

## Federal Scan Center at the Library of Congress (FedScan) FAQ – Frequently Asked Questions

### Background

In an effort to offer major digitization, digital preservation, and access services for the federal library community, FLICC/FEDLINK recently established an agreement with the Internet Archive (IA) to provide high-volume digitization services to its federal library clients using IA's proprietary Scribe scanning workstations at a facility located at the Library of Congress. Other scanning facilities will also be available for FEDLINK customers located outside the Washington, D.C. metropolitan area.

In partnership with member institutions of the Open Content Alliance (OCA), IA has undertaken mass digitization of public domain resources. Materials are scanned, processed through a rigorous quality assurance protocol, permanently hosted on IA servers, and made available to the public for full text searching and retrieval, free of charge, through an IA web site. FEDLINK members will receive copies of the image files and metadata and will retain rights to these files. The Library of Congress has initiated the first task order under the contract. Other FEDLINK members will be able to procure these services very soon through individual task orders.

Additionally, the Library intends to work with IA and its OCA partners to document best practices for the digitization of brittle and otherwise fragile material and presentation of foldout materials encountered collections, and investigate the development of an open source page turner. Simultaneously, the Library will work with all parties to identify structural metadata elements which should be captured during the scanning process to facilitate navigation in the page turner and allow for more precise and inclusive searching.

### Questions

#### 1. What types of materials can be digitized through the scanning center?

Currently, IA is digitizing books and journals on a large scale. IA is also scanning microfilm on a smaller scale at one scanning center; this capability will also be available shortly at the FedScan center.

#### 2. The description mentions public domain items. What about items that are not in the public domain?

Currently, the IA contract allows for public domain materials only. If the agency would like to digitize other materials, the FedScan center is not the solution. Through the current Preservation Services Basic Ordering Agreement, FLICC/FEDLINK offers specialized digitization services that can accommodate materials inappropriate for the FedScan center. Please contact Anne Harrison at FEDLINK on 202-707-4834 or [anha@loc.gov](mailto:anha@loc.gov) to discuss contracting for these services.

#### 3. Can the facility handle sensitive but unclassified information? Examples: documents limited by ITAR, EAR, proprietary information, etc.

One of the requirements is to provide free public access to digitized materials. If the agency does not want to provide this type of public access, then the scanning center may not be a good solution for these sensitive materials. FLICC/FEDLINK will be offering other more specialized digitization services through IDIQ contracts in FY2010-2014. We will issue a Request For Proposal for these contracts next year. For FY2008-2009, the current FEDLINK Preservation Services Basic Ordering Agreement can accommodate sensitive unclassified materials. Please contact Anne Harrison at FEDLINK on 202-707-4834 or [anha@loc.gov](mailto:anha@loc.gov) to discuss contracting for these services.

#### **4. What is the cost?**

The cost to scan from books is \$0.10 per page for bound books, plus the FEDLINK administrative fee. The administrative fee for FY2008 is 5%. In future years, the fee will continue to be lower than the normal FEDLINK fee of 7.75%, and will likely be in the neighborhood of 5%. We are currently working on the pricing for digitizing microfilm.

#### **5. What is the pricing structure, i.e., does the scanning price go down based upon volume or is there one fixed fee? Is it better to send a large quantity of documents at one time or can we send a box at a time?**

Discounts may be offered for high volume, depending on the nature of the materials. The Internet Archive has been digitizing books for many libraries for several years. They can handle either a one-time or periodic delivery of material. For bound volumes, the working minimum is about 5,000 pages or 20 books. For other materials, there is no defined minimum amount, although standard pricing may not apply. Please contact Anne Harrison at FEDLINK on 202-707-4834 or [anha@loc.gov](mailto:anha@loc.gov) to discuss special requirements.

#### **6. Is this destructive and/or non-destructive scanning?**

It is exclusively non-destructive scanning. There are no plans to expand the operation to include scanning methods that require disbinding or causing any damage to the original materials.

