

FACT SHEET

Developments Since the Depository
Library Council Update, April 2006



American Library Association Annual Conference June 22-28, 2006, New Orleans, LA

ANNUAL INTERAGENCY DEPOSITORY SEMINAR –

http://www.access.gpo.gov/su_docs/fdlp/events/interagency.html

- The 2006 Annual Interagency Depository Seminar will be held July 31 - August 4, in Washington, DC. The free seminar, open to depository library staff, provides an opportunity for both new and experienced documents personnel to review basic materials and learn about new products directly from Federal agency staff. Tours of Federal agency libraries are included. For further information about the seminar, contact the Education and Outreach staff by telephone on (202) 512-1119; by fax on (202) 512-0016; or by e-mail at ylouden@gpo.gov.

FALL DEPOSITORY LIBRARY COUNCIL MEETING AND FEDERAL DEPOSITORY LIBRARY CONFERENCE –

- The Fall 2006 Depository Library Council Meeting and 15th Annual Federal Depository Library Conference will be held October 22–25 in Washington, DC, at the Hyatt Regency Washington on Capitol Hill. Registration and hotel reservation information will be announced on the FDLP desktop (www.access.gpo.gov/su_docs/fdlp/index.html) and on the GPO-FDLP-L listserv. The meeting, which is free and open to the public, will cover current issues concerning Federal Government information and the FDLP.

ANNUAL ITEM SELECTION –

http://www.access.gpo.gov/su_docs/fdlp/tools/amendment.html

- The annual Item Selection Update Cycle began on June 1 and will run through July 31, 2006. This is an opportunity for Federal depository libraries to update their selections of tangible and electronic Federal information products.
- Additions to depository library item selection profiles will take effect on October 2, 2006.

REVIEW OF ITEM NUMBER SYSTEM/SELECTION –

http://www.access.gpo.gov/su_docs/fdlp/selection/index.html

- GPO received 177 responses to the questionnaire distributed at the spring Council meeting. A summary of the responses is available at the Web site listed above. GPO will use this data in finalizing new models for depositories in selecting online and tangible publications.

DIGITIZATION OF LEGACY PUBLICATIONS –

<http://www.gpoaccess.gov/legacy/index.html>

- The Joint Committee on Printing has authorized the digitization of legacy publications for a 6 month demonstration project. Priorities for digitization will include legislative and regulatory material that expands the coverage of the most popular *GPO Access* databases.

- GPO is seeking donations of the following titles: *Code of Federal Regulations: 1990-1996; List of CFR Sections Affected: 1990-1996; Public and private laws, 101st - 103rd Congress.*

WEB DISCOVERY & HARVESTING –

GPO's pilot project to automatically harvest online Environmental Protection Agency (EPA) publications continues. The goal of the pilot is to learn about available methodologies and technologies for automated Web discovery, assessment, and harvesting of publications within scope of GPO's information dissemination programs.

The results and findings of the pilot will be used as a first step in developing more specific requirements and specifications for the overall harvesting solution that will be implemented in conjunction with FDsys.

The two vendors involved in the pilot project have each completed their first crawl and harvest of EPA publications. One vendor identified over 12,000 qualifying EPA publications within the scope of the FDLP while the other discovered over 120,000. The difference in publications found appears to be a result of the depth of the vendor's first crawls and the parameters set for each harvest. We expect these numbers to converge by the end of the project.

GPO's analysis of the first crawl results has revealed an error rate in assessment (determination whether an identified publication is in or out of scope of GPO programs) of approximately 50%, meaning that half of the publications harvested are not were not determined to be within scope. While this number is higher than expected, many issues were identified by GPO and the vendors which resulted in modification of harvester rules and instructions that will likely mitigate many of these errors.

In addition, GPO found that PDF, text, or MS Office file formats were correctly harvested in general, where harvested copies of html Web pages were less accurate.

Both vendors have now reconfigured their harvesters to improve publication assessment and are now in the process of completing the second of three crawls to take place during the pilot.

The first crawl of the pilot focused mainly on the harvesting of "surface level" documents, meaning content that resides on Web pages that is easily reachable by Web crawlers. For the second crawl, GPO has identified several query-based databases for each vendor to crawl in order to test their capability of extracting in-scope content from databases.

