RLG Model Request for Proposal (RFP)

for Digital Imaging Services

The Research Libraries Group, Inc.

This document was prepared by Cornell University under contract to RLG. It is a sample (or model) of how the *RLG Guidelines for Creating a Request for Proposal (RFP) for Digital Imaging Services* can be adapted to meet the needs of almost any digital imaging project.

CORNELL UNIVERSITY LIBRARY

REQUEST FOR PROPOSAL (RFP)
FOR DIGITAL IMAGING PRODUCTION SERVICES

July 27, 1998

Please note the following important dates associated with the RFP:

Monday, August 10, 1998, 5PM
Notify Cornell University Library indicating whether you will be responding to this RFP and wish to receive the sample volumes for scanning.

Monday, August 17, 1998
Cornell University Library will supply each vendor with sample volumes and accompanying instructions for the Preliminary Production Test.

Monday, September 21, 1998, 5PM
Vendors must deliver to Cornell University their RFP response, plus the sample volumes and the products of the preliminary test.

Week of October 5, 1998
RFP responses evaluated and Vendor(s) chosen.

Week of October 19, 1998
Contract(s) awarded, and unsuccessful vendors notified.

Friday, November 6, 1998
Shipment #1 sent to vendor(s).

TABLE OF CONTENTS

CORNELL UNIVERSITY LIBRARY

REQUEST FOR PROPOSAL (RFP)
FOR DIGITAL IMAGING PRODUCTION SERVICES

I. INTRODUCTION

Cornell University has been awarded a grant from the Jane Smith Foundation to undertake a project to digitize research library materials related to 19th century New York State and local history. Up to 1,000 volumes (representing 300,000 images) will be selected, scanned, SGML encoded, and made available online via a web-based interface. Cornell is seeking a vendor or group of vendors to perform the digital conversion, metadata creation (directory structuring, file naming, header information, indexing), text conversion (OCR), and SGML encoding/tagging. Cornell will create on-the-fly access derivatives from the master images, therefore this proposal does not include any guidelines for the creation of derivatives.

The purpose of this Request for Proposal (RFP) is to enable Cornell to identify and select a vendor (or vendors) to produce high quality bitmapped images of books and serial literature that can serve as replacements for the deteriorating originals. During a 10 month period (October 1998 - July 1999), Cornell will identify, prepare, disbind, and ship up to 1,000 volumes to the vendor(s) for the production of digital images. The vendor(s) will scan all pages as bitonal (one-bit) images with no gray scale or color scanning at a resolution of true or interpolated 600 dpi. Halftone illustrations and finely detailed line art should be captured so as to suppress aliasing distortions and moiré patterns. Images will be produced in TIFF 6.0 and compressed using ITU Group 4 (formerly CCITT), Intel byte order.

The vendor will also produce associated metadata. In addition, the vendor will convert the digital images to text files via Optical Character Recognition (OCR) process, and encode/tag the resulting text files using Standard Generalized Markup Language (SGML). Cornell also requires the vendor to produce 600 dpi paper prints of the digital files on acid free paper to facilitate quality inspection of the scanned images and to create hard copy replacements for the deteriorating originals.

II. PROJECT OVERVIEW

Cornell offers strong computing and networking resources and has made substantial contributions to the development of digital library initiatives, including participation in multi-institutional projects such as the Making of America (MOA) project. MOA was a joint initiative with the University of Michigan to preserve and make accessible through digital technology a significant body of primary resources related to American social history. Cornell is a member of the Digital Library Federation, and brings to this project valuable expertise in conversion and access to full text materials. Cornell has experience with in-house scanning and is strongly supportive of the development of high quality, third party conversion services that can meet the needs of research libraries.

III. PROPOSAL SUBMISSION INFORMATION

A. Requirements for Submission

1. Submitted proposals must be addressed to:

   Cornell University Purchasing Department
   Attn: Glenn Morey, Purchasing Agent
   120 Maple Avenue
   Ithaca, New York 14850
   (Phone: 607-255-7402/Fax: 607-255-9450)

Technical inquiries should be directed to:

Anne R. Kenney
Associate Director
Department of Preservation & Conservation
Phone: 607-255-6875
Fax: 607-255-9346
Email: ark3@cornell.edu

2. Cornell University *will not* accept proposals and sample digital files received after **5:00 p.m. on Monday, September 21, 1998**.

3. Failure of bidding vendor to follow all proposal submission instructions will be cause for Cornell to disqualify the proposal.

4. All expenses for the preparation of proposals are the responsibility of the bidding vendors.

B. Confidentiality and Retention of Proposals

All proposals submitted become the property of Cornell University. Cornell University will make all reasonable efforts to maintain proposals in confidence and will release proposals only to personnel involved with the evaluation of the project. Cornell will share the names and addresses of all respondents with each potential bidder so as to facilitate collaborative responses. If a vendor fails to respond to any portion of the RFP, Cornell will interpret this as the vendor’s inability to meet specific requirements.

C. Amendments to the RFP

If this RFP is amended by Cornell University, the amendment will be sent to each vendor in writing. Vendors are required to acknowledge each amendment received in writing to the address listed in the RFP.

D. Exceptions to the RFP

Cornell University requires each vendor to provide a list of exceptions taken to this RFP. Any exceptions taken must be identified and explained in writing. An exception is defined as the vendor's inability to meet a mandatory requirement in the manner specified in the RFP. If the vendor provides an alternative solution when taking an exception to a requirement, the benefits of this alternative solution must be explained.

E. Vendor Communications During the RFP Process

1. From the RFP issue date and thereafter, communications between vendors and Cornell University may be in writing, by telephone, email, or by fax. All questions concerning the RFP must be made in writing and must reference the RFP page number and section number. Questions should be concisely stated and be numbered in sequential order. Answers will be returned in writing by Cornell University as quickly as possible. Cornell University will make questions and answers available to all bidding vendors.

2. The project staff may hold vendor conference calls in October 1998 to address questions and to resolve procedural problems that may have developed during the RFP process.

All technical inquiries related to scanning, metadata creation, OCRing, SGML encoding, and printing should be directed to:

Anne R. Kenney, Associate Director
Department of Preservation and Conservation
Phone: 607-255-6875
Fax: 607-255-9346
Email: ark3@cornell.edu

All other questions should be sent to:

Cornell University Purchasing Department
Attn: Glenn Morey, Purchasing Agent
120 Maple Avenue
Ithaca, New York 14850
Phone: 607-255-7402 Fax: 607-255-9450

F. Prime Vendor Relationship

1. Cornell University intends to purchase services from the vendor(s) of the winning proposal(s)--to be known as the Prime Vendor(s). The vendor(s) selected will be responsible for service performance.

2. Subcontracting

Subcontracting of image scanning, metadata creation, OCR, SGML encoding, paper print creation, and inspection may be allowed under this agreement with prior written approval, but Cornell University reserves the right to request information about any subcontracting relationship. In the event of a subcontracting arrangement, the Prime Vendor assumes all responsibility for work performed by the subcontractor.

G. Preliminary Production Test

Cornell will require the bidding vendors to participate in a preliminary production test as part of the RFP process, including scanning, metadata creation, text conversion (OCR), and SGML encoding. The vendor will also provide printouts of the scanned images. All costs associated with the sample test will be borne by the vendor.

On **August 17, 1998**, Cornell will mail each vendor a sample of printed materials that are typical of the material to be scanned in this project with accompanying instructions for metadata creation, OCRing, and SGML encoding. Vendors will receive material of similar volume and composition.

The sample will include one technical target, one book, and one serial volume that are representative of the range of document attributes (text, line art, and halftones) typical of the material to be scanned. In addition, vendor(s) will be provided with an electronic worksheet that is composed of scanning and metadata information specific to the sample documents. This test will be used to evaluate the vendor's capability to:

- scan and generate high-quality and properly formatted TIFF output files
- maintain a proper correspondence between physical page numbers and image numbers within a multipage document
- adhere to the directory structuring, file naming, TIFF header information, and indexing conventions identified by Cornell
- create checkmd5.fil, scandata.txt, dataset.toc files with the contents as specified in the proposal

- generate accurate SGML encoded files (and accompanying auxiliary files) using the specified Document Type Definition (DTD)
- inspect and store output files on CD-R
- deliver high quality (600 dpi) paper prints of the scanned images

Each vendor will deliver to Cornell University the original volumes, test digital files, metadata, SGML encoded files, and the prints along with the RFP response by **5 p.m. on September 21, 1998**. Vendors must respond in writing to Cornell University to suggestions and modifications based on test results.

