

Proceedings of the 1st Annual Federal Depository Library Conference, 1992 April 5-10, 1992

Loading the GPO Tapes

Remarks by Larry Romans
Government Information Librarian,
Central Library, Vanderbilt University
Nashville TN
At the 1992 Federal Depository Conference
April 7, 1992

In 1964, when I first started working with government publications, documents depositories were little fiefdoms, with supplicants coming to the documents staff to plead for access to government information. Those were the days!

Even a basic question could be a treasure hunt. There was no CIS Index, no ASI Index, and Readex was distributing hearings and the serial set on opaque microprint. The Monthly Catalog had just a single index with subject headings that seemed to change from year to year. Patrons were very grateful if you could find anything for them, but most library users avoided documents. We were always up-to-date in processing the documents, because we did not have very many patrons "bothering" us.

The best thing that happened to U.S. government documents was to make these major changes in the Monthly Catalog beginning July 1976: using the MARC format, cataloging according to Anglo-American Cataloging Rules (AACR), and joining the OCLC online cataloging network. This meant that, first, the Monthly Catalog was consistent and more usable, and, second, the cataloging in electronic format could later be used to recreate the Monthly Catalog on compact disc and then to load tapes of GPO cataloging for the Monthly Catalog (hereafter called the GPO tapes) in online catalogs.

The compact-disc version of the Monthly Catalog not only allows staff to find material faster, but also allows patrons to feel, for the first time, that they can identify material by themselves without staff intervention. Even if the patrons do not always find the best government information on their subjects, the sense of being able to access the information on their own increases the chances that they will return to use documents again. While this is a wonderful development, you still have to get patrons to come to the documents depository before they can use the compact disc, unless it is loaded on a network. Loading the GPO tapes on the online catalog greatly increases the number of patrons who get to the documents area. That is because patrons who never considered that there might be documents on their subjects are finding records for documents in the online catalog when they search for books.

Documents at Vanderbilt

The Jean and Alexander Heard Library system at Vanderbilt University has an integrated NOTIS online catalog. There are two U.S. government depositories: the Central Library depository dates

from 1884; the Law Library from 1976. Both are selective depositories, with Central selecting about 60 percent of the items numbers; Law about 20 percent. The Central depository acts as a "subregional": we are responsible for keeping the archival copy for Tennessee of documents from agencies like the Department of State, the Executive Office of the President, and the Congress. We send depository documents to Vanderbilt's Education, Management, Science, and Medical Center libraries. Before we loaded the GPO tapes in 1988, ninety-nine per cent of U.S. documents were not listed in the online catalog.

When we loaded the tapes, the Central documents staff consisted of one professional, one library assistant, and student help, including graduate students who provided reference service at night and on weekends. We were also responsible for international documents. The professional also provided limited service at the Central reference desk and is the Political Science bibliographer. The Law Documents staff consisted of one professional and one half-time library assistant.

As at most libraries, in 1988 government publications were among the least used resources in Vanderbilt's libraries. But our use of documents at Central Library had increased dramatically over the previous few years because of:

1. increased service hours,
2. compact-disc access,
3. increased emphasis on documents in bibliographic-instruction sessions, and
4. service-oriented staff members.

The GD Committee and Its Proposal

In 1987, Flo Wilson, then Assistant Director for Library Systems, appointed a committee to examine whether Vanderbilt should load the GPO tapes in the online catalog. The Government Documents (GD) Committee, which held regular meetings during the next year, included the Central Documents librarian, the Central Documents assistant, the Law Documents Librarian, and representatives from library systems, public services, and technical services. One of the main accomplishments of the Committee was to co-opt the initially skeptical members of the Committee (mostly technical services people). By the time the GD Committee issued its report on Sept. 23, 1988, all committee members supported the conclusions stated in the report.

The goal behind loading the tapes was to make patrons aware of the useful documents resources available on their topics. We wanted to accomplish this by providing access on NOTIS for all U.S. documents received since 1976 that are part of currently-received item numbers. No documents received before 1976 were included in the project, because records are not easily available in machine-readable form for that period.

A number of important issues have to be decided before loading the GPO tapes. I will discuss some of these issues and tell you what we did at Vanderbilt and, occasionally, what we would do if we could do it all again.

Issue #1--Should You Integrate the Documents with Other Items in the Online Catalog or Should You Create a Separate Database?