#### **7. What preparation needs to be done to the documents and/or microforms before sending to the scanning facility? Do the documents need to be barcoded? Binding removed?**

The items need to have a unique identifier and enough bibliographic data to track them throughout the digitization process. Most libraries use barcodes and corresponding bibliographic database records accessible through a Z39.50 connection. However, if a Z39.50 connection is not available, the library may provide the data in some other form such as a spreadsheet. The bindings do not need to be removed. Internet Archive uses workstations with a v-shaped cradle and a pair of cameras for non-destructive scanning. Concerning packing: it is highly desirable for libraries to shrink-wrap or otherwise seal the packing boxes.

#### **8. How do we get materials to and from the center?**

Agencies with a government vehicle driven by an employee or contractor of the agency will deliver and pick up materials using the Library of Congress delivery procedures. We are working out other delivery mechanisms for agencies which do not have a government vehicle or driver available.

#### **9. What formats are available for the output files? What is the PPI for images?**

Output files provided:

- Master Images in JPEG2000 format with pixels per inch (PPI/ DPI) as follows:

Height 14.2 in.

Width 9.4 in.

PPI 300

Height 10.6 in.

Width 7 in.

PPI 400

Height 8.5 in.

Width 5.6 in.

PPI 500

- OCR in an XML format or equivalent technology, as mutually agreed upon between the Internet Archive and the customer.

Output, continued:

- XML for metadata from MARC records.
- XML for operational metadata collected during scanning.
- Searchable PDF files bundled for each title or volume.
- Camera Raw, TIFF, PDF-A, or other format for master image storage is currently not available, but may be possible. Each customer will negotiate with Internet Archive about formats.
- Concerning metadata - one of the purposes of the center is to develop methods for efficiently capturing structural metadata during scanning. Descriptive and structural metadata requirements are negotiated for each customer task order.

**10. Does the facility handle oversized documents? Black and white? Color?**

One of the purposes of the facility is to develop best practices for integrating foldouts and oversized documents into a scanning workflow. This capability will not be available immediately upon opening the center, but is expected sometime this fiscal year. All scanning is in color.

**11. I am concerned about the long-term preservation of the files. What is IA doing about that?**

IA is committed to providing permanent storage of and access to the files - the JPEG2000, PDF derivatives, including OCR output. IA maintains redundant repositories on servers in California which are partially mirrored in two other locations.

**12. How are the scanned images provided to customers?**

Customers may download the image files via FTP, HTTP, RSYNC or OAI harvesting.

**13. Can we get the documents back after scanning?**

Yes. Currently it is expected that the agency will pick them up at the scanning center after the digital images and ancillary files have been reviewed and approved. Other options for delivery and pick up are under consideration; see question 7.

**14. I have some rare books. Is the IA staff trained in the care and handling of rare books? Is the facility suitable for storing rare books?**

Yes, the IA staff at the FedScan center are trained in the care and handling of fragile books. The FedScan center is a secured area within the Library of Congress. Inside the center there is a separate secure room for rare or otherwise valuable materials.

**15. My library is far from Washington, D.C. Are there other IA scanning centers closer to me than the Library of Congress?**

Yes, the Internet Archive has several facilities. FEDLINK is completing a contract modification that would allow customers not in the Washington, D.C. metropolitan area to contract for services at the IA facilities located in Champaign-Urbana, Illinois; Boston, Massachusetts; New York, New York, Chapel Hill, North Carolina; London, UK; and San Francisco and Los Angeles, California.

**16. What is the turnaround time?**

The goal is to complete processing by IA within 72 hours from arrival at FedScan center and images being available for review and approval by the agency.

**17. How do I initiate a task order to use the service?**

First, contact Anne Harrison at FEDLINK on 202-707-4834 or [anha@loc.gov](mailto:anha@loc.gov) to begin a task order. The task order includes a statement of work which specifies in detail the quantity and type of items to be digitized and the outputs you desire. Second, register through FEDLINK's Online Registration System. Select the Vendor ID (XZ) for Internet Archive. After a task order has been completed and approved, a delivery order will be generated and sent to Internet Archive to begin work.

**3/5/2008**