PLEASE INDICATE BY MONDAY, AUGUST 10 BY 5PM WHETHER YOU UNDERSTAND THE REQUIREMENTS OF THE RFP AND YOU WILL BE RESPONDING TO IT, AND WHETHER YOU WISH TO RECEIVE THE SAMPLE VOLUMES FOR THE PRELIMINARY PRODUCTION TEST. USE ATTACHMENT A FOR FAXING RESPONSE.

H. Vendor Award

Cornell University reserves the right to choose the vendor(s) according to the evaluation criteria set forth in Section IV of the RFP. Cornell University also reserves the right not to make an award if it is deemed that no single proposal fully meets the requirements of this RFP. The successful vendor(s) will be notified during the week of October 5, 1998.

The vendor(s) chosen for award should be prepared to have the winning proposal incorporated, along with all other written correspondence concerning this agreement, into a contract and/or purchase order for services. Any false or misleading statements found in the proposal will be grounds for disqualification.

Unsuccessful vendors will be notified in writing by the week of October 19, 1998. Cornell University will answer written questions concerning the evaluation of the vendor's proposal.

IV. EVALUATION GUIDELINES FOR VENDOR PROPOSALS

Cornell University will evaluate vendor proposals using the following guidelines, and will assign a numerical "Range of Compliance" number to each from 0 to 5 (with 0 as "Not Compliant"). This "Range of Compliance" rating will provide proposal evaluators with a clearer insight into the overall strengths and weaknesses of any one vendor:

- Understanding of and compliance with the requirements of the RFP
- Excellence of response
- Ability to meet Preliminary Digital Production Test requirements
- Satisfying mandatory technical requirements
- Ability to meet the required scanning, metadata creation, text conversion, SGML encoding, and printout production within the required project timetable
- Fair pricing of the proposal relative to other proposals received
- Qualifications and experience, evidenced by customer references, resumes of key personnel, etc.
- Guarantee of work, and the nature and extent of vendor support
- Financial stability and other various business issues as outlined in the RFP

V. PROJECT TIMETABLE

The following outlines the project timetable for the services included in this RFP. Please note that Cornell will require the chosen vendor(s) to produce digital files for up to 50,000 pages in monthly shipments within a period of 9 months (November 1998 through July 1999).

Turn-around time for each shipment will be approximately 1 (one) month, counted from the date of receipt of physical volumes at the vendor and the date of delivery of the products (images, metadata, SGML encoding, printouts) to Cornell. Turn-around time for corrections for each shipment will be approximately 2 (two) weeks, counted from the date of notification or receipt of the unacceptable products to the date of delivery of the corrected products to Cornell.

Week of:

Week of October 19, 1998	Vendor(s) chosen and notified, and unsuccessful vendors notified.
November 6, 1998	Shipment #1 consisting of 5,000 pages sent from Cornell to vendor(s).
November 27, 1998	Shipment #1 products returned from vendor(s) to Cornell.
December 10, 1998	Cornell and Vendor conference call to evaluate Shipment #1.
December 11, 1998	Digital images, metadata, and SGML encoding for Shipment #1 inspected at Cornell. Vendor(s) notified of rejected products (images, metadata, SGML encoding, printouts) with retakes due January 8 .
December 14, 1998	Shipment #2 consisting of 40,000 pages sent from Cornell to vendor(s).
January 8, 1999	Shipment #2 products returned from vendor(s). Shipment #3 consisting of up to 50,000 pages sent to vendor(s) from Cornell.
January 22, 1999	Digital images, metadata, and SGML encoding for Shipment #2 inspected at Cornell. Vendor(s) notified of rejected products (images, metadata, SGML encoding, printouts) with retakes due February 5.
January 27, 1999	Cornell and vendor conference call to evaluate progress, and identify and address production problems experienced by either parties.
February 5, 1999	Shipment #3 products returned from vendor(s). Shipment #4 consisting of up to 50,000 pages sent to vendor(s) from Cornell.
February 19, 1999	Digital images, metadata, and SGML encoding for Shipment #3 inspected at Cornell. Vendor(s) notified of rejected products (images, metadata, SGML encoding, printouts) with retakes due March 8.

March 8, 1999
Shipment #4 products returned from vendor(s). Shipment #5 consisting of up to 50,000 pages sent to vendor(s) from Cornell.

March 19, 1999
Digital images, metadata, and SGML encoding for Shipment #4 inspected at Cornell. Vendor(s) notified of rejected products (images, metadata, SGML encoding, printouts) with retakes due April 2.

April 2, 1999
Shipment #5 products returned from vendor(s). Shipment #6 consisting of up to 50,000 pages sent to vendor(s) from Cornell.

April 5, 1999
Cornell and Vendor conference call to evaluate progress and identify and address production problems experienced by either parties.

April 23, 1999
Digital images, metadata, and SGML encoding for Shipment #5 inspected at Cornell. Vendor(s) notified of rejected products (images, metadata, SGML encoding, printouts) with retakes due May 7.

May 7, 1999
Shipment #6 products returned from vendor(s). Shipment #7 consisting of up to 50,000 pages sent to vendor(s) from Cornell.

May 21, 1999
Digital images, metadata, and SGML encoding for Shipment #6 inspected at Cornell. Vendor(s) notified of rejected products (images, metadata, SGML encoding, printouts) with retakes due June 4.

May 24, 1999
Cornell and Vendor conference call to evaluate progress and plan the completion of the project.

June 7, 1999
Shipment #7 products returned from vendor(s).

June 21, 1999
Digital images, metadata, and SGML encoding for Shipment #7 inspected at Cornell. Vendor(s) notified of rejected products (images, metadata, SGML encoding, printouts) with retakes due July 9.

July 9, 1999
All retakes received by Cornell.

PLEASE PROVIDE A WRITTEN CONFIRMATION OF YOUR ABILITY TO ADHERE TO THIS SCHEDULE. IF THE NUMBER OF PAGES/SHIPMENT EXCEEDS YOUR CAPABILITIES, INDICATE HOW MANY PAGES/SHIPMENT YOU WOULD BE ABLE TO HANDLE.

VI. PROJECT RESPONSIBILITIES: CORNELL UNIVERSITY

Cornell project staff will be responsible for selecting, preparing, and shipping to the vendor(s) up to 1,000 volumes (representing up to 300,000 pages) and performing quality control of image capture, metadata, and SGML encoding. The following provides a description of Cornell's project responsibilities, as well as the institutional infrastructure within which the vendor(s) will be working.

A. Selecting Volumes for Scanning.

Cornell University will select up to 1,000 volumes of monograph and serial literature. The materials will consist of volumes ranging in size from roughly 4" x 6" up to 11" x 17" and will average 300 pages, although individual volumes may range from under 100 pages to over 900 pages (less than 1000 pages). The material will contain halftones, line drawings, engravings, charts, tables and other black and white illustrations, as well as occasional foldouts of varying sizes. Any color illustrations present in the material may be rendered in black and white.

B. Preparing Materials for Scanning and Metadata Creation

Project staff members will:

1. Perform page by page collation of each volume to ensure correct order, completeness, legibility of text, and to note any irregularities.

2. Disbind volumes, trim binder's margin to be parallel to the body of the text, repair badly torn pages, and order replacements for missing pages through interlibrary loan. Annotations and marginalia that do not obscure text will be left intact. Any pages that are not to be scanned (therefore not indexed, OCRed or SGML encoded) will be clearly marked. Special instructions and flags shall be provided for these and other anomalies in the volume (e.g., foldouts).