The most divisive issue our committee dealt with was whether to integrate documents into the main online catalog or to maintain a separate database. The three options were:

1. If the GPO tapes are integrated by loading them into the main online library catalog, responsibility for maintaining the records can be shared with existing non-documents

technical services staff, but the documents unit loses control over the records, and changes that need to be made only to documents records will be more difficult to make. Authorization tables cannot be tailored to allow documents staff to work only on documents records, but catalogers will hesitate to have documents staff authorized to change all records.

2. If the GPO tapes are loaded as a separate database, documents records will be retrieved only if someone realizes that there may be documents on her topic and decides to search the documents database. Particularly if there are a number of databases available at the online catalog terminals, a patron is not likely to select documents and, thus, may miss important resources. A separate database promotes the idea that documents are different from other library materials. However, technical services staff may be less anxious about departing from local control of cataloging practices, and computer response-time may be better than if the records were integrated. A separate database is not the politically correct position and is chosen only by documents librarians who are tools of the technical-services commie conspiracy.
3. If the GPO tapes are loaded as a separate subset of the main catalog (in NOTIS lingo, a separate processing unit), a search for library materials will also retrieve documents records, control of the records can remain in the documents unit, and universal changes applicable only to documents can be accomplished easily. Patrons can easily pick out documents because of the different processing-unit code.

We decided on the third option: that the GPO records should be accessible as part of an integrated NOTIS database as a separate processing unit, rather than be loaded into a separate database.

Public-Services Reasons

We felt that bibliographic records for documents should appear along with records for other library material, so that patrons do not have to perform an extra search to access the documents. Since even librarians are not aware of the vast resources available in depositories on almost every subject ranging from art to zoology, we did not expect most patrons to be aware that they might find documents on their subjects and should perform that extra search. We wanted to make documents as much like other library resources as possible.

Technical-Services Reasons

The technical-service reasons for loading the GPO bibliographic records into a separate GD subset or processing unit were:

1. the staffs of both documents departments will receive documents,
2. qualified personnel from all processing units can share responsibility for maintaining the GPO records, rather than any one unit having sole responsibility,
3. some changes after loading may need to be made only for GPO records, and
4. Central documents and Law documents would not need separate records, thus saving space on the NOTIS database. Groups of online records may be easier to manipulate when they are in a separate subset or processing unit than when they are integrated with other records. We hoped to minimize "post-load" processing by developing sufficiently detailed and comprehensive load specifications. However, it seemed best to load records so as to make

"post-load" processing as easy as possible.

Issue #2--Should You Include in the Online Catalog All Documents, Only Those Selected by the Library, or Only the "Important" Documents?

At Vanderbilt, since we want to make people aware of the Documents that we have, we proposed to load only records for the U.S. documents that we select, not for all GPO records. We do not want to create unrealistic expectations by listing documents that we do not have in the library system. In principle we wanted to include all documents we receive; in practice we excluded some "unimportant" ones. These are some of the questions we considered about which documents to include in the catalog:

1. Should you include GPO tape records for documents that are already in the online catalog? We excluded those documents from the tapeload. U. S. documents already on NOTIS have LC class numbers and are housed outside the Documents Department. For serials, we profiled to get future records for documents already listed online, so we will be notified when a serial record is updated. Should you add all documents to the online catalog? Some documents, like provisional Census pieces, may be important enough to list in the catalog, but others, such as separates, reprints, preprints, and posters, may not be. Should you add records for documents that are the equivalent of chapters in a book? We excluded most of these, especially if they are reissued in another form. Examples are slip laws, preprints, and sections of the POMS Social Security Manual. We accidentally included treaties, but we are going to delete them. We included House and Senate reports only because the Law depository was going to include them, and we did not want to create the impression that they were available only in the Law depository. Should you add maps? If most of your patrons do not know the quadrangle names, you may want to add only the generic "mother" record for each state (listed in Administrative Notes). Should you add documents that were issued as part of item numbers that later were dropped? If the item number was deselected, we thought the documents issued under that number were not worth processing, much less cataloging.
2. Should you continue to select an item number for documents that are not worth including in the catalog? Generally, we decided to drop the item number. A thorough review of item numbers was an important part of profiling for the online catalog.

Eventually, some kind of records may have to be created for some of these documents. We are still dealing with some of the problems that these documents may cause: where they should be housed, what classification system should be used, and what processing unit will maintain the records.

