

Proceedings of the 6th Annual Federal Depository Library Conference

April 14-17, 1997

Managing the Depository Database: Some Opportunities with Shared Technology

Part II: BDL and GPO: Creating a Database of Fundamental Depository Information for Web Access by Depository Libraries

Thomas Tyler, University of Denver Library
Denver, CO

Just in Time is a Japanese production and distribution model copied by many U.S. and international manufacturers. It operates on the assumption that, as a producer, you have what you have to have when you need it. Large inventories and the facilities to house and manage them are not needed. Resources are used wisely. The benefit, we are told, is greater efficiency and lower costs in production. Savings can be passed on to the customer through lower prices. In the marketplace, such a producer gains a "competitive edge" and hopefully, greater market share.

What would **Just in Time** mean for the Federal Depository Library Program?

- Separates would never stack up on a shelf or work table awaiting shipping lists
- Electronic representation of Class/Item Number and minimal bibliographic information would be immediately available to:
 1. Create **abbreviated records for the OPAC** with item level information such as bar code number, location, and class number;
 2. Create appropriate **labels** with SuDocs Number, shelving location, property and depository stamp information;
 3. Maintain **statistics** for receipt and processing; and

4. Generate "**pull slips**" for material being superseded or for class corrections.

- Many of the activities associated with dealing with **corrections** in cataloging and classification would be significantly diminished.
- **Typing/data entry errors** for Item Numbers and Class stems would be a thing of the past.
- **Reports** could be generated to simplify the correction of records in the local OPAC or depository database.
- Local information for **new depository items** could be captured and added to your depository's profile for subsequent sharing with depository partners or vendors.
- Recording **rainchecked and claimed** materials could be significantly simplified.

While there are certainly more benefits than those enumerated here, I'm sure that most of us would be at least moderately satisfied if it were possible to achieve the above in a reasonable amount of time.

Actually, we are very close to this nirvana state right now. In the past year LPS has made several of the basic data tools needed for **Just in Time**-type processing available to libraries in electronic format. The Internet and the World Wide Web give us the means to acquire this data. To go further there are a few additional data files that either need to be modified or added to those that are currently available.

Current Depository Library Data

There is an abundance of electronic data available to depository libraries to help in the daily processing of depository documents, but there are some obstacles to the effective use of this data. One or two of the basic tools we need to move to "just in time" are not available at the time we need them or in a form that lends itself to immediate use.

File Availability on the World Wide Web/Internet

When I look at what data is available for use by individual depository libraries on the Internet/World Wide Web, I see the following:

- **Files or Web pages created for viewing with a Web browser.** Each of the 1200+ pages/files served from my Web site - Basic Depository Library Documents (BDLD) [URL: <http://www.du.edu/~ttyler/bdlhome.htm>] - are in this category. As such they are useful for database creation by only the patient and knowledgeable few. When, as is often the case, the files are alternatively available in another format, the data may be acquired for local database creation.

In some cases, as with the GPO Access Monthly Catalog (which is now actually something more and something less than the print or CD-ROM version of the same

title) there is only an online version available for viewing. Because records in the database are retrievable one at a time and because there is no downloadable file containing the same data, use of this valuable resource by most depository libraries for catalog record production is not possible.

- **Files created for downloading.** The List of Classes, Profiles, and Depository Shipping Lists are in this category. The List of Classes is available in CSV (comma separated variables) format, which is very versatile and permits data to be easily imported into spreadsheets, flat file managers and relational databases.

The dBase format used for **Profiles** (the directory of depository libraries) and **Depository Shipping Lists**, is perhaps less forgiving than CSV, but is nevertheless an effective way to allow for data interchange.

- **Neither/Nor files.** Some files we encounter seem to be meant neither for viewing nor downloading. The FDLP Administration Publications versions of the **Superseded List** and **List of Classes** which are served as Web pages fall in this category. Until November 1996, the **Administrative Notes - Technical Supplement** lists for corrections, changes and additions to the **List of Classes** fell in this category. Now this publication fits nicely in category one: files for viewing, but unfortunately there is no good alternative for downloading.

