

1. REQUEST NO. <b>RFQ# 08QFD001</b>	2. DATE ISSUED <b>12/12/07</b>	3. REQUISITION/PURCHASE REQUEST NO. <b>FEDLINK ID: DLCV</b>	4. CERT. FOR NAT. DEF. UNDER BDSA REG. 2 AND/OR DMS REG. 1	RATING
--	-----------------------------------	--	--	--------

5A. ISSUED BY <b>The Library of Congress, 101 Independence Avenue, S. E., RM 318, Washington, DC 20540-9414</b>	6. DELIVER BY (Date)
--	----------------------

5B. FOR INFORMATION CALL: <i>(Name and telephone no.) (No collect calls)</i> <b>Joan Fitts, 202-707-0407</b>	7. DELIVERY <b>FOB DESTINATION      OTHER (See Schedule)</b>
---	---

8. TO: NAME AND ADDRESS, INCLUDING ZIP CODE <b>Electronically Transmitted to Internet Archive</b>	9. DESTINATION <i>(Consignee and address, including ZIP Code)</i> <b>Library of Congress 101 Independence Ave. SE Washington, DC 20540</b>
--	---

10. PLEASE FURNISH QUOTATIONS TO THE ISSUING OFFICE IN BLOCK 5A ON OR BEFORE CLOSE OF BUSINESS (Date) <b>12/19/2007</b> <b>Quotations will be accepted by e-mail or facsimile only</b>	11. BUSINESS CLASSIFICATION <i>(Check appropriate boxes)</i> <table style="width:100%; border:none;"> <tr> <td style="width:50%;">a. STANDARD INDUSTRIAL CLASSIFICATION CODE _____</td> <td style="width:50%;">b. SMALL BUSINESS SIZE STANDARD _____</td> </tr> <tr> <td>c. SMALL      OTHER THAN SMALL</td> <td>DISADVANTAGED      WOMEN-OWNED</td> </tr> </table>	a. STANDARD INDUSTRIAL CLASSIFICATION CODE _____	b. SMALL BUSINESS SIZE STANDARD _____	c. SMALL      OTHER THAN SMALL	DISADVANTAGED      WOMEN-OWNED
a. STANDARD INDUSTRIAL CLASSIFICATION CODE _____	b. SMALL BUSINESS SIZE STANDARD _____				
c. SMALL      OTHER THAN SMALL	DISADVANTAGED      WOMEN-OWNED				

**IMPORTANT:** This is a request for information, and quotations furnished are not offers. If you are unable to quote, please so indicate on this form and return it to the address in Block 5A. This request does not commit the Government to pay any costs incurred in the preparation of the submission of this quotation or to contract for supplies or services. Supplies are of domestic origin unless otherwise indicated by quoter. Any representations and/or certifications attached to this Request for Quotations must be completed by the quoter.

**12. SCHEDULE (Include applicable Federal, State and local taxes)**

ITEM NO. (a)	SUPPLIES/SERVICES (b)	QUANTITY (c)	UNIT (d)	UNIT PRICE (e)	AMOUNT (f)
	<p><b>LC/FEDLINK DIGITIZATION OF PUBLIC DOMAIN MATERIALS -</b> Vendor shall perform work in accordance with the Contract # LC07D7702 and the attached Statement of Work (Attachment A)</p> <p>1. Cost for providing public access to documentation of best practices for safely and efficiently converting high volumes of analog library materials that are typically difficult to scan such as brittle books, books with foldouts, oversized items, and other materials that require special handling.</p> <p>2. Cost for conducting pilot project to test and document the integration of a limited collection of special analog library materials.</p> <p>3. Costs for production of digital images and associated metadata and files shall be based on a fixed fee per page and may reflect discounts for large volume orders.</p> <p><b>TOTAL PROJECT COST</b></p>				<b>\$\$\$450,000 for work done in the period of 4.5 months starting from signing.</b>

DISCOUNT FOR PROMPT PAYMENT			
-----------------------------	--	--	--

**NOTE:** Additional provisions and representations are attached.

14. NAME AND ADDRESS OF QUOTER <i>(Street, city, county, State, and ZIP Code)</i> <b>Internet Archive</b>	15. SIGNATURE OF PERSON AUTHORIZED TO SIGN QUOTATION 	16. DATE OF QUOTATION <b>12/18/2007</b>
--	--	--

