius. Epistolae, 1494

Firmicus Maternus, Julius. De nativitatibus, 1499

Formularium procuratorum, 14--?

Justinus, M. Junianus. Epitome in Trogi Pompeii historias, 1470

Passio Domini Jesu Christi secundum quattuor Evangelia, 1498

Philelphus, Franciscus. Epistolae, 1496

Pigouchet: Vostre (2 leaves), 1498

Pius II. Epistolae familiares, 1496

Plutarchus. Vitae illustrium virorum, 1496

Savanarola, Hieronymus. Prediche quadragesimali dell'anno 1495, 1496

Schedel, Hartmann. Liber chronicarum, 1493

Sebastianus, P. Epistola consolatoria dei cald freddi e tiepidi, 1496

Spiera, Ambrosius. De quadragesimale de floribus sapientiae, 1494

Tritheim, Johann. De scriptoribus ecclesiasticis, 1494

Appendix J. Report by the University of Missouri Libraries Task Force on Imaging, September 23, 1999

"Building a Foundation for the UM Digital Library"

A Report by the University of Missouri Libraries Task Force on Imaging

September 23, 1999 Revised April 28, 2000

Task Force Members: Elizabeth Ader – UMKC Alice Allen – UMC Kurt Kopp – LSO Raleigh Muns – UMSL Nancy Stancel – UMKC LAW Andy Stewart – UMR Scott Walters – UMKC

"Building a Foundation for the UM Digital Library"

Table of Contents

Executive Summary	page 1
Administrative Models	page 3
Selection and Preservation Issues	page 4
Copyright Issues	page 6
Metadata	page 8
Software for Digitizing	page 9
Equipment for Digitizing	page 9
Digitizing Best Practices	page 10
Funding	page 11
On-going Maintenance and Support Issues	page 12
Outsourcing vs. In-House Digitizing	page 12
Appendix A – Notes on Site Visits	page 14
Appendix B – Webography	page 16

"Building a Foundation for the UM Digital Library"

A Report by the University of Missouri Libraries Task Force on Imaging

Executive Summary

The University of Missouri Library Directors established the UM Imaging Task Force to investigate current issues and concerns relevant to the establishment of digital library projects and to provide recommendations to the Library Directors. The Task Force began by studying national trends in digital libraries and reviewing what is being done in other libraries comparable to ours. We made site visits to the University of Kentucky, the University of Michigan, and the Online Archives of California, and attended three excellent workshops on fundamental issues and technologies. During our investigations we became convinced that our resources should not be used to fund a selection of disparate projects. Instead, we should be laying the foundation for a UM Digital Library.

"Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities." (From The Digital Library Federation, "A working definition of digital library" http://www.clir.org/diglib/dldefinition.htm)

The Task Force recommends:

1 The establishment of a UM Digital Library (UMDL), built initially on projects developed within the University of Missouri System, with the possibility of its expansion to include non UM materials.

The University of Missouri Libraries have an existing commitment to resource sharing and cooperation, as represented by the MERLIN consortium, and more recently by the MOBIUS consortium. These consortia arose to save money, and optimize use of information resources. While the UMDL will function within the well-established MERLIN structure, it is nevertheless important that it have its own name and identity. This builds awareness in the user community and can provide a focus for the development of technological expertise. It will also assist the UM Libraries in attracting the considerable resources necessary to make these initiatives successful.

2 The adoption of existing standards for the selection, production, and maintenance of digitized materials.

It is critical to adopt existing digitizing standards to enter in to partnerships with other institutions and to participate in national digital library efforts.

3 The establishment of a standing UM Digital Library Committee (UMDLC) to oversee the implementation, definition, and adoption of standards.

The UMDLC will guide the development of the UMDL. Activities of the committee will consist of recommending to the UM Library Directors criteria for selecting materials to digitize; making specific recommendations to the Library Directors for materials to digitize; recommending to the UM Library Directors standards for digital projects and overseeing the adoption of and adherence to standards. The UMDLC may engage in other activities to promote the long-term development of the UMDL such as the sponsorship of symposia, the survey of candidate materials, and the survey of existing hardware.

4 The establishment of a central digitizing center headed by a full-time director at the Library Systems Office.

The establishment of a digitizing center at LSO will result in economies of scale and ensure quality control over the entire digitization process. The establishment of a digitizing center at LSO does not preclude local digitizing projects.

The appointment of a full-time director allows for coordination of efforts across campuses, and creates an advocate and representative for the UMDL at all levels. The director would also be charged with pursuing external funding opportunities.

5 The identification and implementation of a pilot project which involves materials contributed by all UM Libraries.

A pilot project is a first step to get all of the UM Libraries involved in the creation of the UMDL. Ideally, the initial pilot project would include materials from all of the libraries and would be attractive not only to the entire University, but to the State of Missouri. This would start the project off in the spirit of cooperation and could lead to increased state funding. This report stresses the need for extensive and deliberate planning, however Task Force site visits uncovered the somewhat obvious rule: you have to start somewhere.

