

Report on the Meeting of Experts on Digital Preservation

**U.S. Government Printing Office
Washington, D.C.
March 12, 2004**

I. Introduction and Purpose of the Meeting	1
II. Attendees	2
Participants	2
Observers.....	2
III. Summary of Discussion	3
Introductory Comments	3
Discussion.....	4
Observer Comments	6
IV. Draft Specification Document	7
Plain text	7
Illustrated Text.....	8
Color.....	8
Other Processes	8
Metadata	9
Appendix A: Project Specifications, Revised	10
Table 1. Plain Text and Grayscale Resolution	10
Table 2. Color Resolution.....	11
Table 3. Other Processes	12
Table 4. Metadata Issues.....	13
Appendix B: Sources and Resources	14

I. Introduction and Purpose of the Meeting

On Friday, March 12, 2004, Judith C. Russell, Managing Director for Information Dissemination and Superintendent of Documents, United States Government Printing Office (GPO) convened a meeting of experts on digitization and digital preservation at the GPO in Washington, DC.

The meeting was called as the first activity in an initiative with the federal depository library community to digitize the entire legacy collection of U.S. government documents currently held in depositories, estimated to be about 2.2 million items (excluding microfiche). The intent is to ensure that the collection is digitally reformatted for preservation purposes and that access copies are derived from the digitized preservation copies. GPO is committed to the preservation of this information and to making it available, in the public domain, for permanent public access.

The meeting brought together practicing experts in the field of digital format conversion and digital project development, as well as representatives of funding organizations and other coordinating organizations, to discuss the current standards and specifications for the creation of digital objects for preservation and to put forward a proposed set of minimum requirements for digitizing documents for this project.

To facilitate the discussion, GPO gathered and summarized sample specifications from various digitization projects nationwide, which are presented in Appendix A of this document.

This document summarizes the discussions and presents the resulting draft specification document that was compiled for public comment. Once the comments have been compiled and evaluated, a final document will be issued.

II. Attendees

Participants

Richard Pearce Moses	Arizona State Library, Archives & Public Records
Patricia Cruse	California Digital Library
Timothy Robson	Case Western Reserve University
Richard Urban	Colorado Digitization Program
Nancy McGovern	Cornell University
David Seaman	Digital Library Federation
Bill Comstock	Harvard University
Stuart Snyderman	Stanford University
Martha Anderson	U.S. Library of Congress
Evelyn Frangakis	U.S. National Agricultural Library; New York Public Library
Steven Puglia	U.S. National Archives & Records Administration
Martha Fishel	U.S. National Library of Medicine
Perry Willett	University of Michigan
Cathy Hartman	University of North Texas
Matthew Gibson	University of Virginia
Nolan Pope	University of Wisconsin
Meg Bellinger	Yale University

Observers

Prue Adler	Association of Research Libraries
Clifford Lynch	Coalition for Networked Information
Donald Waters	Mellon Foundation
Robert Martin	U.S. Institute for Museum and Library Services
Helen Aguera	U.S. National Endowment for the Humanities
George Farr	U.S. National Endowment for the Humanities
Larry Brandt	U.S. National Science Foundation

III. Summary of Discussion

Introductory Comments

Following welcoming remarks by Public Printer Bruce R. James, David Seaman, Director of the Digital Library Federation (DLF), spoke briefly, setting the stage for the day's discussion. He mentioned the growing need for standardization as digitization efforts grow larger and more diverse. There is a growing desire among the institutions represented by his organization for collaborative projects that will invest local resources with a larger vision for making good use of the investments already made in skills, equipment, and experience. Efforts to date have been significantly enriched through such collaboration, for example in the area of digital preservation. A great foundation has been built and future efforts will be richer for this information.

As more digital content is created and surrounded with metadata, there is more talk of collaboration and sharing of content and expertise. There's a great deal to be learned from one another, and a strong appetite for expansion of the range and scale of digitized content, demonstrated by the response to proposals such as GPO's. Seaman noted that he and the DLF members are looking forward to hearing from this group what still needs to be done, especially in relation to sharing metadata.

Superintendent of Documents Judy Russell then outlined the proposed project for scanning the legacy collection of the Federal Depository Library Program (FDLP). The project will seek the participation of a variety of partners to scan approximately 2.2 million printed publications (an estimated 60 million pages) now in the collections of federal depository libraries. Microfiche and other materials will be incorporated into the project at a later time.