These are other issues that will affect the accuracy of the records that you load from the GPO tapes:

1. The date that you selected and began receiving an item or that you deselected and stopped receiving an item will not neatly match the years that you specify on the profile. Err on the side of receiving too many records. It is easier to delete a few records than having to amend your profile or derive the records from OCLC or RLIN.
2. You should develop guidelines for records for superseded documents and those you weed after five years.
3. Regardless of how well you profile, some documents you received will not have been assigned the correct item number and will not appear on your tape; other documents that

you did not receive will have been assigned an item number you profiled and will be on the tape.

Issue #3--Should You Add Retrospective or Only Current Documents?

Issue #4--How Accurate Are the Tapes and Have Duplicates Been Deleted?

Issue #5--Which Vendor Should You Select?

These three issues are interrelated. Whether a vendor supplies retrospective tapes and whether the vendor has relatively accurate tapes and has eliminated duplicates probably will influence which vendor you select. OCLC and Marcive supply current tapes: only Marcive provides the 1976-1991 retrospective tapes.

Most libraries want to eliminate multiple records for the same item or serial before the tapes are loaded. Each year an additional record appears in the Monthly Catalog (and thus on the GPO tapes) for all periodicals, annual reports, and some other serials. In addition, a multi-part monograph may have more than one record on the tapes. For example, parts 1 and 2 of a hearing may appear in the Monthly Catalog in one issue and parts 3 and 4 may appear in another issue. There will be two records on the GPO tapes for that hearing. Some serials may have more than one serial record and more than one monographic record. If the duplicate records are not dropped before the tapeload, then they will have to be identified and purged one by one.

OCLC eliminates duplicate records by dropping records with an AVR (for "availability record") in the 949 field. This eliminates a large percentage of the duplicates. Marcive uses retrospective tapes that have been "cleaned up" by librarians from Texas A & M University, Rice University, and LSU. Corrections and changes posted in the Monthly Catalog have been applied, certain fields retagged, multiple SuDocs numbers identified and explanatory notes added, linking fields have been added to serials records with title and corporate author changes, duplicate serial and periodical records eliminated, and names, titles, and subjects have been run against current LC authorities.

Current Marcive tapes continue to be corrected. When Marcive receives a new GPO tape (about once a month), records that are possible duplicates are removed from the tapes sent to participating libraries. These records are then reviewed by a committee of technical services librarians to eliminate true duplicate records. The records sufficiently different to warrant being sent to the participating libraries are added to the tape sent the next month. Since we can not overlay the new records over the old records, at Vanderbilt these new records are duplicates. For such a small number of records, deleting the old records is not a major task.

Problems remain: there are still some multiple monographic and serial records for issues of the same series, wrong item numbers assigned, typos, and so forth. But this is a small fraction of the problems we would have if the cleanup had not occurred.

Issue #6--What Should You Do About Documents that the Library Receives that Are Not Yet on the Tapes, and What Should You Do about Documents that Are on the Tapes but the Library Has Not Received (Such as Those in the GPO Microfiche Backlog)?

Issue #7--Should You Check in Serials or Monographs Online?

There are a number of record-maintenance questions to be resolved:

1. Which unit will be responsible for individual or global changes in authorities, such as names, subjects, and series entries? Which unit will be responsible for changing access points, such as SuDocs number and location, and for correcting typographical errors?
2. Which unit will be responsible for adding copy holdings, volume holdings, order-pay-receipt records, and linked-item records?

At Vanderbilt, since no one on the documents staff is a cataloger, General Technical Services (GTS) is responsible for individual or global changes in authorities. All authorized personnel in our department and in GTS can change access points or correct typographical errors, but the documents staff make most of the changes because we deal with the records on a regular basis. Our department is also responsible for adding holdings, and order-pay-receipt and linked-item records.

You can avoid the two biggest problems that we have encountered with the records if you know about them ahead of time. First, some serial records have more than one SuDocs number in the 086 field. You should be able to specify to your vendor that you want only the most recent SuDocs number or, perhaps, that you want the most recent number first. We did not so specify, so the oldest SuDocs number appeared first on the 086 field. Since the NOTIS program only looks at first SuDocs number, the oldest SuDocs number appears in our NOTIS catalog. There is at least one serial record in our online catalog that has an FS SuDocs number! Since most of our patrons want relatively recent issues, we have to change the SuDocs number record-by-record.