The remainder of this report expands upon and refines these recommendations. It suggests possible administrative models for implementing the UMDL. It also provides more in-depth information surrounding the broader issues of digital libraries.

Administrative Models

There are many ways in which the University of Missouri Libraries could establish the organizational infrastructure to select, plan for, and implement a UM Digital Library (UMDL) ranging from a totally decentralized approach to a completely centralized approach. The options are not limited to decentralized vs. centralized, but includes a continuum of variations between those two extremes. Several alternatives are discussed below followed by the task force recommendations. Each of these options assumes that projects to be done have been approved by the UM Library Directors and will be funded from the pool of money available from the state for digitization projects at the UM Libraries. Projects funded totally by individual libraries or with outside grant funds could, of course, be used however the library chooses although if the project is to be included in the UM Digital Library, it must follow established UMDL standards.

Option 1 – Totally Centralized: Establishment of digitizing center at Office of Library Systems headed by a UMDL Director; projects selected by UM Library Directors from those recommended by committee with representatives from all libraries; LSO staff does conversion; data stored on central server; single search interface; and application of standard digitizing, encoding and indexing methods.

Advantages:

- 1. Lower costs because of non-duplication of staff, equipment, and servers
- 2. Ability to search across databases due to uniform encoding and access.
- 3. Consistent quality control

Disadvantages:

- 1. Lack of local control over digitizing process and products.
- 2. Digitizing expertise centered in LSO, not distributed throughout UM Libraries.

Option 2 – Centralized with provision for local projects: Establishment of digitizing center at Office of Library Systems headed by an UMDL Director; projects selected by UM Library Directors from those recommended by committee with representatives from all libraries; LSO staff does conversion; data stored on central server; single search interface; and application of standard digitizing, encoding and indexing methods. In addition to providing digitizing services and support, LSO would serve as a research center into new digitizing technologies.

Projects approved by the UM Library Directors could be done centrally or locally. Locally produced projects could be housed on the central server and included in the UMDL if UMDL standards have been adhered to.

Advantages:

- 1. Lower costs because of less duplication of staff, equipment, and servers
- 2. Ability to search across databases due to uniform encoding and access.
- 3. Consistent quality control

4. Provides for the development of local digitizing expertise and support for local projects

Disadvantages:

- 1. Lack of total local control over digitizing process.
- 2. Digitizing expertise primarily centered in LSO, not necessarily distributed throughout UM Libraries.
- 3. May result in the duplication of staff, staff training, space, and equipment costs on campuses

Option 3 – Decentralized with some centralization: projects selected by UM Library Directors from those recommended by a selection committee with representatives from all libraries. Each library establishes a digitizing center; uses its staff to do the conversion. All data is housed on a central server provided by LSO and quality control is overseen by LSO based upon mutually agreed upon standards.

Advantages:

- 1. Allows local control of digitizing process
- 2. Digitizing expertise available at all UM Libraries.
- 3. Standard encoding and access methods allow cross database searching.

Disadvantages:

- 1. Duplication of staff, staff training, space, and equipment costs on all campuses
- 2. Mixed organizational structure does not allow for effective oversight of digitizing projects.

Option 4 – Decentralized with standards: Each library establishes a digitizing center; uses its staff to do the conversion; provides access to the data and images on locally maintained servers. Each center will use agreed upon standards for coding, indexing, and access methods.

Advantages:

- 1. Total local control.
- 2. Digitizing expertise available at all UM Libraries.
- 3. Standard encoding and access methods allow cross database searching.

Disadvantages:

- 1. Duplication of staff, staff training, space, and equipment costs on all campuses
- 2. Duplication of server costs
- 3. Additional overhead of tying local servers together for cross database searching
- 4. As data is produced and processed locally, there is the possibility of deviation from standards leading to incompatible databases.

Option 5 – Totally Decentralized: Each library establishes a digitizing center; uses its staff to do the conversion; provides access to the data and images on locally maintained servers; and uses whatever coding, indexing, and access methods it deems most suitable.

Advantages:

- 1. Total local control.
- 2. Digitizing expertise available at all UM Libraries.

Disadvantages:

- 1. Duplication of staff, staff training, space, and equipment costs on all campuses
- 2. Duplication of server costs
- 3. Possibility of incompatible formats, search interfaces, and encoding across campuses

Task Force Recommendation: The task force recommends the adoption of Option 2 because:

- 1. Cost centralization of hardware, software, and staff allows for economies of scale
- 2. System-wide representation a UM wide committee would provide for local input and guide the development of the UM Digital Library.
- 3. Coordination the appointment of a full-time Director is critical to ensure the success of the UMDL

Selection and Preservation Issues

I. Materials Selection

"Materials selection" in this document refers to the process of identifying resources owned by the UM community that are appropriate for digitization. The two purposes of digitizing are: first, to make resources widely available; second, to preserve items in an archival format.