3. Prepare production note that will serve as the first page of each volume.

4. Supply vendor with the following information for each volume in machine-readable form (thereafter referred to as "worksheet"):
   - Total number of pages expected to be scanned
   - Location (starting and ending page numbers) of significant reference structures (Document Structure Labels) that may be present within the work (including but not necessarily limited to title page, table of contents, lists of illustrations, indexes, bibliography or reference listings)
   - A root identifier on which to base file naming for all the pages in the volume
   - Basic bibliographic data (author, title, publisher, date of publication, series notes, subject headings) extracted from NOTIS database and converted to labeled ASCII format
   - Any anomalies such as missing pages

5. Design and construct a database to maintain the aforementioned worksheets electronically. It is highly desired by Cornell to share this database so that the vendor can enter information such as scanner settings used into this shared database. This approach will minimize errors caused by maintaining separate recording systems, and will allow the maintenance of metadata, and production and quality control information in the same database.

C. Shipping Materials to Vendor

Materials will be sorted and packed by Cornell, and costs associated with the shipping and handling of source materials to the vendor(s) within the United States shall be borne by Cornell. A packing slip will be included with each box. Each volume will be clearly marked and accompanied with a project worksheet. In the case of serials, an intellectual volume consisting of all relevant issues may be bound in more than one physical volume. The relevant physical volumes will be identified (e.g., volume 1 of 3) and bundled together. Each box will be clearly marked: "Send volumes back to Cornell when digital files are shipped" or "Keep volumes at (vendor) when digital files are shipped."

D. Performing Quality Control of Image Capture

Although the vendor will be required to perform rigorous quality control procedures (see Section VII B5, Vendor Quality Control Requirements), Cornell will also perform a detailed inspection of the digital files, metadata, text conversion, and SGML encoding.

Vendor(s) will ship digital files, metadata, SGML encoded files, and the printouts of the images to Cornell for inspection according to specifications noted below under Vendor Responsibilities. Cornell will verify media integrity and file readability for all digital files, and will run automated scripts to verify the directory structures and file naming. Cornell staff will perform 100% quality control of all image files by inspecting the printouts produced by the vendor. Technical targets and a sample of the digital images will be viewed on-screen at full resolution using a high resolution monitor. The technical readings will be recorded on a Quality Control form. During quality control, technicians will identify missing/incomplete pages, pages out of sequence, and pages skewed, and will evaluate the image quality of text and illustrations. Any anomalies in image capture will be marked and compared against project worksheets and/or the originals.

For images consisting of text/line art, any or all of the following requirements must be exhibited when examining a 600 dpi paper printout without magnification:

- full reproduction of the page, with skew under 2% from the original (100% of all pages);
- sufficient contrast between text and background and uniform density across the image, consonant with original pages (100% of all pages);
- text legibility, including the smallest significant characters (100% of all pages);
- absence of darkened borders at page edges (98% of all pages);
- characters reproduced at the same size as the original, and individual line widths (thick, medium, and thin) rendered faithfully (98% of all pages);
- absence of wavy or distorted text (98% of all pages).

Magnification may be used to examine the edges and other defining characteristics of individual letters/illustrations. Under magnification the following text attributes are required for 98% of all pages:

- serifs and fine detail should be rendered faithfully;
- individual letters should be clear and distinct;
- adjacent letters should be separated;
- open regions of characters should not be filled in.

For illustrations and other graphics, the following attributes will be evaluated with or without magnification, as needed:

- capture of the range of tones contained in the original;
- consistent rendering of detail in the light and dark portions of the image;
- even gradations across the image;
- absence of moiré patterns and other distorting elements;
- the presence of significant fine detail contained in the original.

E. Performing Quality Control of Metadata and SGML Encoding

In addition to quality control of the content of the image files produced by the vendor, Cornell will verify the correctness of metadata and SGML encoding. Quality control will be based on sampling (except as identified). The sampling for inspection and evaluation will be in accordance with the ASQC Z1.9-1993, *Sampling Procedures and Tables for Inspection by Variables for Percent Nonconforming* (General Inspection Level II) and ASQC S2-1995, *Introduction to Attribute Sampling*[1].

Cornell will check for completeness and correctness of the directory structuring, file naming, and the correspondence of physical page numbers with image numbers. The contents of the checkmd5.fil, dataset.toc, and scandata.txt files will be sampled and inspected to ensure that data are entered as specified in the proposal. Cornell will also run Unix scripts on all of the checkmd5.fil files to verify the directory structuring and file naming (100% quality control).

As noted in section VI B5, Cornell will provide an electronic worksheet indicating the page ranges of selected document structures within volumes, and the vendor is expected to provide corresponding image numbers (preferably entering this data into the shared database). Cornell will verify the accuracy of vendor supplied image number information on worksheets for each volume.

The output of OCR will not be inspected by Cornell as the SGML encoded files' accuracy will be indicative of the text conversion process. SGML encoded files will be sampled and examined for errors to be sure that, on average, the vendor is meeting promised specifications for accuracy. If the sample reveals accuracy below 99.8% (character level), the entire shipment will be rejected. SGML accuracy will be based on character count, including tags, after encoding. The SGML encoded files will also be examined to ensure that the vendor follows the encoding protocols provided by Cornell, and is in conformance with the identified Document Type Definition. The SGML files will be interpreted (parsed) through an SGML parser to verify the correct use of tag and entity indicators. The auxiliary files that accompany the SGML files will also will sampled and examined for accuracy and consistency.

F. Identifying Unacceptable Images, Metadata, and SGML Encoding

Any images, metadata or SGML encoding considered unacceptable by Cornell due to image conversion errors, incorrect data entry, SGML accuracy levels below specifications, errors in SGML encoding/tagging will be returned to the vendor(s) with specific comments identifying the scope and nature of the problem. Cornell will also request replacement of printouts that do not meet the specified quality requirements. If the delivery media or any files on the media are nonfunctional, they will be also returned to the vendor for corrections. The vendor(s) will make all necessary adjustments and reproduce the products (images, metadata, SGML encoding, paper prints) at their expense to achieve an acceptable level of quality as identified in the proposal and confirmed in the sample production test.

VII. PROJECT RESPONSIBILITIES: VENDOR

The role of the vendor(s) is critical to the success of the project. The vendor(s) will be responsible for receiving up to 300,000 pages from Cornell and preparing, scanning, creating metadata, OCRing, SGML encoding, printing, inspecting, and delivering the products on CD-R. In addition, the vendor will be responsible for guaranteeing the quality of the images, metadata, text conversion, SGML encoding, and the printouts produced. The following sections describe vendor project responsibilities and provide instructions for vendor response.

[1] Order information for ANSI and ASQC standards mentioned in this RFP are available via Custom Standards Services' home page at http://www.cssinfo.com

A. Receiving Material from Cornell

The vendor(s) must acknowledge receipt of each item in a shipment, using an annotated copy of the packing slip. If any discrepancies to the packing slip are found, or the originals and/or accompanying metadata instructions are problematic, the vendor needs to inform the institution immediately. All materials will be stored in a secure, dry location at the vendor(s), and great care should be taken in handling the fragile originals.

B. Scanning Procedures

1. System calibration and performance

Vendor shall exercise rigorous quality control to maintain consistency of output as described in ANSI/AIIM MS44-1988 (R1993), *Recommended Practice for Quality Control of Image Scanners.* Vendor shall ensure that the scanning system is free of dust and other distorting particles, that it maintains calibration throughout each shift, and that the appropriate technical targets are used. These targets shall include, but are not necessarily limited to:

RIT Alphanumeric Test Object

The 3 X 3 (1.9, 1.3, .7 Density Target) (Catalog No. A200-610-D1.9) may be ordered from:
T&E Center
Rochester Institute of Technology
Lomb Memorial Drive
Rochester, NY 14623-5604
(716) 475-2739

(Note: MS44-1988 does not mention the RIT Alphanumeric Test Object.)

AIIM Scanner Test Chart #2

The target (Catalog No. X4410) may be ordered from:
AIIM Bookstore
1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910-5603
(301) 587-8202

2. Capability to manage textual data for multipage volumes

Bound volumes for scanning consist of both books (usually published in a single physical volume, on a specific topic, and as a closed-ended entity) and journals or serials (usually published on an open-ended basis over an extended period of time in many physical volumes). The physical volumes of serial publications may or may not correspond to the logical publishing unit of the serial. For instance, a publisher might issue one "volume" per calendar year, but that volume might be divided into several physical volumes for binding purposes. On the other hand, a sparse or infrequently published serial might be bound with several published "volumes" collected together in one physical volume.