116 Sheridan Avenue San Francisco, CA 94129	17. NAME AND TITLE OF SIGNER <i>(Type or print)</i> <b>Brewster Kahle, Digital Librarian</b>	18. TELEPHONE NO. <i>(Include area code)</i> <b>415-561-6767</b>
--	---	---

**STATEMENT OF WORK FOR LIBRARY OF CONGRESS SLOAN GRANT PROJECT  
FEDLINK ID: DCLV**

## 1. BACKGROUND INFORMATION

The Library of Congress (the Library or LC) was awarded a grant by the Alfred P. Sloan Foundation to digitize thousands of public domain books from the Library's collections and to make the digital files freely available to the public. The grant also tasked the Library to develop, test, and share best practices for cost-effective, non-destructive scanning of large volumes of bound materials, with emphasis on handling brittle books, books with foldouts or other special format characteristics, and other library materials that are typically difficult to scan. In addition, the Library is tasked to work with the Internet Archive to develop an open source page-turner, and other applications to support digitization, creation of metadata, and public access.

**2. SCOPE OF WORK:** This task order is for the purchase of services and digital files of scanned materials under Lots 1 and 2 of the Contract #LC07D7702, LC/FEDLINK DIGITIZATION OF PUBLIC DOMAIN MATERIALS ("Contract").

### 2.1. LOT 1 – Development, Testing and Public Release of Best Practices and Support Applications.

**(A) Best Practices.** The Library is working on developing best practices – effective processes, standards, and technologies – for safely and efficiently converting high volumes of analog library materials that are typically difficult to scan such as brittle books, books with foldouts, oversized items, and other materials that require special handling for scanning. Best practices documents will include: work breakdown structures; materials handling instructions; standards for images, metadata, and quality assurance; and specific strategies for risk management and problem resolution. Internet Archive's role in this development is to make the best practices documents accessible on a public website (e.g., the Internet Archive website or a website associated with the Open Content Alliance).

**(B) Pilot Project.** A pilot project to test and document the integration of a limited collection of special analog library materials, defined in the above project, with the scanning of standard materials. Internet Archive will develop workflows incorporating brittle books and other materials requiring special handling in a high volume, high throughput scanning environment. In order to advance the understanding of how to incorporate these special materials into the normal scanning center operation, the Library shall hold the Internet Archive harmless for items designated as brittle which are damaged during scanning, provided the Internet Archive adheres to the strict handling procedures prescribed by the Library.

### 2.2. LOT 2 – Digitization and Free Public Access.

**(A) Digitization of Public Domain Materials.** Pursuant to section B.4. of the Contract, the Library agrees to purchase the services and deliverables indicated under Lot 2 of the Contract, namely to convert standard analog library materials, create required metadata, and make the resulting digital files available to the Library for download and provide access to the public via the Internet.

**(B) Material to be Digitally Reformatted.** The Library will select approximately 10,000 – 15,000 bound books in the public domain taken from the following collections: History of America and United States (LC Classification E), U.S. Genealogy and Regimental Histories (LC Classification CS), and Photography (LC Classification TR). It is expected that approximately ten (10) percent of the items will be deemed to be brittle.

**(C) Scanning.** The Internet Archive shall use non-destructive methods and equipment for scanning items unless other procedures are specifically agreed upon with the Library in writing, as specified in Sections B.4.1.3., and B.4.1.4. of the Contract.

**(D) Requirements for Digital Images.** Items shall be digitized according to the specifications in section B.4.1.4.b of the Contract. Under this task order, the Library and the Internet Archive agree to the following additional Output Formats:

- (1) The Internet Archive will deliver OCR in one XML format: ABBYY.
- (2) Agree to pursue XML for descriptive metadata in MODS and METS. See <http://www.loc.gov/standards/mods/> and <http://www.loc.gov/standards/mets/> for further information about these standards

- (3) XML for administrative metadata collected during scanning.
- (4) XML structural metadata. The Library may acquire structural metadata to support efficient navigation of the digital material. Internet Archive will consent to modifying 1 Scribe Scan Station to accommodate this proof-of-concept pilot for Library materials, not to exceed 10% of total output. Minimally, the metadata shall include: title page, table of contents, chapter or section (one level), and back of book index. The metadata shall be represented in XML. The Library may agree to shift the metadata activity to a post-scanning, OCR-based process. The Library retains the option of adding metadata to a captured item record or expanding metadata within an item record, if needed. The Internet Archive shall collaborate with the Library to develop a suitable metadata structure.