At the conclusion of the EPA pilot project in September, harvested content will be delivered to GPO. At that time, in-scope harvested EPA material will be incorporated into the FDLP and Cataloging and Indexing workflow and will be classified and cataloged.

ESSENTIAL TITLES (TITLES THAT NEED TO REMAIN IN TANGIBLE FORMAT) –

http://www.access.gpo.gov/su_docs/fdlp/coll-dev/proposed-estitles.html

- Proposed methodology for determining an essential title was posted to the FDLP Desktop the first week of June. The methodology is based upon the collection development decisions of

depository coordinators by using item selection profiles. The proposal provides for an essential titles list for the aggregate of libraries as well as for lists by library type.

- GPO is seeking comments until August 1st. You may submit comments by using the form at the above URL.

FEDERAL DEPOSITORY LIBRARY HANDBOOK –

http://www.access.gpo.gov/su_docs/fdlp/handbook/index.html

- Volunteers from the depository library community have worked diligently to create draft chapters for the new *Federal Depository Library Handbook*. This new Handbook will combine and update the existing *Federal Depository Library Manual* and *Instructions to Depository Libraries* into one online document. The draft chapters are posted on the Handbook web page at the URL cited above.
- Comments on draft chapters 1, 4, 8, 9, 10, 11, and 14 are being reviewed and incorporated as appropriate. Additional draft chapters will be posted for comment as they become available. Please use the web form on the Handbook web page to submit comments. We hope to post a final version of the majority of the chapters by October.

INTEGRATED LIBRARY SYSTEM –

<http://catalog.gpo.gov/>

- GPO launched the enhanced version of the *Catalog of U.S. Government Publications* (CGP) in March 2006. This is the online public access catalog (OPAC) module of GPO's new integrated library system.
- The new and improved CGP currently offers more than 500,000 records to both historical and current Government publications, from 1976 forward. There is an advanced search function for professionals as well as a more basic search function for the general public.
- New features of the CGP include "Other Catalogs to Search" on the advanced search page that provide specialized access to unique groups of publications in the Catalog such as access to the U.S. Congressional Serial Set, periodicals, and Internet publications. The "Locate in a Library" feature within the displayed record will find a Federal depository library that has the particular publication if it is in the current item selection profile. The updated New Electronic Titles feature is being tested and is scheduled for implementation later this summer.

CATALOGING CONTRACT –

- Vendor proposals have been received for procuring contracted cataloging services for Library Services and Content Management for pre-1976 material and titles in a small cataloging backlog. A formal evaluation process has commenced, following Federal procurement regulations and guidelines. Once this process is complete, an award can be issued. It is expected that this process will take approximately 90 days, with an award before the end of the fiscal year.

CD-ROM DATA MIGRATION –

- GPO is working to devise a data migration strategy to ensure that information on the CD-ROMs that have been distributed to depository libraries remains accessible. GPO staff conducted a review of a sample of CD-ROMs from three agencies: Department of Education, Department of Justice, and US Geological Survey. Discs were prioritized based on whether the content was found on the Internet in order to ensure that content that was not available elsewhere was addressed first. The list of titles reviewed will be released in June for comment from the depository community. Comments received will help GPO staff to identify appropriate discs for in-house testing.

- Additionally, GPO has been working with the National Institute for Standards and Technology (NIST) to test potential application of emulation technology to make selected discs accessible over the Internet. Future analysis will include testing the scalability of this application, as well as cost/benefit analysis of virtualization vs. migration.

LOCKSS –

http://www.gpo.gov/su_docs/fdlp/lockss/

- Since launching the LOCKSS Pilot Project in June 2005, GPO has made the current volumes of nine Government e-journals available to the 21 participating libraries. One remaining title, *Environmental Health Perspectives*, is currently being harvested.
- The first “real world scenario,” designed to test how LOCKSS caches respond when content is removed from a publisher’s website, has been conducted and data is being analyzed.
- The second “real world scenario” will allow participants to watch how damaged content repairs itself. Preparations are underway, and the test will be conducted in early June.