3. Setting up the volumes

Prior to scanning, each volume should be reviewed to determine the presence and nature of illustrations, the page dimensions, and the physical condition. The inspection will ensure that the scanner settings selected (threshold, filters, screens, TRC maps, page trim) will provide the best possible image capture for a given volume. These settings should be recorded on the accompanying worksheet.

4. Scanning the volumes

a) Using a flatbed raster scanner, the vendor will scan all pages as bitonal (one bit) images (one page per image file) with no gray scale or color scanning at a resolution of true or interpolated 600 dpi. Because of the age of the material, it is anticipated that all pages must be manually placed on the platen and that an automatic document handler can not be used.

b) The RIT Alphanumeric Test Object, AIIM Scanner Test Chart #2, and a Production Note will be scanned at the beginning of each volume using the scanner settings chosen for that volume. The vendor shall use the version of the RIT target (1.9, 1.3, .7 density) that most closely corresponds to the density of the original volume. The vendor will verify that line 15 on the RIT target can be read in all four quadrants. If line 15 can not be read at the settings deemed optimum for the volume, the vendor shall note on the worksheet the smallest line pattern than can be read in all four quadrants. On the AIIM Scanner Test Chart #2, the Bodoni 4 point lower case letters should be clear and distinct, the diagonal line should be smooth and straight, and distinct halftone wedges representing the dynamic range present in the source document should be rendered free of moir★ patterns at the appropriate screen mesh (normally 133) in either regular or enhanced mode. Again, if these criteria can not be met at the settings deemed optimum for the volume, the vendor shall note on the worksheet this information.

c) Vendor staff will scan the accompanying technical targets and production note as well as all pages of each disbound volume (including blank pages), unless otherwise instructed by the presence of instruction slips and the information provided on the worksheet.

d) Each page shall completely fill the scan area. A page trim should be applied to the images of a volume so that page edges are not detected.

e) The pages will be aligned on the scanner platen in a manner to ensure little or no skew of the text from the original page (no more than 2%). Skew is measured from the two corners of the document image parallel to the longitudinal edge of the projected image frame.

f) Foldouts that are 11" x 17" or smaller will be scanned at their original size, front and back. It is important that the TIFF header information accurately reflect the size of foldouts so that proper printing instructions (with respect to the paper size used for printing) can be generated. For foldouts that are larger than 11"x17", the original will be reduced in size via photocopying and then scanned as above, with TIFF header information indicating the photocopy and original page dimensions. If significant information is lost in the reduction photocopy, the vendor will scan the foldout in sections at its original size. In any case, the original will be preserved and will be tipped into the printed digital replacement.

g) Centerfolds (i.e., a two page, uncut plate) will be treated as foldouts, both in the scanning and binding processes. Standardized pagination and naming for these pages will appear in the electronic worksheet.

h) On occasion, content which Cornell wishes to preserve appears on the endpapers of a volume. One such case is for bookplates, which are usually on the inside front cover, and which will generally be scanned as the verso of the production note. In other cases, important content appears across either the front or back endpapers. In such cases, the work will be disbound so as not to cut the endpaper, and the resulting double width "page" should be scanned as a foldout. All cases involving scanning of endpapers will be considered special circumstances warranting clear and obvious mention and handling instructions on the accompanying worksheet for the volume. Page numbering and file designations will treat each side of the sheet as a single image. The pages of a foldout are invariably not included in the logical pagination.

i) Cornell will provide a printed production note for each volume which will be formatted in such a way that when it is scanned it will appear centered left-to-right and the center will fall above the centerline of the page. The vendor will scan the sheet on both sides and insert the resulting images at the beginning of the relevant volume. These two images will constitute the initial leaf of the volume when it is printed and bound. Following the "canned" text of the production note, specific mention will be made if image or page sizes have been altered in the digitization process or if any pages in the logical pagination are not included in the digital copy.

j) The pages will be aligned on the scanner platen in a manner to preserve the front to back registration of the text on the recto and verso of a page leaf at the time of printing. No adjustments in duplex printing commands should be required.

k) The digital files and targets will be properly oriented, ordered, and named to reflect the presentation of the original volume. Images should be scanned in the orientation they appear in the bound volume. In other words, the orientation field should always indicate that the image was scanned "upright" (i.e., with Orientation: row 0 top, col 0 lhs) whether the image is portrait or landscape oriented.

5. Quality Control

a) Vendor staff will perform quality control to ensure that each page is fully rendered, properly aligned, and free of aliasing/distortions. Inspection and quality control data shall always be recorded on the worksheet accompanying each volume.

b) When necessary (e.g., poor image capture of an illustration), the staff will re-scan from the original text and insert the image(s) into the proper image file sequence.

PLEASE RESPOND WITH A DESCRIPTION OF YOUR EQUIPMENT, SOFTWARE, AND SERVICE CAPABILITIES FOR CREATING DIGITAL FILES TO THESE SPECIFICATIONS, INCLUDING YOUR QUALITY ASSURANCE PROCEDURES. DESCRIBE ENHANCEMENT CAPABILITIES FOR CAPTURING HALFTONE INFORMATION AS BITONAL IMAGES. INDICATE LOCATION(S) WHERE SCANNING WILL TAKE PLACE.

C. Printing Paper Copies from the Digital Images

The vendor will print paper copies of digital images to facilitate Cornell's quality inspection and to create hard copy replacements for the deteriorating originals following these specifications:

- Paper stock must meet ANSI Z39.48-1992 *Permanence of Paper for Publications and Documents in Libraries and Archives* requirements for permanence and durability.
- Image stability depends on proper adhesion of print to page; machine and toner requirements are defined in: Norvell M.M. Jones, *Archival Copies of Thermofax, Verifax, and Other Unstable Records*, National Archives Technical Information Paper No. 5 (Washington, D.C., 1986).
- Printing resolution should be 600 dpi or higher.
- There should be less than 1% variation in print size from the original.
- Duplexing should result in recreating the original front-to-back registration, within 2/10th of an inch.
- The printing process should not introduce any skew.
- Printing should be uniform, resulting in sharp contrast between text and background, with no banding or streaking.
- The image should be printed centered on the paper with front to back registration as close as possible to the original document. Trimming and binding will be done by the institution.

DESCRIBE YOUR ABILITY TO MEET CORNELL'S PRINTING NEEDS INCLUDING YOUR QUALITY ASSURANCE PROCEDURES. DESCRIBE PRINTING HARDWARE AND SOFTWARE THAT WILL BE USED FOR THIS PROJECT.

D. Metadata Creation

1. Directory Structuring and File Naming

The vendor is expected to return to Cornell sufficient information to allow determination of which image file (file name) corresponds to the physical page number for each page in each volume. Cornell will provide a unique name for each work to serve as the basis for naming the image files created from it. Cornell prefers that file names be devised such that the contents of a file can be reasonably determined from the file name, thus avoiding the need to use intermediate tables to decode file names. Cornell expects to load the generated TIFF files onto Unix file servers. If vendor wishes to use a file system other than Unix (e.g., MS-DOS), please describe the file naming conventions to be used and their compatibility with Unix file servers.

In general, the image files for books should be named with the root name provided by the library and an extension indicating the image sequence within the work. Image files from serials should be named with the root name provided by Cornell and a series of extensions indicating volume number, issue number and image sequence number. The vendor will follow the following file directory structuring and naming conventions:

Serials:

Root Identifier	Volume	Issue #	Page #s
century	v0000005	i0000001	0001000a.tif
			0002000b.tif
			0003000c.tif
			0004000d.tif
			0005r001.tif
			0006r002.tif
			....
			checkmd5.fil
		i0000002	00010053.tif
			00020054.tif
			00030055.tif
			....
			checkmd5.fil
			....
			dataset.toc
			scandata.txt

Monographs:

	(placeholder)	(placeholder)	
Root Identifier	Volume #	Issue #	Page #s
0001roeh	v0000000	i0000000	0001000a.tif
			0002000b.tif
			0003r001.tif
			0004r002.tif
			0005r003.tif
			....
			checkmd5.fil
			dataset.toc
			scandata.txt

Directory Structuring and File Naming Guidelines

1) All directory names will be eight characters and unique within the context of the project. Lowercase should be uniformly used in naming directories and files.