**(E) Library-Internet Archive Agreement Concerning Metadata.** Details of the structure and elements of metadata associated with the scanned materials shall be agreed upon between the Internet Archive and the Library and shall be specified in the task order project plan.

**(F) Additional Requirements or Standards.** Based on the application of best practices for particular types of materials and use, the Library and the Internet Archive may agree, in writing, upon additional requirements and standards for digital images.

### F.1. Bound Materials

**F.1.1. Book Covers.** Covers shall be scanned for certain books. When covers are required to be scanned, an instruction will be provided in a note included on the target. If both front and back covers are to be scanned, the front cover image shall be numbered to precede the images for the inside pages and the back cover shall be numbered to follow them.

The general rules for cover scanning are as follows:

- Do scan covers (front and/or back) when the cover includes typography or illustrations that are original to the book, and when a legible image can be produced.
- Do not scan covers of books that will not render a legible image.
- Do not scan covers that are devoid of any marking except the Library's call number.

**F.1.2. Inside Pages.** The images of the inside page shall come after the images for the target and the front cover (if any). The first page of the book to be scanned shall be the first page containing significant information. Examples include a page containing a copyright stamp that precedes the title page, the title page itself, or end papers containing significant information, such as a map. Scanning of the remainder of the book shall continue in sequence, omitting blank pages. However, pages that contain no printed information but that contain handwritten inscriptions, notes, marginalia or other written ephemera shall be scanned. End papers shall only be scanned if they contain significant information, such as a map. End papers that are merely decorative shall not be scanned. Blank pages or blank pages with stray pen or pencil marks shall not be scanned.

### F.2. Cropping

**F.2.1 Cropping of Bound Manuscript materials.** The Library requires presentation of the entire original sheet or page. In no event shall the actual document be cropped. Researchers using Library of Congress digital documents often wish to be reassured that the entire document has been captured. This is especially desirable for unbound manuscript documents. A border zone approximately ¼ inch or less of the surface behind the scanned document shall be provided whenever possible.

For some combinations of document sizes and scanning equipment, capturing such a margin may not be possible for all four edges of the page. Therefore, the Library desires a 3-inch margin wherever possible, and requires at least that the entire sheet or page is captured.

**(G) Quality Assurance.**

**G.1 Quality Assurance Plan.** The Internet Archive shall provide a quality assurance plan as part of the project plan for this task order. The quality assurance plan shall include the Library’s standards for acceptance of digital images and associated files. The Library will have open access to review the quality assurance plans and the execution of quality assurance measures throughout the project.

**G.2 Review by the Internet Archive.**

The Internet Archive shall perform manual quality assurance review in the range of 12% – 20% of the imaging output by qualified staff, depending upon the number of books in a given set.

<b>Books in set</b>	9-15	16-25	26-50	51-90	91-150	151-280
<b>number to QA</b>	3	5	8	13	20	32

See attached QA document for further details.

(2) The Internet Archive shall perform automated file validation on 100% of the image input files. For specific review process methodology, please see attached QA document.

(3) Review shall include identifying the following problems:

- Cropped text
- Blurred page(s)
- Missing page(s)
- Front cover missing
- Back cover missing
- Book was scanned twice; this copy darkened.
- Washed-out text (bad light-comp)
- Evidence of scanner (fingers/shadows/etc) visible on page
- Glass not centered in gutter; text is distorted or cropped
- Book and metadata do not match
- Tissue pages marked incorrectly
- Anomaly in image format is under investigation
- Left/right pages are reversed

For further detail on Internet Archive Error Codes and Classes, please see attached QA document.

**G.3. Library Review.**

- (1) The Library will perform either 100% quality assurance review or periodic, statistical, quality assurance review of the Internet Archive’s imaging files to identify imaging problems such as blur, missed or duplicate pages, obscured page content, skew, crop, exposure, artifacts, or other image-related quality problems.
- (2) If 100% quality review is undertaken by the Library, a 1% error rate is considered acceptable. If a statistical quality assurance review is performed, all images reviewed must be deemed acceptable. The Government Quality Assurance Surveillance Plan (QASP) will be negotiated with the Internet Archive at the time of contract award and will incorporate the Internet Archive’s proposed Quality Assurance Plan. An information copy of the Government QASP is provided in Section D as Attachment D.1.
- (3) After such reviews, the Library will provide the Internet Archive with a detailed report that includes: severity, error type, error position bias (top, bottom, left, or right).

