

Background

In an effort to offer major digitization, digital preservation, and access services for the federal library community, FEDLINK established an agreement with the Internet Archive (IA) to provide high-volume digitization services using IA's proprietary Scribe scanning workstations at a facility located at the Library of Congress. In partnership with FEDLINK member institutions, IA has undertaken mass digitization of public domain resources and other government publications and documents. Materials are scanned, processed through a quality assurance protocol, permanently hosted on IA servers, and, if appropriate, made freely available to the public for full text searching and retrieval through an IA web site.

Questions

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

Currently, IA is digitizing books, periodicals, newspapers, unbound sheets, archival materials, and large format flat materials. IA is scanning microforms at another scanning center available to FEDLINK members.

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

In this case, IA does not provide public access to the materials.

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. 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.

5. What is the cost?

The cost is \$0.10 per page for most bound books, plus the FEDLINK administrative fee, (6.75%; 6.0% for more than \$100,000). Foldout pages are \$2.00 per image; newspapers are \$0.25 per page. For items that are small, oversized, or require special handling, per-page prices range from \$0.12 to \$0.30 per page.

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

The Internet Archive can handle either 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.

7. Does IA use destructive 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.

8. 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 through the digitization process. Most libraries use barcodes and bibliographic records accessible through a Z39.50 connection to the library OPAC. However, if that is not possible, the library may provide the data in some other form such as a spreadsheet. The bindings do not need to be removed.

9. How do we get materials to and from the center?

Agencies may deliver and pick up materials using the Library of Congress delivery procedures. Contact FEDLINK to discuss other delivery mechanisms for agencies which do not have a driver available.

10. 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) as follows:

PPI 300:	Height 14.2 in. x Width 9.4 in.
PPI 400:	Height 10.6 in. x Width 7 in.
PPI 500:	Height 8.5 in. x Width 5.6 in.
- EPub, MOBI, and Daisy formats.
- OCR in an XML format or equivalent technology, as mutually agreed upon between the Internet Archive and the customer.
- XML for MARC data (if applicable) and for operational metadata collected during scanning.
- Searchable PDF files bundled for each title or volume.

Note: TIFF, PDF-A, or other format for master image storage is currently not available. Descriptive and structural metadata requirements can be negotiated for each customer task order.

11. Does the facility handle oversized documents? Black and white or color scanning?

Yes, this capability is now available for \$2.00 per image; \$0.25 for oversized pages. The maximum size for a single image is approximately 18 x 24 inches at 300 ppi. All scanning is in color.

12. 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 masters, PDF derivatives, OCR output, and metadata. IA maintains redundant repositories on servers in California which are partially mirrored in two other locations.

13. 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.

14. 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 shipping and delivery, such as commercial carriers, are possible.

15. 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.

16. 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 the quantity and type of items to be digitized. Second, register through FEDLINK's Online Registration customer account management system and complete an inter-agency agreement (IAA), or an IAA 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

October, 2012