2) All journal names and monograph names will be eight characters and will be unique as assigned by Cornell. Eight characters will be used, with the first four being a zero padded incrementing number, and the last four being the first four letters of the author's last name. The worksheets will identify the unique name to be used to name the top level directory for a volume as the "Root Identifier." The library will also supply unique names using a scheme.

3) Any document types or individual files that have an innate sequence, or for which it may be desirable to process in an ordered manner, should incorporate an initial leading zero padded, incrementing number with a sufficient number of digits to encompass the collection, into the naming scheme. This will insure that the documents will sort as expected utilizing the natural sort order of all ASCII-based computer systems.

4) Under the "root identifier" level for each serial name there will be a directory level for volumes (labeled v#######) and one for issues (labeled i#######). For serials these directories will indicate the actual volume and issue number. For monographs, they will serve only as dummy directory levels so that the basic directory structure for monographs and serials are the same. Volume and issue numbers for monographs will always be "00000000" and "00000000" respectively. Serials occasionally have supplements which are not assigned issue numbers. However, supplements will be treated as issues in the directory structure, but will be labeled "s#######" instead of "i#######"

5) All file names will be in the ISO 9660 format. Image file names will have an extension of .tif. The file names themselves will be divided into two logical components. The first four characters will contain a leading zero padded sequentially incremented image sequence number, starting with 000. This project does not include any volumes with more than 999 pages. The final four characters will contain a representation of the page number as printed on the page, formulated according to the following rules:

   a) Every image should have designated a logical page number or appropriate tag to accompany it. This page number or tag will be used in the TIFF header and, amended if necessary to accommodate ISO 9660 file name restrictions, in the image file name.

b) For the purposes of these instructions, the word "pagination" will refer to the logical sequential pagination of a series of pages. For example, a page without a printed number on it which is located between pages imprinted 2 and 4 can be assumed to be page 3. Similarly, four pages without page numbers printed on them followed by pages 5, 6, 7, etc., can be assumed to be pages 1, 2, 3 and 4.

c) All pages which are included within the logical pagination should be designated with their actual page numbers.

d) The first two pages of every volume will be the production note and its verso, which may contain a book plate. These will always be designated 000A and 000B, respectively. If there are more pages before the logical pagination begins, they will be designated 000C, 000D, 000E, etc.

e) Pagination that appears as Roman numerals in the original will be translated into Arabic numbers and appended with a leading "R" for file names (e.g., page vii becomes page R007, etc.). In the absence of printed page numbers, it is to be assumed that Roman numerals continue until logical Arabic pagination commences. In the situation where sequential pagination continues through a change from Roman numerals to Arabic numerals, the Arabic numerals will be assumed to start at the change in type of document content. Typically this can occur in transitions from the index to editorial content. All pagination information will be delivered to the vendor on the digital worksheet.

f) When there are pages in the material which are not included in the sequential pagination (commonly occurring with plates) the pages will be designated by the number of the preceding paginated page appended with a trailing letter which will increase sequentially for each page (e.g., 0031, 0032, 032A, 032B, 0033, 0034).

g) Occasionally, there are pages that are included in the logical pagination which are not included in the material to be structured/scanned. These page numbers should also be deleted from the structure pagination.

h) When pages are numbered incorrectly in the original material, the correct logical pagination should be used in the TIFF header and file name, unless otherwise specified in the pagination instructions which will accompany each volume.

i) When pagination restarts result in duplicate page numbers in the same volume, the longest section will have its pages recorded unamended. Shorter segments will be recorded with a letter preceding the number to differentiate it from similarly numbered pages in the same volume (e.g., A001, A002, A003, A004; B001, B002, B003, B004). Note: Letters which will not be used in this situation are I, L, O, and R. This procedure does not need to be used for a Roman Numeral section if it is the only one in the volume. If there is more than one, the shorter section[s] will be differentiated in the same manner as Arabic Numbers (e.g., RA01, RA02, RA03, RA04; RB01, RB02, RB03, RB04)

j) Page numbers which actually contain letter prefixes will be recorded according to the same rules as standard Arabic numbered pages, except that punctuation between the prefix and the number should be dropped. Thus a page from Appendix A which is labeled A-9 should be recorded as 00A9.

k) Page numbers containing characters which are not permitted in ISO 9660 file names should be recorded with an underscore character in place of the illegal character. For example, page 22.6 should be recorded as 22_6. In situations such as this, the _unmodified_ page number should be recorded in the TIFF header (see the section on the TIFF header for more detailed instructions).

l) Adornments around page numbers, such as if the page number is both preceded and followed by a dash, asterisk, parenthesis, square bracket, etc., should be ignored (not entered).

m) Any pagination situation that falls outside those described above will be noted in pagination instructions which will accompany the volume. These instructions will include how the pages in question should be designated in the TIFF header and file name. If the vendor discovers a situation which is not described above and not commented on in the work sheet, they will contact Cornell for instructions about how to proceed.

6) The file checkmd5.fil will contain a listing of all the files in an image directory except checkmd5.fil itself. This file will be used during automated quality checking of files to a certain CD-R. Since file directories for serials are organized at the issue level, checksum files should encompass all the files in an issue for serials. Since the level of image organization for monographs is at the physical volume level, checksum files for monographs should encompass all the image files in a physical volume. The checksum algorithm used should be the 128-bit md5 (Message Digest 5) algorithm described in RFC 1321. (One source of RFC 1321 is http://ds.internic.net/rfc/rfc1321.txt). These will be compared with the checksum generated on the output media itself to help verify correct file writing.

```
e.g. 0001000a.tif      57edf4a22be3c955ac49da2e2107b67a
```

7) The file dataset.toc will contain a dump of all the bibliographic information recorded for the logical volume level for serials and at the physical volume level (i.e. the "title" level) for monographs.

8) The file scandata.txt will contain a dump of the TIFF header "Image Description" field. These files should be recorded at the logical volume level for serials and at the physical volume level (i.e. the "title" level) for monographs. The structure and contents of this file will be discussed later.

9) Cornell uses Unix operating system that uses only line-feed as the line-end character and does not employ a file-end character. Consistent use of line-end and file-end characters will be required of the vendor, since line oriented processing and counting procedures will not work correctly. Cornell will require the vendor to provide all data files with Unix formatting conventions, which will be given to the vendor.

Directory Structuring and File Naming Quality Control

1. Vendor staff will inspect directory structuring, file naming, and the contents of the auxiliary files for each volume to ensure consistency of input, page sequencing, and data integrity. Inspection and quality control data shall always be recorded on the worksheet accompanying each volume. When necessary, the staff will correct the structuring or naming.

PLEASE VERIFY YOUR ABILITY FOR STRUCTURING AND NAMING AS DESCRIBED AND PROPOSE A MECHANISM TO MANAGE THE STRUCTURING AND NAMING PROCESS INCLUDING QUALITY ASSURANCE PROCEDURES. DESCRIBE HARDWARE AND SOFTWARE TO BE USED DURING THE PROCESS.

2. Indexing

Following scanning, Cornell wishes to provide bibliographic access to the image files at a basic level. Cornell will extract bibliographic information for monographs and serials from the Cornell Library Online Catalog (NOTIS) and convert the MARC files into labeled ASCII files. This textual data (worksheet) can be presented to the vendor on diskette or via an FTP site. However, as mentioned earlier, the preference of Cornell is to maintain a shared database with the vendor in entering and receiving data to minimize data entry errors. For example, the vendor will be able to enter the shipment number, which will be assigned by the vendor, and image sequence numbers into this database. The delivery method and medium will be decided based on the vendor preferences and related experience.

The vendor will enter the bibliographic data provided by Cornell for each monograph and serial into dataset.toc files. The data will be ready for the vendor to transfer to dataset.toc files and there will be no need for reformatting or restructuring. The data need to be recorded as UNIX formatted ASCII text. The dataset.toc file will contain a dump of all the bibliographic information recorded for the logical volume level for serials and at the physical volume level (i.e. the "title" level) for monographs.