Potential items for selection for digitization must go through a systematic collection development review process. Once selected, digitized items should be enhanced with value-added metadata. In selection decisions, UM must first determine what areas and levels of staff expertise are available. Equally important is determining who will be responsible for the selection and physical evaluation of materials, as well as selecting the best method of access for the user. This team (UMDLC) should include library subject specialists, catalogers, archivists, computer/network systems people, etc.

Items are selected for digitization to: (Source: <u>http://images.grainger.uiuc.edu/dii/backgrnd.htm</u> (University of Illinois-Urbana-Champaign))

- 1. Preserve and to make accessible fragile and under-utilized visual resources.
- 2. Promote the use of digital images throughout the campus and scholarly community.

Selection Criteria for Digitizing Projects (Source: <u>http://www.columbia.edu/cu/libraries/digital/criteria.htm</u> (Columbia University))

The criteria stated below should be considered when selecting materials for digitizing before any new project is initiated. It is important to note that digital collections do not replace original materials.

- 1. The product has sufficient intrinsic value to ensure ongoing use by a defined constituency on and/or beyond the UM community.
- 2. There is no identical or directly similar digital product that can meet these needs.
- 3. There is advocacy for the project and/or a reasonable expectation that the product will have immediate utility for members of the UM community.
- 4. Intellectual property rights are managed in accordance with applicable laws, and any necessary restrictions to access can be implemented through current institutionally-supported mechanisms.

Added Value Derived from Digitizing Projects (Source: <u>http://www.columbia.edu/cu/libraries/digital/criteria.htm</u> (Columbia University))

1.Digital capture will enhance intellectual control through creation of new finding aids, links to bibliographic records, and development of indices and other tools.

2. The ability to search widely, manipulate digital objects, and study them in new contexts will increase and enrich use.

- 3. The widespread dissemination of local or unique collections will encourage new scholarly use by providing enhanced resources.
- 4. Digital capture will enhance use through improved quality of images, e.g., through improved legibility of faded or stained documents.
- 5. Digitization will allow the flexible integration and synthesis of a variety of formats, or of related materials scattered among many locations, thereby creating a "virtual collection."

II. Preservation Issues (Source: <u>http://ahds.ac.uk/manage/framework.htm#sec4</u> (A Strategic Policy Framework for

Creating and Preserving Digital Collections))

Academic data archives typically make information about their holdings available through [online] catalogs, and may adopt one or several distribution scenarios. Preservation policy consists of a suite of strategic and procedural decisions which together with other aspects of data management help to ensure that the content, context and authenticity of a data resource survives through time and changing technologies with minimal loss in its information content, functionality, and accessibility. Decisions involve the adoption of a preservation strategy or combination of strategies that include:

- 1. **migration** (data is stored in software-independent format and migrated through changing technical regimes);
- 2. **technology preservation** (data is preserved along with the hardware and/or software on which it depends);
- 3. **emulation** (the look, feel, and behavior of a data resource is emulated on successive hardware/software generations);
- 4. **long-term preservation** is highly contingent on decisions made at the time the resource is digitized and during its subsequent management, and also rests on available funding and technologies. It is also undertaken to maintain future access and use of the resource and is therefore closely linked and potentially contingent upon data use.

Items selected for digitization must undergo a systematic review process by the UMDLC. Selected items must have an intrinsic value and/or are not identical or directly similar to any other digital product. Items selected should be expected to have immediate value for members of the UM community.

Copyright Issues

The relationship of copyright to digitizing projects is less one of specifics about the law, and more one of developing an overall management philosophy of the relationship between the materials which are to be digitized and the idea of copyright. The administration of digital projects requires that first, an overarching philosophy be established, and that second, existing copyright law be applied within the context of that established philosophy or policy.

Though a defined body of intellectual property law and regulation exists and can be consulted, by its nature copyright law is ambiguous, malleable, and open to diverse interpretation. Evolving case law in this area is constantly defining and changing the copyright field. This does not mean that decisions about copyright cannot be made, rather, it means that there is a broad spectrum, if not kaleidoscope, of issues that may or may not arise on a case by case basis.

Many institutions have developed extensive resources about intellectual property for their local use. Typically, these sites do three things:

- 1. List links to intellectual property resources online (e.g., United States Code)
- 2. Promulgate the institution's policies regarding intellectual property rights
- 3. Gives institutional interpretations and distillations of existing laws and regulations

Some prototypical sites already exist within the UM system which do these three things, and it seems straightforward to assume that a digitizing project incorporate the existing institutional policies with regards to copyright. A decision could also be made to devote resources to developing and refining policy and tools that directly address this issue. The libraries could create a local copyright site where participants in future digitizing projects would be able to examine this issue, and propose with some understanding specific materials to digitize. In lieu of a "resource site," it is still necessary that the libraries have an established policy to which internal reference can be made when evaluating digitizing project proposals.

The closest thing to a definitive statement about intellectual property that can be stated is "Everything has an ownership status." Often that status is: unclear. For example, according to current copyright law, items which have been printed that are older than 75 years should unambiguously be in the public domain. A conservative copyright management philosophy might consist of restricting digitizing projects only to those monographs.

