



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 are also 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 are provided access to 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 are able to procure these services through individual task orders.

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 scanning microforms on a smaller scale at one scanning center; this capability may also be available through the FedScan service. IA also has the capability to scan unbound sheets.

2. The description mentions public access and public domain items. What about items not in the public domain?

The library may decide not to provide public access to items. In this case, it may not be required that the items be in the public domain.

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

Yes. The library may choose not to provide public access to their materials.

4. What is the cost?

The cost is \$0.10 per page for most bound books, plus the FEDLINK administrative fee, which is 6.75%. Foldout pages are \$2.00 per page. For items that are odd sizes, there may be a higher per-page price.

5. Is it better to send a large quantity of documents at one time or can we send a box at a time?

The Internet Archive 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.

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 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. In this case, Internet Archive may charge a one-time metadata setup fee. 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 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 may 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.
- XML for metadata from MARC records.
- XML for operational metadata collected during scanning.
- Searchable PDF files bundled for each title or volume.
- TIFF, PDF-A, or other format for master image storage is currently not available, but the JPEG2000 master may be converted to another format.
- Concerning metadata - 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 is now available for \$2.00 for foldout or oversized page. The maximum size is 18 x 24 inches. All scanning is in color.

11. What is IA doing about long term digital preservation?

IA is committed to providing permanent storage of and access to all the files - the JPEG2000, PDF derivatives, OCR output, and metadata. 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 from IA servers via FTP, HTTP, RSYNC or OAI harvesting. The Library of Congress is happy to share scripts for OAI harvesting.

13. Do we get the items 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.

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. Customers not in the Washington, D.C. metropolitan area may contract through FEDLINK for services at the IA facilities located in Indiana, Illinois; Massachusetts; New Jersey, New York, North Carolina, and 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 to begin a task order and to discuss any special requirements. 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 and complete an IAG (or IAG amendment if already registered). Select the Vendor ID (XZ) for Internet Archive. After a task order has been completed and approved, a delivery order will be generated, at which point Internet Archive will begin work.

Other questions? Contact Anne Harrison on 202-707-4834; anha@loc.gov