The bibliographic information will include:

Serial fields:

Notis ID (id)
ISSN (in)
LC Number (lc)
Document Identifier (S for serial, M for monograph) (di)
Journal Title (jt)
Other Journal Name(s) (ot)
Place of Publication (pl)
Publisher Name (pn)
Volume Number (vl)
Issue Number (no)
Date of Publication (volume and number level) (yr)
Actual Page Numbers of Journal Volume (ap)
LC Subject Headings (su)
Additional Information (such as special volume information) (ad)

Monograph fields:

Notis ID (id)
ISBN (ib)
LC Number (lc)
Monograph Title (mt)
Other Monograph Title (ot)
Primary Monograph Author (au)
Secondary Monograph Author(s) (sa)
City of Publisher (ci)
Publisher Name (pn)
Date of Publication (yr)
Monograph Series Note (se)
LC Subject Headings (su)
Additional Information (ad)

Indexing Quality Control

Vendor staff will inspect the indexing data entered in dataset.toc and record the quality control data on the worksheet accompanying each volume. When necessary, the staff will make corrections to ensure that bibliographic data presented by Cornell is input correctly.

PLEASE VERIFY YOUR ABILITY TO RECORD THIS INFORMATION IN DATASET.TOC FILES. PLEASE PROPOSE A MECHANISM TO MANAGE THE PROCESS DESCRIBED ABOVE, WHETHER SUCH DATA IS PRESENTED TO YOU ON DISKETTE, VIA FTP, OR THROUGH A SHARED DATABASE. CORNELL IS SPECIFICALLY INTERESTED IN YOUR IDEAS RELATED TO MAINTAINING A SHARED DATABASE. ALSO EXPLAIN YOUR QUALITY ASSURANCE PROCEDURE SUGGESTED FOR INDEXING AND DESCRIBE HARDWARE AND SOFTWARE TO BE USED DURING INDEXING.

3. Organization of the TIFF Header

TIFF Specifications

Generated TIFF files must not make assumptions about default values of TIFF header fields that could affect the ability of TIFF readers to properly display the file. Each TIFF file should explicitly state the values of the following header fields:

NewSubfileType
ImageWidth
ImageLength
RowsPerStrip
StripOffsets
StripByteCounts
X Resolution
X Position
Y Resolution
Y Position
Resolution Unit
BitsPerSample
Compression
PhotometricInterpretation
Orientation

Cornell also requires additional information recorded in the TIFF header:

Date and Time of Scan
Scanning Station and Operator Identifier (a code can be used to maintain privacy)
Copyright or Source Statement
Image Description

Some of the information necessary for TIFF headers will be provided by Cornell as part of the bibliographic data prepared for dataset.toc files. Other information, such as shipment number, will be assigned by the vendor. The data in the Image Description tag should be structured as follows:

For serials:

Shipment number [assigned by the vendor]
Serial name (full name) [derived from MARC jt field]
Year of publication [derived from MARC yr field]
Root identifier [top level directory name for this work]
Volume # [derived from MARC vl field]
Issue # [derived from MARC no field]
Image sequence number [first four characters of file name--assigned by the vendor]

Printed page number (or appropriate tag) [last four characters of file name]
(Exception: When the ISO 9660 naming convention requires that the record of the printed page number be changed before being made part of the file name for the image, the _actual_ printed page number--that is, the one containing the illegal ISO 9660 character--will be recorded here)
Document Structure label [assigned by the vendor from Cornell instructions]
(Note: There will usually be only one document structure label per page. However, on occasion, there may be more than one. Multiple document structure labels should be entered within the same pipe-delimited subfield and separated by an underscore--e.g. |TOC003_LOI001|. Such cases will be clearly flagged on the worksheet.

For monographs:

Shipment number [assigned by the vendor]
Monograph title [derived from MARC mt field]
Monograph author [derived from MARC au field]
Year of publication [derived from MARC yr field]
Root identifier [top level directory name for this work]
Volume # [fixed string 'V0000'] (for multi-volume monographs)
Issue # [fixed string 'I000']
Image sequence number [first four characters of file name--assigned by the vendor]
Printed page number (or appropriate tag) [last four characters of file name]
(Exception: When the ISO 9660 naming convention requires that the record of the printed page number be changed before being made part of the file name for the image, the _actual_ printed page number--that is, the one containing the illegal ISO 9660 character--will be recorded here)
Document Structure label [assigned by the vendor from Cornell instructions]
(Note: There will usually be only one document structure label per page. However, on occasion, there may be more than one. Multiple document structure labels should be entered within the same pipe-delimited subfield and separated by an underscore--e.g. |TOC003_LOI001|. Such cases will be clearly flagged on the worksheet.

Notes on the TIFF Header Contents

In general, elements in the Image Description tag should be delimited by an ASCII "pipe" or "vertical stroke" character. However, the sequence of elements starting with "root identifier" down to "image sequence number" should be separated by an underscore, as should multiple instances of document structure labels (see section IV). For example,

Image Description: "|1|The Century; a popular quarterly.|1895|CENTURY_V0000049_I0000003_0485_0480|UNSPEC|"

Image Description: "|1|Farm woodwork|Roehl, Louis Michael|1919|
0001ROEH_V0000000_I0000000_0013_0009|UNSPEC|"

Document Structure Labels

The following coding scheme should be used to label appropriate document structures:

Production Note	PNT
Title page	TPG
Table of contents	TOC
List of illustrations	LOI
List of tables	LOT
Bibliography	BIB
References	REF
Comprehensive Index	IND
Subject Index	SUI
Author or name index	PNI
Special index	SPI
Volume index	VOI
Mixed text/illus. index	MIX
Errata	ERR
Blank pages	BLP

In each case the three letter abbreviation should be followed by a zero padded three digit number representing the sequence within the particular document structure. Thus, the pages of a five page table of contents should be labeled TOC001, TOC002, TOC003, TOC004 and TOC005. If the page from Roehl's "Farm Woodwork" shown above was the third table of contents page, the TIFF header Image Description tag would look like this:

Image Description: "|1|Farm woodwork|Roehl, Louis Michael|1919|
0001ROEH_0009_0013|TOC003|"

Vendor(s) will not have to make any determinations about which pages should receive which document structure labels. This information will be included on the electronic worksheet accompanying each physical volume. If there is any potential ambiguity about which label to use, the worksheet will indicate the proper selection. For instance, LOI will be used for lists of illustrations, diagrams, maps, charts, figures, plates, etc. In some cases, more than one document structure may appear on the same page. If so, multiple labels should be recorded in the final subfield of the Image Description tag, separated by underscores (e.g. a page which has both the List of Illustrations and List of Tables might have |LOI001_LOT001| in the last position of the Image Description tag. Cases of overlapping document structure labels will be clearly flagged on the worksheet accompanying the volume.

For pages with no specific document structure label (the vast majority), insert the string UNSPEC.

Document structure labels should be properly associated with their respective physical page numbers and file names. Vendor is expected to record image number ranges corresponding to page number ranges for document structure labels on the provided worksheet for each volume. The TIFF header contents should be such that no errors are generated when the file is read by TIFF readers designed for the particular revision TIFF file being used. Vendor(s) will verifying the accuracy of data entry.

TIFF Header Quality Control

Vendor staff will inspect the TIFF headers and record the quality control data on the worksheet accompanying each volume. The TIFF headers need to be read through a TIFF viewer capable of viewing the headers to ensure that all the information is readable and there will be no error messages indicating inaccurate entry of data. When necessary, the staff will make corrections to ensure that TIFF headers are accurate.

PLEASE VERIFY YOUR ABILITY TO RECORD THIS INFORMATION IN THE TIFF HEADER, AND DESCRIBE YOUR QUALITY ASSURANCE PROCEDURES FOR ENTERING TIFF HEADERS. DESCRIBE HARDWARE AND SOFTWARE TO BE USED DURING THE PROCESS.

E. Text Conversion and OCR Processing

All the images created during this project will be converted to text (ASCII format) through OCR processing after passing the quality inspection of the vendor. The ultimate goal is to encode the textual information for indexing and other text manipulation purposes using SGML standard. As a result, users will be able to search the scanned monographs and serials by keyword, title, author, title word, issue number, etc. through the interface that will be provided by the library.