Second, some documents are not sent in the format (paper vs. microformat) that we expect. When we set up locations for the profile of item numbers, we wanted to distinguish "Documents" as a location for paper copies and "Documents Microfiche" for microfiche copies. Unfortunately, even if the List of Classes and the item number card specify that documents issued under a certain SuDocs stem are paper, you will sometimes get a microfiche copy, and vice versa. This results in thousands of documents that have an incorrect location listed because of format. Luckily, most of these "wrong-format" documents can be identified by looking at the 074 (item number) field. If a document that is supposed to be sent in paper format is actually sent in microformat, usually the 074 field will have "mf" or "microfiche" after the item number. The vendor should be able to program your tape load so that the paper location will be given unless the 074 field indicates microformat. The only place that I know this will not work is for hearings, which are issued in both formats. Other than hearings, our sample of documents from the Departments of Health and Human Services and of Transportation suggests that 80 to 90 per cent of the format location problems could be corrected by the "074 fix."

There may be a significant lag from the time you receive a monographic document until its record appears on the GPO tape and is loaded into your online catalog.

1. You may receive documents well before they appear on the tapes.
2. You may load records for items that are not yet on the shelf. This is a big problem for hearings on microfiche.

Receiving documents before they appear on the tape load may not be a great problem for patrons because they are not likely to ask for documents that are not in your catalog or in the Monthly Catalog. However, depository regulations require that documents be available in a timely manner. In addition, delays in receiving tape records for documents you already have creates a big problem if you are checking in documents online. For monographs, if you process the document before the record appears in your catalog, you have a choice:

- (a) check in the way you did before you started getting the tapes,
- (b) create preliminary records that you will have to delete or link to the tape record when it is added, or
- (c) keep your documents in a holding area until the tape record appears in the catalog.

At Vanderbilt, we decided on option (a): to continue to check in for monographs and monographs in series in our shelflist, since many publications are received weeks before the records for them are added to NOTIS. Alternatives (b) and (c) require handling the documents or their records more than once. As we monitor the timeliness of the records, we may change this decision.

We recommended that serial records and monograph records be handled alike, except that check-in records for periodicals and some serials might be created and maintained online.

Loading a record before receiving the document is a problem because patrons expect you to have everything that is listed in the catalog. The main problem is with hearings, which we receive in microfiche. Patrons have been very understanding, and can go to our Law Library, which receives hearings on paper--usually before the records appear online.

Issue #8--Should You Load All of the GPO Records at Once or in Stages?

1. If you develop and load a test tape of 100-200 item numbers for documents of various types, frequencies, and locations, you may see some problems that you had not anticipated.
2. If you decide to load in stages to minimize the impact on your system, you may want to do the profiling for all stages at the beginning, even if you give the vendor only the profiling for the first stage. Otherwise, you may have to duplicate effort by going through the entire item list again when you profile for subsequent stages.

Stages of Implementation at Vanderbilt

We planned to develop a profile for a test tape of 100-200 item numbers, for monographic and serial titles of various types, formats, and frequencies, but we got busy with other things and did not develop the test profile. That turned out to be a big mistake. After loading the test tape, we could have corrected some problems before loading the full retrospective tapes. So instead of making corrections to a few hundred records, we may have to make corrections to thousands of records.

At Vanderbilt, rather than loading all post-1976 records at once, we loaded the tapes in two stages. In the first stage, we loaded the retrospective and current tapes for about half of the item numbers we select (mostly social-science-oriented material). In the second stage, we loaded the retrospective and current tapes for the other half (mostly natural-science-oriented material).

Issue #9--What Will Loading the Tapes Cost?

1. What will be the charge for the retrospective records? For the ongoing monthly tapes? The profiling cost is \$200 with OCLC. Any new locations that you add to your OCLC catalog profile cost \$40 an hour to add. The profiling cost is \$75 with Marcive. Your generic or default location (ours was "Central Documents") is free with Marcive, and each additional location code (one we used was "Central Documents Microfiche") that you profile costs \$250. Some specialized locations, like Documents Reference Microfiche, have so few titles that it is not worth \$250 to include that location in your profile. Instead assign it a more general location

and change the few titles to the specialized location on your online catalog. For current GPO tapes, Marcive charges an annual fee rather than a fee per record: \$2,200 for depositories that select 42 per cent of the item numbers or more; \$1,300 if you choose less. OCLC products and services are generally purchased through a regional network, which usually adds a surcharge. For current GPO tapes, OCLC charges regional networks 4 cents per record and 12 cents per record to set holdings on the OCLC database. Some networks add a surcharge up to 4 cents per record. OCLC charges a library subscribing directly (not through a network) 15 cents per record to set holdings. The minimum charge for Marcive's retrospective tapes is \$2,000. Selective depositories pay 5 cents per record. You can get all records, without multiple holding codes, for 4 cents a record. Retrospective authorities processing is 1 cent per record (\$150 minimum). OCLC can set holdings on their database from Marcive tapes for 7 cents per record. 2.