In contrast, items that have been produced, but not printed (e.g., correspondence) even if more than 75 years old are currently in an intellectual property limbo.

It is recommended that copyright status categories be delineated. For example:

Monographs

Published more than 75 years ago Published less than 75 years ago, but copyright never renewed Under copyright but permission granted to digitize and disseminate Under copyright, but no identifiable rights holder

. . .

The copyright category and copyright status of material would be information that would be evaluated in conjunction with the broader question of the subject material that is to be digitized. In a sense, the issue of copyright is a technical one, and the topic of content is a philosophical and political one. Addressing copyright is a question of "can we do this?" whereas addressing content is a question of "why do we do this?"

Risk and Benefit

The idea of risk is a crucial element of considering copyright and digitizing projects. Since there is so much inherent ambiguity as to the copyright status of many items, it follows that the philosophy of an institution regarding the acceptance of risk will determine the population of items that can be implemented in digitizing projects. If you know an item is in the public domain, you know that you can digitize it. If the intellectual status of an item is ambiguous, you must choose whether or not to proceed (accepting a certain amount of risk) or to avoid risk altogether and discard the item from any project.

There is an obvious dichotomy between risk and benefit. The more risk one is willing to accept, the richer the body of materials that can be utilized. An inability to accept ANY risk will result in

a small pool of items (public domain monographs) many of which are being, or will be, digitized by other institutions.

From a site visit to the Online Archive of California (OAC) we have an illustrative example of how risk can be addressed: The OAC has decided that within the definition of "Fair Use" that they can digitize parts of collections of photographs which are of unidentifiable provenance (and thus ownership). By digitizing only a small percentage of all the photographs of a given collection, the OAC posits that they are "paraphrasing" the entire collection.

The interpretation of existing copyright laws and regulations is part of the process of risk acceptance. The types of materials that fall under this risk assessment model are primarily those items held in special collections, archives, and manuscript repositories. Incidentally, those are precisely the kinds of materials that often are considered for digitizing projects.

Another aspect of risk analysis involves the fairly new concept established by state law regarding the intellectual property status of the image of a celebrity. Tennessee State law establishes rights to the heirs of Elvis Presley regarding the use of his name and image. Other states, California for one, have similar laws that must be taken in to account when considering digitizing.

A total risk avoidance policy would prohibit digitizing ANY photograph of an individual without first obtaining written legal release from them. In practice, this would mean that photographs of people would be unlikely to be digitized. The U.S. National Park Service has adopted this policy in their own digitizing projects which explains why most US National Park Service web sites are rife with digitized photographs of bridges and trees.

An example of a more rational risk acceptance policy would be one that recognizes that historic photographs of unknown individuals who are not famous, and are taken on a public street, could be digitized. Another risk mitigation policy could be established so those items could be removed from digitizing sites when requested to do so by rights owners. At the same time, this would be an opportunity to request from the rights owner the permission to put the item on the Internet.

Again, there is no simple single answer regarding risk, but the issue should be addressed and embodied in the establishing policies for system wide digital projects.

Users and Producers

Intellectual property involves the rights of those who produce information, and the rights of those who utilize it. Classically, libraries have been users of information and our policies, rules, and culture revolve around issues of those who wish to use information. In practice, this is expressed by a desire to legally get as much as possible and pay as little as possible (ideally, "free"). "Fair Use" is an expression of library activity in the area of copyright as a user of information.

Digitizing projects on the other hand, imply that the libraries, and others, will be producing online information that will have an ownership status. It must be clarified as to who controls

future permissions to utilize library produced or directed digital archives, and their elements. If the libraries fund other University units' digitizing projects, the ownership status must be established to and agree upon before proceeding. The libraries will be owners of information, and what that entails must be clearly defined.

Metadata (Administrative) and copyright

Metadata refers to information about information. For example, library catalog records are metadata about library holdings. The most daunting aspect of digitizing projects is the establishment of metadata formats and actual metadata content surrounding the digital products. Administrative metadata, specifically copyright information, is a critical element of all digital projects. Information regarding the copyright status of a work (if unknown, that may be stated), permissions for reproduction, who the rights holders are, restrictions on use, should be imbedded in the metadata for a given item.

The Dublin Core metadata element set

http://purl.org/DC/about/element_set.htm

contains fifteen elements for establishing metadata records for online digital resources. Item number fifteen is :

Rights Management Label: RIGHTS

A link to a copyright notice, to a rights-management statement, or to a service that would provide information about terms of access to the resource. Formal specification of RIGHTS is currently under development. Users and developers should understand that use of this element is currently considered to be experimental.

A philosophy and understanding of the bigger picture of copyright in the context of digitizing projects is a necessary precondition before ever dealing with the ever-shifting nuts and bolts of applying copyright issues. As the IDEA of cataloging is a foundation for organizing library holdings, and the library catalog is the practical expression of that idea, so to must we establish the underlying IDEAS of handling digital projects and find practical expression after that in how we actually handle our digital archives.