This initiative will provide GPO with acceptable archival digital masters for permanent preservation as well as providing the basis for the development of a wide variety of derivative access products. This initiative will ensure that the legacy collections now available only in print and microform are fully a part of the electronic federal depository library collection of the future. It will also create opportunities for aggregation and enhancement of the digital content to create new products and services, by GPO and other federal agencies, by libraries and by the private sector.

The focus for today's meeting is the creation of digital master files, from which derivative files can be made. The legacy collection digitization initiative has the twin purpose of preserving the content permanently and improving public access through derivative files, but the first and essential step is creating appropriate digital preservation masters.

Effective coordination and the use of common specifications are essential to the success of such a large scale project with many contributors. GPO is in a position to serve as the coordinator of this project and to facilitate establishment and implementation of specifications that will establish a baseline for digital preservation master files, a minimum threshold that must be met by all participating institutions. With an appropriate specification for digital preservation masters, generations of derivative files can be produced as technology advances.

GPO and the depository library community must develop other guidance for the project as well, such as addressing metadata specifications and governance, but the digital preservation master file specification is a logical place to begin.

The legacy collection digitization initiative is intended to create a comprehensive digital collection. The project will be like the development of a quilt by a community. The digitization will be done in many places and consist of many pieces, but the coordination and consistency of the digitization and the associated metadata will make it a collection—a quilt, not just a patchwork of scattered pieces. If we scan the material correctly the first time, we will only need to do it once and we should be able to use the resulting files for other outputs.

Discussion

To prepare for the meeting, GPO gathered digitization specifications from the participating institutions and prepared a set of tables to facilitate discussion of the information that was gathered. The projects and policies illustrated in the tables (see Appendix A) reflect some of the work being undertaken by institutions represented at this meeting. The list is not exhaustive or comprehensive, but is intended to convey a cross section of practices currently in use by leaders in preservation digitization. It also served as a starting point for discussion of specifications for the GPO project to digitize the legacy collection of Federal government documents.

Tables 1 and 2 focus on scanning specifications, with plain text and grayscale resolution in Table 1, and color resolution in Table 2. Table 3, Other Processes, addresses image cropping and optical character recognition (OCR). Table 4 summarizes approaches to metadata.

The goal for the day was to arrive at a baseline specification to be utilized by institutions cooperating in the GPO project. The purpose of the specification is to assure the creation of digital master files suitable for preservation and of sufficient quality to support the production of multiple generations of derivative products for access. The projects and policies of the experts' institutions illustrated the current range of variables in specifications for existing projects in the field. While no absolute consensus emerged from viewing the data in the tables, it was immediately clear that the variables fall within a fairly narrow range. Significant divergences were attributed either to the characteristics of a particular project, or to overall policies that were set at an earlier time and not recently revised to reflect more current practices.

The experts discussed the groupings of variables in each table in turn and ultimately arrived at suggested values for a new row on each table, labeled "GPO," which was understood to constitute a proposed minimum specification for the digitization of the legacy collection. In discussing technical issues around resolution, file type and compression, the experts were asked to concentrate on the creation of objects specifically envisioned for preservation, with the assumption that copies for access will be derived from the preservation masters and will have different specifications.

In arriving at proposed minimum specifications for preservation quality digital master files, a number of observations were made:

The institutions and individual projects represented in the tables have arrived at their functional specifications for digital image files based on the specific characteristics of the original documents or objects to be scanned, the nature of the intended use, and the requirements/limitations locally for handling, processing, and storing objects. Each institution has evolved a process for evaluating the specific factors when setting the individual project plans.

Concerns were expressed about the dangers of individual partners in such a large and complex project following specifications without exercising informed judgment as to specific characteristics of the source material and desired outcomes. Factors such as intended use, condition or characteristics of the printed documents, whether the original object is to be preserved or not, and particular needs of the intended user community all must inform the development or modification of digitization specifications. The lack of such context might lead to the creation of digital masters which would be minimally functional, but still fall short of being a faithful digital reproduction. The need for informed judgment in planning and coordinating this effort will need to be articulated in the project's governance framework, which will be taking shape over the next several months.

While all agreed that an ideal specification is a valuable tool in coordinating a large cooperative project of this kind, some expressed concern that GPO not set a bar so high that it would exclude smaller institutions from making contributions that they may be in a unique position to make. There was a clear sense that particular projects whose specifications deviate from those GPO implements will need to be evaluated for inclusion and that alternative specifications may need to be developed for specific formats, such as posters or large maps. There must be a mechanism in place, alongside the baseline specification, which allows for variants to be proposed to GPO, evaluated, and approved or rejected. This will also be a function of the governance structure.