**G.4. Correction by Internet Archive.** The Library shall notify the Internet Archive in writing of any digital images or files deemed unacceptable. The Internet Archive shall re-scan or re-process any digital image or file the Library deems unacceptable within 45 days of download from Internet Archive’s website, in accordance with and subject to the provisions of Article B.4.1.5.b of the Contract.

**(H) Reporting.**

**H.1. Progress Reports.** The Internet Archive shall provide timely, periodic Progress Reports, which shall include production figures, in a format and on a schedule as agreed to by the Internet Archive and the Library.

**H.2. Daily Testing and Monthly Report.** The Library shall provide the Internet Archive with an image target. Before each daily shift, the Internet Archive shall validate the imaging capabilities of each scan station by digitizing the image target and reviewing the resulting file. The Internet Archive shall inform the Library if any scan station is producing image characteristics below expected values. The Internet Archive shall provide the Library with a monthly detailed report on the digital imaging performance of each scan station.

**H.3. Schedule for Digitization, Receiving Reports, Return Reports.** As specified above, the Internet Archive shall provide the Library with Receiving Reports, Return Reports, and a Schedule for Digitization.

**H.4. Quarterly Summary of Scanning Activity.** Quarterly reports shall be submitted to the Library as specified in Section B.5.4. of the Contract.

**(I) Project Plan.** The Internet Archive and the Library shall document a project plan for digitizing the Library’s materials. At a minimum, the plan shall address: the type of materials to be scanned, required image formats and metadata elements, the schedule for digitization, and the plan for quality assurance.

**(J) Phase two of the scanning center operation -** Pending the availability of funds, requirements may be the same as phase one.

**2.3. PROJECT MANAGEMENT AND COORDINATION WITH THE LIBRARY.**

**(A) Key Personnel - Contractor.** Pursuant to section 5.4. of the Contract, the Internet Archives agrees to provide the services of the following individuals:

(1) Project Manager.

Robert Miller, [Robert@archive.org](mailto:Robert@archive.org), +1-415-561-6767

**(B) Key Personnel - Library.** Library personnel, who will coordinate with the Internet Archive personnel, include the following individuals:

(1) Project Manager: Mike Handy, [mhan@loc.gov](mailto:mhan@loc.gov) 202.707.8338

(2) Project Officer: Aaron Chaletzky, [aach@loc.gov](mailto:aach@loc.gov) 202.707.6026

(3) Technical Monitor: Anne Harrison, [anha@loc.gov](mailto:anha@loc.gov) 202.707.4834

**2.4. SUMMARY OF DELIVERABLES.**

**(A) Lot 1 – Development, Testing, and Public Release of Best Practices and Pilot Project.**

(1) Distribution of Best Practices Documentation.

– Free public access

(2) Pilot Project Documentation.

– Receiving Report

– Return Report

– Progress Report

**(B) Lot 2 – Digitization and Free Public Access.**

- (1) Digitization and Access
  - Digital images and associated files
  - Structural, administrative, and descriptive metadata
  - Free public access
  - Library access
  - Preservation copy (as needed)
- (2) Project Management
  - Library-specific Project Plan
  - Schedule for Digitization
  - Quality Assurance Plan
  - Receiving Report
  - Return Report
  - Progress Report

**(C) Management Reports.**

- Library Monthly Report of Scanning Equipment Tests
- Library Quarterly Summary of All Scanning Activity

**2.5. PERIOD OF PERFORMANCE; PAYMENT SCHEDULE.**

**(A) Period of Performance.** All tasks under this task order are to be performed by the Internet Archive on a level-of-effort basis, over the period of performance commencing [no more than thirty (30) days] after the Library's written acceptance of the Internet Archive's proposal in response to this RFQ and continuing for a period of 4.5 months thereafter.

**(B) Payment Schedule.** The Library shall pay to the Internet Archive a fee of \$100,000 per month, payable ~~[in arrears]~~ ~~[in advance]~~ for each month during the period of performance. The Internet Archive will submit monthly invoices to the Library in accordance with Article F.3 of the Contract, and such amounts shall be paid by the Library within thirty (30) days after receipt of invoice. For the avoidance of doubt, as a level-of-effort task order, payments are not tied to the deliverables or the individual tasks set forth in Articles 2.1 and 2.2.