Metadata

Metadata can be simply defined as data about data or information about information. Its primary purpose is to provide a standard description format for data so that applications using the same metadata can easily exchange data. In the world of information retrieval it allows search engines to retrieve data from many sites as long as the same metadata is used. There are three types of metadata: administrative; structural; and discovery and retrieval.

There are three major reasons to develop metadata standards: to make information universally available to users; to facilitate the sharing and interchange of information; and to preserve information (make it safe from changes in hardware and software.

According to Howard Besser (presentation at the School for Scanning Workshop in Chicago June 2, 1999) the two major problems facing those building digital collections are longevity and interoperability. By longevity he means the problem of being able to read or decode a piece of digital encoding years after its creation. Two examples he gave were the problem of trying to read a document written in Wordstar version 1 and to read the data on an 8 inch floppy. By interoperability is meant the ability to share data between applications and digital collections.

Administrative metadata is information about the digital object recorded at capture time. Types of administrative metadata would include source type, scanning date, resolution, type of image, bit depth, copyright status, etc. Structural metadata records information that is relevant to presenting the digital object to the user. This would include metadata to define the object (book, photo album, etc.) and sub-objects (pages, chapters, file size, dimensions, etc.). Discovery metadata is information describing the digital object for retrieval purposes. Three discovery metadata standards are widely used at the present time: the Text Encoding Initiative (TEI), the Dublin Core (DC); and the Encoded Archival Description (EAD).

The Text Encoding Initiative is an SGML based document description language and has been widely used to mark up electronic texts for searching. An XML based version has recently been made available. For more information see www.tei-c.org.

The Dublin Core consists of fifteen elements that can be used to describe resources. These are title, subject, description, source, language, relation, coverage, creator, publisher, contributor, rights, date, type, format, and identifier. It was originally developed to describe electronic resources on the web but has been well received by many researchers because of its relative simplicity (as compared to MARC or TEI). For more information on the Dublin Core visit http://www.oclc.org/oclc/research/projects/core/documents/index.htm.

The Encoded Archival Description language was originally developed at Berkeley and has since been adopted by the Society of American Archivists and the Library of Congress. Like the TEI it is also SGML based. It is a widely used standard for archival finding aids. Finding aids are inventories, registers, indexes, or guides that are created by archival and manuscript repositories to provide detailed information about specific collections. For more information on the EAD, see http://lcweb.loc.gov/ead/.

The Task Force does not at this time recommend any specific metadata set, but feels that a standard set should be adopted so that all collections on the UMDL can be easily searched using the same search tools and syntax.

Software for Digitizing

Necessary software for digitizing projects will fall in two categories. The first category consists of software utilized for creating the actual digital products including scanning software, and digital image software such as Adobe PhotoShop. Commercial software of this type typically costs from "free" (if included with purchased hardware), to less than \$1000 per license for commercial packages such as Adobe PhotoShop. This category of software is usually resident on single workstations as described elsewhere in this document.

The second category for software is that which is utilized for the ongoing maintenance and access of the digital products and attendant meta-data. Typical of this class of software is Inso Corporations' Dynatext and Dynaweb software. Commercial versions of products like this can cost as much as \$50,000 to license.

Equipment for Digitizing

There are several types of equipment to be considered for digitizing projects: workstations; scanning equipment; and servers.

Workstation: the exact specifications for a scanning workstation would vary depending on the type of imaging and image processing (e.g., OCR) to be done. The consensus is to buy the most that can be afforded. Following are specifications for a workstation that would be suitable for just about any digitizing application:

Monitor - 21" CPU - 600 mhz. Memory - 512 meg. Video memory - 16 to 32 meg. Writeable cdrom or dvd storage Advanced sound capability

Basically, this is currently the top end of what is available and at University of Missouri pricing from Dell Computing would cost about \$4,500.

Scanning equipment: there are many types of scanning equipment available and which one to purchase depends upon the material to be scanned. Following are the major categories of scanning devices with estimated cost ranges.

Flatbed scanners – this is the most common type of scanner and is capable of bitonal, grayscale, and color scanning. The maximum practical document size that can be scanned is 11" by 17" (newspaper size flatbed scanners are available but at a cost exceeding \$100,000). The price range is \$100 - \$50,000+

Sheetfeed scanners – these scanners use the same technology as flatbed scanners, but the doument to be scanned is fed over a stationary CCD array and light source. The price range is \$2,000 - \$40,000.

Drum scanners – these are widely used for architectural drawings and similar large size documents that can be fed through the machine. They are not suitable for fragile or bound material. Price range is \$10,000 to \$100,000.

Digital cameras – these are coming into widespread use for many imaging applications such as three-dimensional objects, large sized and fragile material. The price range is \$30,000+.

Microfilm scanners – allows the creation of digital images from microfilm. The price range is \$12,000 - \$200,000+.

Planetary overhead scanners – these scanners are useful for oversized and fragile material as they can scan images up to 17" by 23". Allows bitonal scanning only at the largest document size although grayscale capability is now available for up to 11" by 17". Price range is \$12,000 to 15,000+.