2. Will additional computer capabilities have to be purchased? At Vanderbilt we had to add a DASD (direct access storage device--like a hard drive for a mainframe computer). In addition, we acquired two dedicated terminals, one for staff and one for patrons. Money has also been needed for supplies, such as computer paper and barcodes. 3.

Will additional technical-services staff have to be hired to deal with the online records? Will additional public-services staff have to be hired to deal with increased use of the material?

We really needed to hire two staff members, one each in technical services and in public services to deal with the increased number of records and patrons. Instead, we were funded to add one staff member--half-time for technical services and half-time for public-services.

Adding a staff member made expanding our small office space critical. Luckily some money became available at the right time, so the Library Administration spent about \$75,000 to build a new office. 4.

Will money have to be paid to other units to train documents staff?

We have been lucky, because Jean Wright, our systems Research and Development Librarian, has been training us to deal with the records. She acts as a liaison between GTS and the depository. Thus, we have not had to pay for training.

Issue #10--How Should You Publicize the Tapes?

While use of our government publications has not increased exponentially, use at most other depositories after loading the tapes has at least doubled. Loading the records online will publicize itself as patrons retrieve citations to documents as they perform searches for books and other materials. Not only may you want to minimize publicity, but also you may want to load the tapes in stages to lessen the effect on use at any one time.

Issue #11--How Should You Barcode and Check Out Documents?

The options for barcoding the documents are straightforward:

1. Do not barcode any documents. This option may be appropriate if you are not going to circulate your documents or if you will circulate them manually rather than online. Barcode documents as needed. This option avoids a big barcoding project, but it will take longer to check out material, because you will have to barcode each piece to be checked out.
2. Barcode all documents at once. This option requires a barcoding project, but makes it easier

to check out material.

At Vanderbilt we chose the second option: barcode documents as needed. One reason for our decision was that a large percentage of our documents never circulate. A second reason is that the most time-consuming part of checking out a document online is creating the item record rather than attaching a barcode. We create a linked item record for those documents that are already in the online catalog and an unlinked item record for those that are not in the catalog. We then barcode the item and put on a date-due slip.

Issue #12--Which Unit Processes the Documents and Which Checks Them Out?

I have surveyed a large number of libraries that have loaded the GPO tapes. Most depositories that processed their own documents before going online continue to do so; most depositories that checked out their own documents before going online continue to do so. Government publications are checked out at the main circulation desk for many libraries that did not circulate documents before going online or that circulated very few documents. At Vanderbilt the documents are checked out at the main circulation desk after the documents staff create item records and barcode the documents.

CONCLUSION

Providing greater access to documents by creating online records will have a relatively small purchase price for the tapes. There will be a much larger and somewhat hidden ongoing price, primarily for cleaning up the database, creating item records, providing reference service, and shelving more items that have been used. However, if you are going to spend \$100,000 a year for depository space, materials, and staff, it makes sense to spend a little more so that the government publications are actually used. A relatively small amount of money spent to make documents accessible to patrons through an online catalog makes more library resources truly available to patrons more inexpensively and efficiently than virtually any other comparable expenditure.

This is a condensed and edited version of a presentation given at the 1992 Federal Depository Conference in Arlington, Virginia, on April 7, 1992. It benefited greatly from papers prepared by Barb Hulyk, Julia Wallace, Gary Cornwell, and others. They are responsible for any errors that remain. Diana Ukleja of OCLC and Joan Chapa of Marcive provided helpful information. Mike Morgan, Anne Reuland, and Dale Manning proofread this paper and suggested stylistic changes. I am grateful to Sherre Dryden and Bill Robnett for providing Vanderbilt funds to attend the conference and, most especially, to Gretchen Dodge and Diane Grantham for condoning my professional activities. LR--6/8/92

NOTE: Mention of individual companies, products, or services does not constitute an endorsement by the Government Printing Office.