The goal must be to produce digital objects which can be preserved as well as adapted into copies for access, and to accomplish this "with the first scan" since the resources may not be available to rescan the same content at a later time using higher specifications. Scanning needs to be done initially at a threshold that will be, to the best of our ability to predict, adequate to support preservation as well as future iterations of derivative products through which GPO will coordinate access to the public through the FDLP. This will necessitate both bibliographic control and a registry of digitization projects.

In some of the sample projects, images were cropped and/or de-skewed based on the specific requirements of the institution (e.g., the output needed to preserve the "look and feel" of the original). It was noted that with some "rare books" an effort was made to show the page edging. In general, de-skewing is not appropriate for digital preservation masters and should be avoided since it causes re-sampling errors. It is preferable to de-skew the derivative files as they are created.

While the specifics of Optical Character Recognition (OCR) specifications of the participating institutions are often applicable to access products rather than to preservation masters, it was clear that planning for a "scan once" scenario for the legacy digitization project must, again to the best of our ability, include OCR and support subsequent processes to derive access files. The

choice that must be made is whether the OCR is merely for retrieval purposes or intended to accurately represent or replace text.

Furthermore, while the OCR results will be preserved, they are distinct from the preservation master images and the two are linked by metadata. OCR and scanning may, in fact, be completely separate processes, performed at different times or under different settings, even by different institutions. OCR will work best with a good quality TIFF. However, the issues of type size and contrast, for example, will affect the ability of the OCR software to accurately interpret the image. OCR is presently most effective for text with clean, uniform type. Statistical and other complex tabular formats, as well as complicated layouts in some legal publications, pose technical challenges that will probably be addressed as technology advances. For example, legal resources and scientific materials utilize very specific symbols that are not well read by OCR. Additional care needs to be taken with doing OCR on these resources. Various approaches and the advances in OCR software were mentioned, and the general opinion is that high accuracy rates are ever more attainable.

The final topic of conversation for the day was metadata. It was acknowledged by GPO that metadata requirements for the proposed project will need to be discussed in depth, probably at another gathering of experts, and vetted by the community. Participants stressed that metadata is the “glue” that will hold the project together, absolutely key to the success of any digitization project, particularly one of this scale, and that metadata is potentially the most expensive part of this process. Descriptive, administrative, and technical/structural metadata must be of sufficient depth to support the various processes, uses, and preservation of the material.

A number of metadata efforts are underway in the larger community, such as the OCLC PREMIS (Preservation Metadata Implementation Strategies) the Library of Congress’ METS (Metadata Encoding and Transfer Standard) and MODS (Metadata Object Description Schema), and numerous local schema; it may be some time before they converge, so it will be necessary to adopt best practices and review these decisions over time.

Observer Comments

In addition to the invited experts, a number of observers attended and commented on the GPO project and the proposed specifications. Their comments included the following:

There is a need for large-scale projects of this type and for wide ranging cooperation to accomplish them. The GPO project will become a model for projects to digitize large library collections of materials other than Federal government information.

This project will require an integration of a wide variety of textual and tabular (financial and statistical) data, as well as non-textual data and images, including various kinds of color representation.

In defining the project there is a need to address the tension between object-by-object (or even page-by-page) variation, which may affect downstream outcomes such as indexing or OCR, and the need for a production line approach to accomplish the large volume of work. Since the

objective is to create high quality preservation master files, the emphasis must be placed on the varied characteristics of specific documents, or portions of documents, rather than on high volume, low cost throughput.

It was noted that other organizations, such as JSTOR, that have experience in large-scale digitization efforts need to be brought to the table. Established programs, including initiatives of Coalition for Networked Information (CNI) and Digital Library Federation (DLF), and projects such as JSTOR may provide paradigms that would assist in the planning and implementation of this project. Such initiatives have created demand and competition in the marketplace and have influenced the way in which funding agents such as National Science Foundation (NSF), National Institutes of Health (NIH), and the Andrew W. Mellon Foundation evaluate and encourage solutions and proposals.

IV. Draft Specification Document

This section defines benchmark specifications to guide the proposed digitization of an estimated 2.2 million printed publications currently housed in Federal depository libraries. The specifications will assure that scanned images will be suitable for long-term preservation and appropriate for processing into access editions, which are searchable.