The technology of scanning equipment is constantly changing and prices are continually coming down or greater capability is provided for the same price. This section was intended as a general survey of the types of scanning equipment available and relative prices for each.

Servers: Although one can serve up digital objects on a desktop computer, there are obvious limitations to doing so. This report assumes that digitizing projects selected by the UM Libraries will be meant for system-wide viewing requiring a server with large storage capability, processing power, and bandwith. Such servers cost \$60,000 at a minimum.

Digitizing Best Practices

The best practices and digitizing formats will change as technology advances. This is what is currently recommended for digitizing images.

Although there is some disagreement among the experts in the field on exactly what the "best" practices for digitizing are, there is widespread agreement that one should create the highest quality image possible consistent with current technology and budgetary constraints. The basic theme is "digitize once, create many derivatives". Because of the time and expense involved in creating images from the original source (and in the case of fragile materials, the need to keep handling at a minimum), one would like to avoid at all costs having to go back and re-digitize.

Generally, although not universally, accepted practice for various types of material follows:

Black and white text – 600 dpi bitonal Black and white text with drawings – 8 bit grayscale or 24 bit color at 400 - 600 dpi Text with color – 24 bit color at 400 - 600 dpi Black & White and color photographs – 24 bit color at 400 - 600 dpi Three dimensional objects – 24 bit color at maximum resolution The images should be stored in TIFF format. Scanning at the above settings will result in large files and would in most cases not be used for presentation. Once the original or archival image is made, derivative images (jpeg or gif) can be produced for presentation on the web.

A much more complete discussion of imaging standards can be found at http://www.columbia.edu/acis/dl/imagespec.html.

Funding

Digital library projects are expensive. Startup costs of purchasing equipment, hiring and training staff, designing access mechanisms, etc., are substantial even before any collections are made available to users. Subsequently, there will be significant ongoing costs for hardware and software upgrades, retraining on ever-changing technologies, and programmatic expansions to meet the rising expectations of users as they become familiar with the benefits of digital library collections. The size of the investment required dictates two strategic directions which appear to have been followed by most of the institutions involved in digital library initiatives. The first is that these projects must be considered as long-term investments which must unfold over a period of several years, at least, to meet their goals. Planning and budgeting cannot be short term or year-to-year. The second strategic direction is the recognition that few, if any, libraries have the resources to implement digital libraries by themselves, or from existing funding. Almost all of the libraries whose digital initiatives we have examined rely on multiple sources of funding.

The most frequent sources of funding for digital library projects are:

- 1. Grants (government, private foundations, corporate)
- 2. Joint funding from other campus entities
- 3. Inter-institutional cooperation

The most ambitious grant funding project to date has been the Digital Library Initiative, funded by the National Science Foundation, the National Endowment for the Humanities, and several other governmental agencies. Two phases of grant funding have been awarded to date. These are large awards, sometimes in the millions. Usually, though not always, they go to large research libraries. The Library of Congress, with support from Ameritech, made awards in a sequence of three years, for the American Memory Project. The project supported digitization of music, photos, manuscripts, books, etc. in a wide range of subject areas. The Andrew Mellon Foundation has funded the Making of America projects. On a smaller scale, states have funded start-up and developmental digital projects such as the Colorado Digitization Project, which has received 2 LSTA grants from the Colorado State Library, as well as a grant from Regional Library Systems of Colorado. Private corporations will sometimes fund hardware and software acquisitions. The University of Kentucky, for example, was given \$34,000 worth of software by the Inso Corporation. Kodak supported a digital imaging project at the University of Illinois. Cornell's Institute for Digital Collections received funding from several private donors as well as \$150,000 in equipment from Intel. Many large grants have been awarded, as one might expect, to institutions that were willing to be trailblazers in the field of digital libraries. It is possible that, as more institutions begin digital projects, grant funding will be harder to come by. All the more reason to begin without delay to explore the opportunities.

Several large academic digital library projects have been developed cooperatively with other campus entities. The University of Michigan, one of the most extensive and successful digital library organizations, is jointly funded by the University Library, the School of Information, the Information Technology Division, and the Media Union, with additional funding from several other campus sources. It was a cooperative effort from the beginning. Indiana University's Library Electronic Text Resource Service resulted from a partnership between the University Library Digital Initiative, created to provide an infrastructure supporting the "collecting" of digital resources, received internal funding from the University for a five-year startup. Another form of institutional support might be the inclusion of digital library funding in the university's development priorities. The web page for the University of Virginia Library's Electronic Centers carries this statement: "The Electronic Centers are an important part of the University's Capital Campaign. <u>Read how you can help us</u> continue our innovative work."

A third model for achieving impressive results without having to bear the entire cost is to join forces with one or more other institutions to work on a joint project. An example of this is the Digital Scriptorium, a joint project of the UC-Berkeley Bancroft Library and the Special Collections Library of Columbia University to digitize and make available on the World Wide Web their holdings of medieval and early Renaissance manuscript holdings. The project is intended as a prototype that will eventually allow participation of other institutions. Part of its work is to do an economic study that "will allow us to propose a sound business plan to ensure the project's long-term viability...Without such long-term viability, the effects on scholarship will be limited, regardless of the Digital Scriptorium's technical success."