Cornell will provide the page count and average number of characters per page. The output of the text conversion needs to be recorded in ASCII format. If a file compression method will be proposed by the vendor, self-extracting mechanisms are preferred.

The desired OCR accuracy level is 99% (maximum 1 errors in every 100 words: this translates to 99.8% accuracy at the character level if each word averages five letters). The text conversion should be conducted on 600 dpi, unprocessed image files. Cornell recognizes that the percentage of each page that is OCRed correctly varies greatly depending on the quality of the image and the typeface. The pages that contain multiple type face and size, such as title pages and advertisements, may have a lower accuracy level (a threshold accuracy level will be identified by Cornell and the vendor). However, there is enough consistency among the materials that will be scanned (such as page headers and footers) to enable the vendor to identify a method to capture consistent patterns for markup and text manipulation.

The vendor’s OCR system should be able to:

- Block sections of the object so that images surrounded by text can be omitted, columns may be defined and arranged in a certain order

- Attain corrected OCR accuracy rate of 99% (will require manual corrections)

The purpose behind the text conversion is not to recreate the appearance of the original (with original page numbers, running headers, etc.) but merely to provide searchable text for accessing images. When graphic information that can not be captured by OCR is present, a note needs to be embedded in the ASCII document indicating the existence of the image files. Whenever the image file includes text that can be converted without extensive manipulation, e.g., a map with place names, this textual information needs to be converted and recorded in ASCII format. The vendor will adhere to the text conversion guidelines provided by RLG in determining how to handle non-ASCII characters and graphical elements, and linking illustrations to the text during the OCR process (http://lyra.rlg.org/scarlet/tech.html).

The text conversion process will provide ASCII versions of the image files for SGML encoding. The SGML encoding will be done following the Text Encoding Initiative (TEI) Lite Document Type Definition (DTD), Version 1.6, for texts (ftp://www-tei.uic.edu/pub/tei/lite/). The vendor will insert the indexing provided by Cornell (as identified in Section VII.D) into a document header at logical volume level for serials and at the physical volume level (i.e. the "title" level) for monographs.

All the document pages of a volume need to be concatenated into a single SGML file that includes markup that divides the content into broad categories (introduction, body, conclusions, etc.) as specified by the TEI Lite DTD. The vendor will also insert page numbers on each page and retain references to non-text images. In addition to adhering to the TEI Lite DTD guidelines, the vendor will follow the encoding guidelines provided by RLG (http://lyra.rlg.org/scarlet/tech.html).

Cornell will provide unique file names for all the encoded serials and monographs in the accompanying worksheets. The vendor will use the identified file names and deliver the encoded files on a CD-R (image files and text files will be stored separately). Several types of associated files will be provided for each encoded serial and monograph. There will be *one* for each full-text document and will be named as identified in the worksheet.

Page Information Group Files will contain a list of all the page information group tags and their contents, in the order in which they appear in the document.

Reference Files will include a list of pointers to internal and external references (e.g., all occurrences of tables, illustrations) in the encoded documents.

Omission Report Files will record all the anomalies encountered by the vendor that falls outside of the provided guidelines (e.g., any omitted text due to illegibility)

Entity Files will list all the entity values used in the document and the filenames associated with the entity value.

The vendor response needs to provide the following information for both OCR and SGML encoding software:

Text Conversion and SGML Encoding Quality Control

Vendor staff will inspect the text conversion and encoding to check that the vendor is meeting agreed upon specifications for accuracy. The vendor needs to parse the encoded files using an SGML parsing software to ensure that it confirms with the TEI Lite DTD. In addition, all the auxiliary files accompanying the encoded files need to recorded and named accurately. This quality control data need to be recorded on the worksheet accompanying each volume. When necessary, the staff will make corrections to ensure that encoding meets the identified quality standards.

PLEASE DESCRIBE YOUR OCR AND SGML ENCODING CAPABILITIES, INCLUDING SOFTWARE USED AND YOUR MEANS FOR GUARANTEEING ACCURACY LEVELS. EXPLAIN HOW YOU WOULD SUPPLY OCR/SGML ENCODING OUTPUT FILES. ALSO INDICATE LOCATION(S) WHERE SCANNING WILL TAKE PLACE.

F. Output/Delivery Media

Cornell would like to receive all files from the vendor on CD-R media. The CDs should comply with the ISO-9660 standard. Material scanned for serials and monographs should not be mixed on any output CD. Also, pages from individual physical volumes should all be together on the same CD. The outputs of the SGML encoding will be delivered on a separate CD-R, clearly marked to indicate its contents. All rescans, corrected metadata, and encoding will be integrated in one CD-R.

G. Shipping Digital Files

All deliveries of products (images, metadata, SGML encoded text, printouts) will be made within one month of receiving the physical volumes from Cornell. Vendor(s) will be responsible for all costs associated with shipping and handling of the products. Vendor(s) will pack the delivery medium containing the products, prepare a packing slip, insure each shipment, and ship via overnight courier the material to Cornell to:

Anne R. Kenney, Associate Director
Department of Preservation and Conservation
214 Olin Library, Cornell University
Ithaca, NY 14853

At their sole discretion, Cornell shall have the option to specify, on a per shipment basis, whether original materials are to be retained by the vendor until Cornell has fully inspected and accepted the products. The original documents do not need to be shipped via overnight courier. They will be sent via two-day delivery to decrease shipping costs.

In the case that the original material is to be returned along with the products, the volumes will be sent in the same delivery shipment as the products, and shall be accompanied by an annotated copy of the original packing list. Cornell will inspect the contents of each shipment to be sure that all items are present, complete, and in original page order.

In the case that the original material is to be retained by the vendor until Cornell has fully inspected and accepted the products, all original material in a shipment will be held at vendor's location, until such time as Cornell has inspected and accepted the products. At such time, all original items included in a shipment will be returned to Cornell together in a single delivery shipment which contains an annotated copy of the original packing list. Cornell will inspect the contents of each shipment to be sure that all items are present, complete, and in original page order.

The vendor(s) shall be responsible for all costs associated with shipping and handling of the return of the source materials and digital files to Cornell.

H. Subcontracting

Subcontracting of scanning, metadata creation, OCRing, SGML encoding, printout creation, recording, and inspecting are permitted under this contract. In the event of a subcontracting agreement, the vendor(s) shall assume responsibility for work performed by the subcontractor.

I. Guarantee of Digital Image, Metadata, SGML Encoding, and Printout Quality

The vendor(s) must be willing to guarantee the quality of digital images, metadata, SGML encoding, and printouts. The vendor must also agree to correct these products that are redeemed unacceptable by Cornell due to vendor errors, at no additional charge to Cornell, and adhering to the specified project timeline.

PLEASE STATE YOUR WILLINGNESS TO GUARANTEE THE QUALITY OF THE PRODUCTS THAT DO NOT MEET THE CRITERIA IDENTIFIED IN THE RFP (DUE TO VENDOR ERROR). ALSO , EXPRESS YOUR WILLINGNESS TO MAKE THE NECESSARY CORRECTIONS AT NO ADDITIONAL COST TO CORNELL.

VIII. PROJECT MANAGEMENT AND FORMAL CONDITIONS

The success of this project will depend on how well the participating institution and the chosen vendor(s) manage their respective project responsibilities. In addition to each party complying with its appropriate technical specifications and production timetables, another factor--the establishment of key expectations between Cornell and the chosen vendor(s)--is also critical to the project's success. Cornell has identified the following key expectations:

A. Communication

Cornell University and the vendor(s) shall each designate representatives who will be available upon request to field questions and to discuss any aspect of the project. The Cornell University project representative will be Anne R. Kenney. As identified in the project timetable, there will be periodic conference calls between the vendor and Cornell to facilitate communication and to provide a platform for exchanging ideas to resolve problems that might be encountered during the project.

B. Delivery schedule

Cornell and the vendor(s) shall establish a regular delivery schedule of physical volumes from Cornell to the vendor(s) and products from the vendor(s) to Cornell. Any changes in the delivery schedule shall be communicated by the party initiating the change within 2 (two) weeks prior to the schedule change. Cornell must be able to change or cancel order releases against this agreement or blanket order any time before shipment without penalty.

C. On-Site Inspection of Vendor Facility

At any time during the agreement period, the vendor(s) shall permit representatives from Cornell to inspect its facility during the course of the vendor(s)' normal working hours.