The goal of the GPO project is to produce faithful digital reproductions sufficient to enable preservation and facilitate the production of a variety of derivatives for access. A fuller definition of faithful digital reproductions is contained in *Benchmark for Faithful Digital Reproductions of Monographs and Serials*, version 1, December, 2002 issued by the Digital Library Federation Benchmark Working Group (2001-2002) <<http://www.diglib.org/standards/bmarkfin.htm>>, with which these specifications closely conform.

Page dimensions, font sizes, and other characteristics may influence resolution and/or file size and manageability in particular instances. The ultimate goal of a faithful digital reproduction for preservation and derivation purposes may be achieved with differing specifications. If an individual project deviates significantly from these minimums, adequate documentation describing and justifying the deviation should be provided. A mechanism for submitting such documentation and seeking approval for inclusion of projects with differing specifications will be addressed in the governance model for the GPO initiative.

Digital preservation masters for the GPO legacy collection digitization project should meet the following minimal requirements:

Plain text

Preferred: 400 dpi, 8-bit grayscale single page TIFF, uncompressed or employing lossless compression.

Minimum acceptable: 600 dpi, bi-tonal, single page TIFF, uncompressed or employing lossless compression.

The minimum specification may be employed in cases in which the original document includes type which is of uniform size, 8 points or larger, is in sharp contrast to the background, and in which the background is clear and uniform.

The preferred specification must be employed in cases in which the original document includes type, which is 7 points or smaller, where there is wide variation in type sizes and style, where there is variation in print contrast or intensity, or where the background is uneven, damaged, discolored or stained.

Illustrated Text

400 dpi, 8-bit grayscale single page TIFF, uncompressed or employing lossless compression.

Images may be cropped and/or de-skewed according to need, as dictated by the nature of the original.

Color

400 dpi, 24-bit RGB single page TIFF, uncompressed or employing lossless compression.

Images may be cropped and/or de-skewed according to need, as dictated by the nature of the original.

Preferred: Color is accurate to original without color management, or output includes an ICC display profile for color management.

Other Processes

Scanning should create a single TIFF file per page, capturing all pages including those intentionally left blank. Scans should not be cropped or de-skewed for the preservation master to avoid a loss of resolution. If necessary, such corrections can and should be made in the derivative files. In addition, a machine-readable text is required, whether created by optical character recognition (OCR) or other means such as re-keying. Uncorrected OCR files should be retained for preservation purposes. OCR files should be corrected to at least 99% accuracy for access.

The desired outcome or use of derivative products will dictate the application of various processes. In situations for which there is a desire to retain the visual integrity of the original page for presentation purposes, with machine-readable text in the background, adjustment for skew and cropping of edges is not appropriate. If, however, the output is to be strictly an electronic edition, with text rendered without reference to the original visual makeup of the page, or if the visual image is purely a point of reference (such as a thumbnail), cropping and de-skewing may improve the quality and accuracy of the text rendering.

Metadata

It was agreed that good metadata is essential to the success of such a complex digitization project. As a general guideline, images produced should be accompanied by detailed metadata, which describes the content and identifies the technical environment and specifications used to create the images, as well as establishing the relationships between and among preservation objects and any derivatives for access. For the digital preservation masters, a fully formed submission information package should be produced as described in the *Reference Model for an Open Archival Information System* (ISO 14721:2003) <http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=24683>, with attributes to be defined by GPO and its partners.