Exploring funding resources, whether from grants or from cooperation with other departments and institutions, has benefits that go beyond the financial. The digital library community is a highly interconnected one. Like the World Wide Web it depends on to make its resources available, the digital library community is an ever-larger and more complex network of relationships and cooperative endeavors. It is critical to be tied into this network, to give and to receive. Wendy Lougee, Associate Director of Libraries for the University of Michigan's Digital Library Program, told the UM Task Force that the major aspect of her job was building and maintaining relationships. "We had a mandate to get out and evangelize." Having the right relationships also makes it easier to get technical advice and tap into the considerable expertise available within colleague institutions.

The other side of the coin (so to speak) in terms of funding is planning and accountability for use of the funds. Detailed budgets should be prepared and approved each year, to allocate funding for new projects, ongoing projects, hardware and software upgrades, personnel costs, etc. The budgets are an expression of the planning process that must be constantly under way to make our digital library projects successful.

On-going Maintenance and Support Issues

Once digital collections have been created, an ongoing and growing maintenance obligation is incurred. The one-time cost of image-capture accounts for less than half of the total expense of a project. It is important to establish the degree of institutional support and funding that is likely when priorities are being determined.

A critical element of long-term digitizing initiatives, and often overlooked, is the commitment to long-term maintenance of the sources. It's easy (and cheap) to do a small pilot project, but what happens down the road? Who's committing to maintaining the machinery and software to keep the growing array of items available? As the size of the digital collection grows, this obligation becomes a critical element of digitizing initiatives. Annual budgets must take into account the overhead necessary for maintaining digital archives, and the servers with which to deliver access.

It must also be noted that if the UMDL is to maintain resources created by projects not initiated by the UMDL, ongoing maintenance costs will still be incurred. Grant proposals by such outside projects should be directed to include requests for funding in their grants to contribute to this ongoing maintenance cost.

Outsourcing vs. In-House Digitizing

The question of whether to produce UMDL digital images in-house or to contract for their production by a vendor (outsource) is a complicated one. An exhaustive list of advantages and disadvantages of each choice is given in <u>Digital Imaging for Libraries and Archives</u> by Anne R. Kenney and Stephen Chapman (pp. 140-141). A recent overview of outsourcing in general, accompanied by a brief discussion of imaging, may be found in *Library Technology Reports* (v. 34 (5), September, 1998, pp. 559+). This article suggests that institutions outsource to concentrate on core activities, save money, reduce a temporary backlog, acquire expertise, or utilize special equipment. Cost, expertise, and equipment appear to be most relevant to the establishment of the UMDL.

In-house costs – Costs of start-up purchases (equipment, hardware, and software) can be accurately estimated. The same may not be true for ongoing costs. The actual per-image cost is based on questions such as:

- 1. How many images will be scanned in the first year?
- 2. What are the staff costs?
- 3. Will the production rate increase in subsequent years?
- 4. Will per-image costs increase with more complex processing?

Outsourced costs – If expected digital outcomes can be specified, then the cost of contracting for the production of digitized images can be stated precisely. Bids (or responses to RFPs) will contain a per-page cost for creating and encoding digital representations of selected documents.

Standards and monitoring – For both in-house and outsourced processing, it is essential to clearly specify project goals, image quality, and image formats at the outset. Given the diverse nature of potential UMDL collections, it may be difficult to generate and maintain accurate and workable specifications for each of the possible types of objects to be digitized (text, grayscale, color, audio materials, manuscripts, etc.) for outsourcing vendors. In addition, it is important to note that technologies and standards will evolve over time. This will affect the decision to outsource.

The primary reasons organizations choose to outsource are cost savings and efficiency of processes (re-engineering). Since there are no known costs for UMDL projects, there is no real basis for comparing outsourcing costs to in-house costs. Similarly, re-engineering is inappropriate for the UMDL. "Before considering whether outsourcing is a viable alternative, an institution must have a good understanding of the near- and long-term goals of an imaging initiative, assess the collections to be converted and benchmark conversion requirements, define metadata requirements and users' needs, locate potential vendors, evaluate vendor claims and products, adopt policies and procedures for various functions, and define institutional and vendor responsibilities. As institutions develop a firm sense of their requirements and a confidence in what vendors can provide, outsourcing imaging services becomes a viable option." (Digital Imaging for Libraries and Archives by Anne R. Kenney and Stephen Chapman, p.141)

The library literature generally recommends outsourcing when a high production volume is wanted or needed. The Task Force does not view the UMDL as being at this stage of development. Theoretically, it would be possible to build the UMDL entirely with outsourced equipment and expertise. However, the Task Force believes that the opportunities for continued learning and retaining expertise are greater with in-house, hands-on experience.

