

## Federal Scanning Center at the Library of Congress (FedScan) Frequently Asked Questions

### Background

In an effort to offer major digitization, digital preservation, and access services for the federal library and information center community, FLICC/FEDLINK established an agreement with the Internet Archive (IA) to provide high-volume digitization services to its federal clients using IA's proprietary Scribe scanning workstations at a facility located in the Library of Congress. Other scanning facilities will 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 resources in the public domain or otherwise not protected by copyright. Materials are scanned, processed through a rigorous quality assurance protocol, hosted on IA servers, and made available to the public for full text searching and retrieval through an IA web site. In addition, FEDLINK members will receive copies of the image files and metadata and will retain applicable rights to these files. The Library of Congress has initiated the first task order under the contract. Several other FEDLINK members have procured services through individual task orders.

The Library of Congress intends to work with IA and its OCA partners to document best practices for the digitization of brittle and fragile material and presentation of foldout materials encountered in the selected collections and investigate the development of an open source page turner. Simultaneously, the Library of Congress 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 bound materials (books, periodicals, technical reports) on a large scale. IA is also set up for digitization from microfilm.

#### 2. What about items for which the copyright status is not clear?

The IA contract allows for only works that are not protected by copyright. However, FEDLINK offers other digitization services through the Preservation Services basic ordering agreement that can accommodate materials inappropriate for FedScan. 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 FedScan handle sensitive but unclassified information?

Examples: documents limited by ITAR, EAR, proprietary, 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 to certain items, then the scanning center may not be a good solution for these sensitive materials. FEDLINK offers other specialized digitization services through the Preservation Services basic ordering agreement. 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 and microfilm is \$0.10 per page. Scanning foldouts, which requires special equipment and handling, is priced at \$2.00 per page. There may be an additional one time set up fee for loading bibliographic records if the agency is unable to provide standard Z39.50 access to its bibliographic data. FedScan is a FEDLINK Transfer Pay service for which the normal 6.75% administrative fee applies.



**5. What is the pricing structure, that is, does the scanning price decrease based upon volume or is there a fixed fee? Is it better to send a large quantity of documents at one time or send smaller quantities periodically?**

Discounts may be offered for high volume, depending on the nature of the materials. The Internet Archive has been digitizing books for many institutions 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 microfilm, there is no defined minimum amount, although standard pricing may not apply for very small projects. 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 cause damage to the original.

**7. What preparation needs to be done to the documents and/or microfilm before sending them to the scanning facility? Binding removed?**

The items need to have a unique identifier and enough bibliographic data to track them throughout the digitization process. Most agencies use barcodes and MARC records accessible through a Z39.50 connection to the online catalog. If a Z39.50 connection is not possible, the agency may provide a file MARC records or even unique ID's in a spreadsheet. In these instances, Internet Archive will load the provided records onto an IA server for an additional one time set up fee.

The bindings do not need to be removed. IA uses workstations with a floating v-shaped book cradle and a pair of cameras for non-destructive scanning.

Concerning packing: it is highly desirable for the agency to shrink-wrap or otherwise seal the packing boxes or trucks.

**8. How do we get materials to and from the FedScan 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. FEDLINK is working out other delivery mechanisms for agencies which do not have a government vehicle or driver available.

**9. What about the specs – Is the scanning black and white or color? Are there size limitations? What formats are available for the output files?**

Image capture is in color.

Output files provided:

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

PPI 300: Height 14.2 inches, Width 9.4 inches

PPI 400: Height 10.6 inches, Width 7.0 inches

PPI 500: Height 8.5 inches, Width 5.6 inches

- OCR in plain text and in an XML format.
- XML for metadata from MARC bibliographic records.
- XML for operational metadata collected during scanning.
- Searchable PDF bundled for each title or volume.

**10. Does the facility handle oversized documents?**

One of the purposes of the facility is to develop best practices for integrating foldouts and oversized documents into a scanning workflow. Currently, foldout pages can be accommodated.



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

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

**12. I am concerned about the long-term preservation of the files. What is Internet Archive doing about that?**

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

**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 digitization process is complete. Other options for delivery and pick up are under consideration, which would enable federal agencies outside the D.C. area to use the Library of Congress facility.

**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 are trained in the care and handling of rare and 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 other materials of high value.

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

Yes, IA has several facilities. FEDLINK is completing a contract modification that will allow customers to contract for services at the IA facilities located in Champaign-Urbana, Boston, New York, Chapel Hill, San Francisco, Los Angeles, and London.

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

The goal is to complete processing and provide images available for approval within 72 hours from arrival at the FedScan center.

**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. Second, register through FEDLINK's Online Registration system. Select the Vendor ID for Internet Archive: XZ. After a task order has been completed and approved, a delivery order will be generated and sent to IA to begin work.

**18. Is it possible to see a sample before initiating a task order?**

Yes. You may provide up to 5 items to be scanned as a test. You will be given access to the resulting test files to review. Contact Anne Harrison at FEDLINK on 202-707-4834 or [anha@loc.gov](mailto:anha@loc.gov) to arrange for a test scan.