Appendix A: Project Specifications, Revised

Table 1. Plain Text and Grayscale Resolution

Institution	Project name (if applicable)	Plain Text Resolution	Plain Text Bit Depth	Plain Text Compression	Illustrated text b/w Resolution	Illustrated text b/w Bit Depth	Illustrated text b/w Compression	Format of preservation object
California Digital Library		300-600 ppi	bitonal	lossless	300-600 ppi	8-bit grayscale	lossless	TIFF
Case Western Reserve University Libraries		400 dpi	8-bit grayscale	none	600 dpi	8-bit grayscale	none	TIFF
Colorado Digitization Program		600 dpi	1 bit bitonal	n/s	3000-5000 pixels across long dimension	8 bit grayscale or 24 bit color	none or lossless	TIFF
Cornell University		600 dpi	bitonal	Lossless	400-600 dpi	8 bit grayscale or 24 bit color	None or lossless	TIFF
Digital Library Federation		600 dpi	1 bit or bitonal	none	300 dpi	8 bit grayscale	none or lossless	TIFF
Harvard	printed monographs and serials	600 dpi	bitonal	none	600 dpi	8-bit grayscale or 24-bit color	Bitonal: (TIFF Group IV) Color or grayscale: uncompressed (TIFF)	TIFF
Library of Congress	American Memory	200-400 dpi	bitonal	ITU Group IV	200-300 dpi	8-bit grayscale	JPEG	TIFF
National Agricultural Library		600 dpi	bitonal	ITU Group IV	600 dpi	bitonal, single bit	ITU Group IV	TIFF
National Archives and Records Administration		300-600 dpi	bitonal or 8 bit grayscale	none	300 dpi	8 bit grayscale or 24 bit colr	none	TIFF
National Library of Medicine	PubMed Central Back Issues	600 dpi	1 bit	CCITT Group 4	300 dpi	8 bit grayscale	Packbits	TIFF
Stanford University	Joint Comm on Atomic Energy Hearings	400 dpi	8-bit grayscale	LZW	400 dpi	8-bit grayscale	LZW	TIFF
University of Michigan		600 dpi	1 bit	ITU Group IV	400 dpi	8-bit grayscale	none	TIFF
University of North Texas	Travel at the Turn of the Century				400 dpi	grayscale	none	TIFF
University of Virginia Electronic Text Center		400-600 dpi	24-bit color	none	400-600 dpi	24-bit color	none	TIFF
University of Wisconsin Libraries		600 dpi	bitonal	CCITT Group IV	300-400 dpi	4- or 8- bit grayscale	none	TIFF
Yale University	Economic Growth Center Digital Library (EGCDL)	300 dpi	24-bit color 1-bit black and white	None	300 dpi	24-bit color or 1-bit black and white	none	TIFF
GPO	Legacy Digitization Project	400 dpi (preferred); 600 dpi (acceptable)	8-bit grayscale (preferred); bitonal (acceptable)	uncompressed or lossless	400 dpi	8-bit grayscale	uncompressed or lossless	TIFF

Table 2. Color Resolution

Institution	Project name (if applicable)	Color Resolution	Color Bit Depth	Color Compression	Format of preservation object
California Digital Library		300-600-ppi	24-bit	lossless	TIFF
Case Western Reserve University Libraries		600 dpi	24-bit	none	TIFF
Colorado Digitization Program		3000 pixels across long dimension	24 bit color	none or lossless	TIFF
Cornell University		400-600 dpi	24 bit color	none or lossless	TIFF
Digital Library Federation		300 dpi	24 bit	none or lossless	TIFF
Harvard	printed monographs and serials	250 – 400 dpi depending upon size of characters or other meaningful details	24 bit	none	TIFF
Library of Congress	American Memory	200-300 dpi	24-bit	JPEG	TIFF
National Agricultural Library		400 dpi	24 bit	none	TIFF
National Archives and Records Administration		300 dpi	24 bit color	none	TIFF
National Library of Medicine	PubMed Central Back Issues	300 dpi	24 bit color	Packbits	TIFF
Stanford University	Joint Committee on Atomic Energy Hearings	n/a	n/a	n/a	TIFF
University of Michigan		400 dpi	24 bit	none	TIFF
University of North Texas	Travel at the Turn of the Century	400 dpi	24 bit	none	TIFF
University of Virginia Electronic Text Center		400-600 dpi	24 bit	none	TIFF
University of Wisconsin Libraries		300-400 dpi	24 bit	none	TIFF
Yale University	Economic Growth Center Digital Library (EGCDL)	300 dpi	24 bit	none	TIFF
GPO	Legacy Digitization Project	400 dpi	24 bit	uncompressed or lossless	TIFF

Table 3. Other Processes

Institution	Project name (if applicable)	Image cropped (Yes/No)	Image deskewed (Yes/No)	OCR performed (Yes/No)	OCR Corrected (Yes/No)	Desired OCR accuracy
California Digital Library		NO	NO	NO	NO	n/s
Case Western Reserve University Libraries		NO	YES	NO	NO	
Colorado Digitization Program		NO	NO	NO	NO	n/s
Cornell University		YES	YES	NO	NO	n/s
Digital Library Federation		NO	NO	NO	NO	
Harvard	printed monographs and serials	YES	YES	YES	NO	n/s
Library of Congress	American Memory	NO	YES	YES	NO	n/s
National Agricultural Library		NO	NO	YES	NO	n/s
National Archives and Records Administration		NO	NO	NO	NO	n/s
National Library of Medicine	PubMed Central Back Issues	NO	YES	YES	NO	n/s
Stanford University	Joint Comm on Atomic Energy Hearings	YES	NO	YES	NO	97-99%
University of Michigan		YES	YES	YES	NO	
University of North Texas	Travel at the Turn of the Century	NO	NO	YES	YES	99.998%
University of Virginia Electronic Text Center		NO	NO	NO	NO	n/s
University of Wisconsin Libraries		YES	YES	YES	NO	n/s
Yale University	Economic Growth Center Digital Library (EGCDL)	NO	YES	YES	NO	98.5%
GPO	Legacy Digitization Project	NO	NO	YES	NO	99%