Appendix A

Notes on Site Visits

Lessons learned at the University of Kentucky:

- They planned a digital library from the beginning; it coincided with the planning and construction of their new building.
- A key development was the hiring of Eric Weig who had recent technical expertise
- They got a grant from Inso for the DynaText software which would have cost them \$34,000
- The project that began at the University of Kentucky has now expanded into the Kentucky Commonwealth Virtual Library
- They tried to select projects that were unique to UK, e.g. digitizing a journal about Appalachia
- For their first projects they selected doable, unique projects that had the biggest chance for success
- They got faculty interested at an early stage
- They had good collaboration with subject specialists
- Start small and successful, make building blocks and grow incrementally; you do not have to have a lot of money in order to begin
- Don't put all your eggs in one basket; train others
- After you have succeeded, it is easier to get resources
- Most libraries start their "digital libraries" with commercial resources such as databases and then move to creating resources locally; UK did it the opposite way
- Buy software, don't try to write your own
- They have done projects in a variety of formats-text, audio, video, photos
- KCVL will hire a consultant to advise them on rights issues
- They are working with catalogers; there is an Internet cataloging group, they are participating in OCLC CORC

Lessons learned at the University of Michigan

- There is great value in establishing cooperative relationships with other campus entities, although it means a lot of politicking, jockeying for power and money (everyone is after the same resources)
- They began some cooperative projects, such as the work with museums on imaging, without worrying about standards. Now they have to go back and think about these issues because they affect the ability to search across collections
- Input from librarians, e.g. on content issues, is as vital as that from technical people
- In your budgets it is important to leave some "flex money" for "opportunity-based" projects
- They will share their expertise with other libraries through their SGML Server Program

- Once they started their digital library projects, other units on campus learned that they had expertise and came to them
- They wrote a number of proprietary programs themselves
- They are working with catalogers, may put some collection level records into the OPAC
- There is more out there to be done than any one player can accomplish; it is important to build a strong infrastructure and then look for ways to bring everything together (that is what the Digital Library Federation works on)
- It is good to have an annual budget of \$2,000,000

Lessons learned at the Online Archive of California (Berkeley visit)

• The mechanism they used to bring different UC campuses (other institutions as well) together for the project was to (1) do a multi-campus survey of digitizable materials, (2) establish subject categories based on survey

results, (3) initiate a project based on a subject area to which all campuses can contribute.

- Interestingly, they did NOT do any projects initially, rather, they accepted input from established centers (special collections, libraries, archives).
- Similar to Michigan, they would actually do projects for others, as well as accept outside projects adhering to standards; they could send their material to the Online archive, or they could do it themselves and send the products.
- Attempting to integrate non-UC digital collections and made ties to non-UC Institutions.
- Used the concept of "paraphrasing" photo collections: would digitize a small part of a photo collection and call the "paraphrasing" of the entire collection as fair use.
- Used EAD for creating meta-data which was developed at Berkeley.
- Vast bulk of what they've done to date is just meta-data, online finding aids of existing materials which still requires individuals to go to the actual holding sites. Digitizing the actual items en masse is just beginning.
- Philosophically determined it was more useful to promulgate finding aids to existing collections even if you couldn't afford to create and maintain digital resources, you could tell people where the "real" resources exist.

Appendix B

Webography

Colorado Digitization Project http://colorado.digital.coalliance.org/

Digital Image Access Project (Duke University) http://scriptorium.lib.duke.edu/diap/diap_info.html

Digital Initiatives at RLG http://lyra.rlg.org/digital/

the Digital Library Federation http://www.clir.org/diglib/

The Digital Library Tool Kit http://www.sun.com/products-n-solutions/edu/libraries/digitaltoolkit.html

<u>Digitising History -</u> A Guide to Creating Digital Resources from Historical Documents <u>http://hds.essex.ac.uk/g2gp/digitising_history/index.html</u>

Imaging Information http://sunsite.berkeley.edu/Imaging/

LC/Ameritech – Lessons Learned – National Digital Library Competition http://memory.loc.gov/ammem/award/lessons.html

LC's American Memory project http://memory.loc.gov/ammem/collections/

Making of America http://www.umdl.umich.edu/moa/

Moving the Digital Library from "Project" to "Production" http://jpw.umdl.umich.edu/pubs/japan-98.html#Heading19

National Digital Library Competition http://memory.loc.gov/ammem/award/98award/award98.html

Online Archive of California

http://www.cdlib.org/guides/oac/

Technical Recommendations for Digital Imaging Projects <u>http://www.columbia.edu/acis/dl/imagespec.html</u>

TEI Text Encoding in Libraries - Draft Guidelines for Best Encoding Practices <u>http://www.indiana.edu/~letrs/tei</u>

Text Encoding Initiative's Guidelines for Electronic Text Encoding and Interchange

http://www.hti.umich.edu/docs/TEI/

- University of Kentucky http://www.uky.edu/libraries/EIAMC
- University of Michigan http://www.umdl.umich.edu
- Univ. of Va. Electronic Text Center http://etext.lib.virginia.edu/