AUTHENTICATION –

- GPO is engaged in a major authentication initiative designed to assure users that the information made available by GPO is official and authentic and that trust relationships exist between all participants in electronic transactions. This authentication initiative, which employs Public Key Infrastructure (PKI) technology, will allow users to determine that the files are unchanged since GPO authenticated them.
- GPO is procuring the necessary tools and capabilities to automate the application of digital signatures on Adobe Acrobat Portable Document Format (PDF) files. A Request for Proposal (RFP) was issued in late 2005 and the resulting bids have been evaluated and a contract award is pending from Procurement.

DISASTER RECOVERY –

- GPO has completed the first phase of the Disaster Recovery project, which includes a full global offsite backup of *GPO Access* data, static content delivery of the gpo.gov and bookstore.gpo.gov domains, and a Proof of Concept for five *GPO Access* applications (*Federal Register*, *Code of Federal Regulations*, *Congressional Record*, *Congressional Bills*, and *Congressional Hearings*) that emulate the current search system while providing faster performance and improved search results.
- The remaining *GPO Access* applications are available on the site in case of disaster, but have not been reformatted for optimum use. A second phase of the DR project will result in the migration of the remaining WAIS applications, at which time GPO will have a complete backup system that will offer assured availability to all of GPO’s Web services and also maintain an archival copy of GPO’s data for redundant, offsite storage. Work is expected to begin this summer.

FUTURE DIGITAL SYSTEM –

<http://www.gpo.gov/projects/fdsys.htm>

- GPO’s Future Digital System (FDsys) will manage, preserve, provide version control and access to, and disseminate authentic Government information. In FDsys the Master Integrator will function much like a general contractor in a remodeling job and will work collaboratively with GPO to identify solutions for the FDsys functions and to integrate the various components, technology and applications. The Master Integrator Request for Proposal closed in late May, and the source selection process has begun. A contract award is expected in July.
- Version 2.1 of the FDsys requirements has been released and is available on the FDsys website.

- A cross-functional team from the Office of the Chief Technical Officer, Information Technology and Systems, and Library and Content Management Service are collaborating to integrate the ILS with FDsys. Recent accomplishments include data conformity testing and a systems security evaluation of the ILS X-server. The team has identified a series of next steps and action items to progress toward a full integration of the systems.

PUBLIC-PRIVATE PARTNERSHIPS (SALES) –

<http://www.gpo.gov/bidopps/majoracquisitions.htm>

- GPO is seeking innovative relationships with the private sector to create new business models for its Sales Program, and is interested in employing the same best business practices and processes to improve service to FDLP partners, International Exchange Service (IES) libraries, and By-Law recipients. GPO released a Request for Proposal on April 12, 2006 and expects to receive initial proposals by July 10, 2006.

DISTRIBUTION AND OTHER STATISTICS –

- During the first five months of FY 2006 (October 2005- April 2006) GPO distributed a total of 3,237,100 tangible copies of 5,819 titles (this includes print, microfiche, CD's, DVD's and maps). This is an increase of 477,791 copies and 805 titles compared to the same period in FY 2005, when GPO distributed 2,759,309 copies of 5,014 titles.
- Through June 1, 2006, 8,166 bibliographic records and 6,000 PURLs have been created.
- Through June 1, 2006, 9% of depository titles are tangible-only, where there is no electronic equivalent. About 25% of our titles are produced in both tangible and online formats. 91% of depository titles are available online.

GPO ACCESS STATISTICS –

<http://www.gpoaccess.gov/>

- In May, 1,537 online titles and 522 PURL links to agency titles were added to *GPO Access*, for a total of 2,089 new online titles. For fiscal year 2006, 7,763 online titles and 4,026 PURL links to agency titles were added to *GPO Access*, for a total of 11,789 new online titles. These additions bring the total number of titles on *GPO Access* to 201,212 and the total number of titles linked from *GPO Access* to 44,361 for a grand total of 246,573 titles available through *GPO Access* as of May 2006.
- Retrievals from *GPO Access* since its inception in 1994 are now 2.65 billion. 255 million documents have been retrieved through April in FY2006 and average nearly 36.5 million a month, or 1,214,000 a day.
- In FY06, 9% of the titles are tangible-only, with no electronic equivalent. About 25% of the titles are both tangible and online. 91% of the titles for FY06 are available online.