D. Default

Cornell shall notify the vendor in writing concerning potential conditions of default--e.g., unsatisfactory service or poor workmanship. Failure of the vendor to correct the conditions identified by Cornell at its own expense, or to come to an amicable solution with Cornell within thirty days, shall constitute default.

E. Cancellation for Non-Compliance

Cornell University or the vendor(s) shall have the option to cancel the agreement upon thirty days written notice to the other party for performance that is not in compliance with all instructions and specifications stated within this document.

F. Errors

The vendor shall correct any errors identified during the institutional inspection process at no additional charge to Cornell. The vendor(s) shall reproduce the unacceptable products within 14 (fourteen) working days of the vendor having received the item(s) for correction. Any extra transportation or mailing costs resulting from such errors shall be paid by the vendor(s).

G. Invoices

1. The vendor(s) shall provide detailed invoices for each completed shipment within 7 (seven) working days of delivery of a shipment to Cornell. Invoices shall reflect the price structure delineated in the agreement. The invoices must include:
   - the total number of pages scanned, and the charge per image,
   - the number of images named, the number of directories structured, the number of images for which TIFF headers are generated, the number of records indexed (the number of auxiliary files), and the itemized price per image/record,
   - the total number of pages OCRed and SGML encoded, and the charge per 1,000 characters for text conversion and encoding
   - the total number of printouts printed, and the charge per page printed
   - the total charge for the shipment and insurance

In addition, it should also include the Cornell Purchase Order Number and any other itemized charges. Invoices will be paid upon the acceptance by Cornell of the products following inspection as described elsewhere in this RFP.

2. Invoices shall be sent to:

   Cornell University
   Invoice Processing
   P.O. Box 4040
   Ithaca, New York 14852-4040

PLEASE INDICATE IN WRITING YOUR AGREEMENT WITH THESE KEY PRINCIPLES.

IX. RIGHTS OVER PRODUCTS

Cornell University Library will retain all physical and property rights over the original volumes, digital images, and text and SGML encoded files that are products of the project described in this proposal (including the Preliminary Production Test). The vendor(s) will be performing work for hire and therefore will be expected to sign a legal waiver regarding any use of the products delivered to Cornell.

SIGN THE AGREEMENT PRESENTED IN ATTACHMENT B AND RETURN IT WITH YOUR RESPONSE.

X. PRICING

A. Prices quoted by the vendor(s) shall be firm for the duration of the project.

B. Prices quoted by the vendor(s) shall be net, unless otherwise specified by Cornell.

PLEASE SUPPLY A SCHEDULE OF PRICES FOR SERVICES REQUIRED IN THIS RFP USING ATTACHMENT C. PRICES MUST BE PROJECTED FOR THE ENTIRE PERIOD OF THE PROJECT (NOVEMBER 1998 THROUGH JULY 1999). YOU MUST INCLUDE:

A projected base bid per page. The scanning cost should include all the services specified in the RFP, with the exception of OCRing, SGML encoding, and printing hard copies of the images (printout). Pricing should include scanning, directory structuring, file naming, TIFF header information input, and the creation of dataset.toc, checkmd5.fil, and scandata.txt files. In addition, the price should also include set up, quality control, media, shipping, insurance, and other additional costs. Assume that volumes will be delivered disbound with the inner margin trimmed perpendicular to the text. Cornell may consider scanning some volumes where the binding may not be removed. If you can handle such material, please provide a separate bid/page for bound volumes.

Text conversion and SGML encoding/tagging of text converted files. Provide text conversion (OCR) and SGML encoding pricing per 1,000 characters for implementing TEI Lite DTD at 99% accuracy level. Price should include return of encoded text, auxiliary files, inspection, correction, supplies, and special handling. Indicate the pricing assuming that up to half of the pages to be OCRed may be in two column format requiring manual zoning. If the vendor prefers, OCR and SGML encoding prices can be provided separately.

Printouts of the digital images. Provide pricing per page to create 600 dpi paper prints of digital images on acid free paper meeting ANSI Z39.48-1992 standard.

XI. STATEMENT OF CAPABILITY, EXPERIENCE, AND COMPANY VIABILITY

The vendor(s) shall have the experience and capability to undertake this project.

Business Summary
Provide a brief business summary (one page or less) on your company. The summary should included whether you are a public or privately held, how many years you have been in the business, what your annual sales are, how many full and part time employees you have, examples of any business you have done with Cornell, and your D&B number.

References
Please provide three (3) customer references including contact name, phone number, and a brief description of your business relationship with them. These references should be for products and services similar to what is requested in this RFP.

Technical Support

- Do you provide a local or 800 number for no charge technical support?
- What is the phone number for technical support?
- Where are technical support personnel located?
- How many support staff are there?
- What are their hours of availability?
- What services are available?

PLEASE ANSWER THE VENDOR BUSINESS PROFILE QUESTIONS STATED ABOVE AND PROVIDE ANY OTHER RELEVANT INFORMATION ABOUT YOUR BUSINESS, HARDWARE/SOFTWARE USED, AND WARRANTY AND SERVICE PROFILE.

ATTACHMENT A: FAX RESPONSE FORM

PLEASE INDICATE BY **MONDAY, AUGUST 10 BY 5PM** WHETHER YOU UNDERSTAND THE REQUIREMENTS OF THE RFP AND YOU WILL BE RESPONDING TO IT, AND YOU WISH TO RECEIVE THE SAMPLE VOLUMES FOR THE PRELIMINARY PRODUCTION TEST. USE THIS ATTACHMENT FOR FAXING YOUR RESPONSE.

Vendor Information:

Address (for mailing the Preliminary Production Test):

Phone Number:
Fax Number:
Email Address:

____ YES, I UNDERSTAND THE REQUIREMENTS OF CORNELL'S RFP AND WILL RESPOND TO IT

____ NO, I WILL NOT RESPOND TO CORNELL'S RFP

______________________________ ______________________________
Signature Date

____
Printed name

PLEASE FAX THIS FORM TO:

Cornell University Purchasing Department
Attn: Glenn Morey, Purchasing Agent
FAX: 607-255-9450

ATTACHMENT B: RIGHTS OVER PRODUCTS

ASSIGNMENT OF RIGHTS

I understand that any copyrightable work ("Work") which I develop in the course of my employment by Cornell University ("Cornell") constitutes "work for hire" as defined in 17 U.S.C. Section 201(b) of the federal Copyright Act and all ownership rights to such Work belong to Cornell as my employer.

Should such Work not constitute a "work for hire" under copyright law, I hereby grant, transfer, assign and convey to Cornell and its successors and assigns, the entire right, title and interest in my Work or any part thereof, including but not limited to the right to reproduce, to prepare derivative works, to distribute by sale, rental, lease, lending or other transfer; to perform publicly, and to display the Work, as well as the right to secure copyrights or patents and renewals, reissues, and extensions of any such copyrights or patents in the United States of America or any foreign country;

Whether a copyright in the Work will be maintained or registered in the United States of America or any foreign country shall be at the sole discretion of Cornell;

I agree to cooperate fully with Cornell in the preparation and execution of all documents necessary or incidental to this assignment and the protection and preservation of rights herein granted to Cornell.

______________________________ ______________________________
Signature Date

____
Printed name

ATTACHMENT C: PRICING FOR THE SERVICES AND PRODUCTS

Description	Quantity	Unit Price	Total Amount	Notes*
Scanning				
Unbound Volumes (≤8.5 x 14”)	900 volumes 270,000 pages			
Unbound Volumes (>8.5 x 14”)	50 volumes 15,000 pages			
Bound Volumes (≤8.5 x 14”)	30 volumes 9,000 pages			
Bound Volumes (>8.5 x 14”)	20 volumes 6,000 pages			
Text Conversion and SGML Encoding (cost/1,000 characters)				
at 99% accuracy	2,500 char/page 750 million char			
Paper Printouts				
≤8.5 x 11”	925 volumes 277,500 pages			
>8.5 x 11”	75 volumes 22,500 pages			
GRAND TOTAL	1,000 volumes 300,000 pages			

*** If not meeting the specifications as described in the proposal, list exceptions:**