Table 4. Metadata Issues

Institution	Project name (if applicable)	Treatment or Disposition of Original	Metadata Expressed in	Metadata Encoded as	Metadata Types
California Digital Library		n/s	Local Schema	n/s	All
Case Western Reserve University Libraries		Varies by project	n/s	n/s	Varies by project
Colorado Digitization Program		n/s	n/s	n/s	All
Cornell University		n/s	n/s	n/s	Varies by project
Digital Library Federation		n/s			All
Harvard	printed monographs and serials	Varies	Local Schema (Administrative and Structural metadata) MoAII (structural metadata)	XML	Administrative/technical/structural
Library of Congress	American Memory	preserved	Local schema	MARC, SGML	All
National Agricultural Library		n/s	Local schema	n/s	All
National Archives and Records Administration		n/s	n/s	n/s	n/s
National Library of Medicine	PubMed Central Back Issues	Original disposed	n/s	n/s	n/s
Stanford University	Joint Committee on Atomic Energy Hearings	Retained	Local schema	n/s	All
University of Michigan		retained			All
University of North Texas	Travel at the Turn of the Century	Preserved	local schema	xml	All
University of Virginia Electronic Text Center		n/s	Local		Administrative/technical/preservation
University of Wisconsin		Varies by project	Dublin Core	SGML	Varies by project
Yale University	Economic Growth Center Digital Library (EGCDL)	Returned	Dublin Core and Data Documentation Initiative	XML	Descriptive
GPO	Legacy Digitization Project	Varies by project	TBD	MARC, Other	TBD

Appendix B: Sources and Resources

Institution	Sources or Documents Cited	Web Address (if applicable)
California Digital Library	California Digital Library Digital Image Format Standards (July 9, 2001)	www.cdlib.org/about/publications/CDLImageStd-2001.pdf
Case Western Reserve University Libraries	Correspondence	
Colorado Digitization Program	Western States Digital Imaging Best Practices, Version 1.0 (January 2003)	www.cdphheritage.org/resource/scanning/index.html
Cornell University	Establishing a central depository for preserving digital image collections, part 1: responsibilities of transferee, Version 1.0 (March, 2001)	www.library.cornell.edu/preservation/IMLS/image_deposit_guidelines.pdf
Cornell University	Moving Theory Into Practice Digital Imaging Tutorial (2002-2003)	www.library.cornell.edu/preservation/tutorial/contents.html
Digital Library Federation	Benchmark for Faithful Digital Reproductions of Monographs and Serials, Version 1 (December 2002)	www.diglib.org/standards/bmarkfin.htm
Digital Library Federation	Digital Library Standards and Practices	www.diglib.org/standards.htm
Harvard University	Library Preservation at Harvard: Digitization	preserve.harvard.edu/resources/digital.html
Library of Congress	Conversion Specifications for Contracted Scanning Services	memory.loc.gov/ammem/techdocs/conversion.html
National Agricultural Library	Statement of Work: National Agricultural Library Preservation Digitizing (June 2001-September 30, 2002)	
National Archives and Records Administration	Records Management: Transfer Permanent Electronic Records to NARA	www.archives.gov/records_management/initiatives/erm_products.html www.archives.gov/records_management/initiatives/scanned_textual.html
National Library of Medicine	Correspondence	
Stanford University	Correspondence	
University of Michigan	Correspondence	www.umdl.umich.edu
University of North Texas	Correspondence	
University of Virginia Electronic Text Center	Archival Digital Image Creation	etext.lib.virginia.edu/helpsheets/specscan.html
Yale University	Correspondence	ssrs.yale.edu/egcdl
University of Wisconsin Libraries	Correspondence	http://uwdcc.library.wisc.edu/documentation.html