What is needed is uniformity, standardization and a wider appreciation of the potential use of depository-related data files. The comma delimited format is perhaps the most versatile because there are no parameters beyond number of fields per record that have to be considered in the downloading process. dBase formatted records are probably the next most useful format, but only if there is standardization and consistency in the number of fields, their names, lengths, and properties.

GPO - Rebuilding the Depository Information Infrastructure

Item Lister:

This valuable Web page was introduced by GPO/LPS in December 1996. It permits depository libraries to download their current list of item selections. Changes made to the original offering and formally released in February now permit libraries to select output modes that are appropriate to immediate needs, whether for viewing or for downloading.

List of Classes:

44 USC 1904, (Classified list of Government publications for selection by depositories) is the statutory basis for the List of Classes:

The Superintendent of Documents shall currently issue a classified list of Government publications in suitable form, containing annotations of contents and listed by item identification numbers to facilitate the selection of only those publications needed by depository libraries.

The List of Classes is a very changeable database. My experience is that annually, between 10 and 20 per cent of the records are changed in some way each year.

With the release of the April electronic version, GPO has completed one year of regular, timely publication of this important tool in a usable electronic format.

LPS staff efforts to clean up the List of Classes are certainly appreciated. As much of the useful data in this database is hidden or embedded within the free-form "title" area (e.g., frequency, format, report series designation, notes, etc.), it would be helpful if LPS would separate these data elements out so they can be used in an automated setting.

Closer correlation of List of Classes data with other resources of the depository program might also be considered. Some examples:

- Title equivalency when appropriate with cataloged titles;
- Separate data elements for OCLC and other standard numbers; and
- Provision for URLs (Uniform Resource Locators).

While the law requires that the List of Classes indicate currently "active" items that may be selected by depositories, it would be helpful if the database included inactive items also. This information is generally required for some aspects of local record processing and is necessary for database control at the local level. For many depositories, keeping track of additions and changes to the List of Classes database is considered next to impossible because so many print and electronic sources have to be consulted. A single authoritative source would certainly be welcome.

The "unauthorized" BDL D List of Classes - Additions & Changes attempts to monitor the more common resources where change information is to be found. While I have not been able to include Shipping List notice of changes as frequently as I would like, what I've noticed is indeed interesting. **No single source can be relied on to reflect all changes.**

Administrative Notes - Technical Supplement:

The most important sections of **Administrative Notes - Technical Supplement** for local database control are:

1. Classification/Cataloging Update
2. Update to the List of Classes: Miscellaneous;
3. Update to the List of Classes: New Item Numbers; and
4. Update to the Superseded List.

While these files have been available in electronic form for several years, they require significant massaging to make them usable in a database environment.

A CSV (comma separated variables) format would be appropriate (and useful) for these files.

Inactive List:

This important publication needs to be available to depositories. Currently the BDL version is the only net-available Inactive List that I know of. Local depository processing database applications need this data to work with older cataloging records and with recently deactivated item numbers that continue to be distributed or cataloged due to backlogs or format reproduction.

Superseded List:

This file is available in several formats and the text version can be readily converted to database use.

Electronic Shipping Lists:

Electronic versions of Depository Shipping Lists are now available from three Web sites in as many formats.

- GPO Access serves shipping lists in dBase format from the Federal Bulletin Board.
- BDL provides HTML format shipping lists suitable for printing and useful for in-house processing.
- The University of Texas-Arlington serves Shipping Lists in an a-la-carte (or user selectable) format.
- Commercial vendors (e.g., Marcive and Bernan) also provide this data via FTP or diskette.

While the commercial shipping list products at best reproduce only the errors on the paper versions, the new electronic Shipping Lists probably contain many more errors that are beyond the usual transposed numbers or mistakes in transcription.

Unlike the paper counterpart, electronic shipping lists illustrate the following rule: **More errors can be made by fewer people in less time.** Many of these errors seem to be related to the production environment.

- Date information that should appear as 6 digits in the format "yymmdd" sometimes appears in the currency format expressed to two decimal points;
- Titles are sometimes truncated due to faulty column width settings in the production software, and
- Item number data, which should be formatted as text, is formatted as numeric and thus leading zeros are stripped off.

Other difficulties in using the electronic shipping list data in a database environment are related to administrative decisions about which data fields are included and how they have been specified:

- Uppercase/truncated title data appears for microfiche shipping lists, and
- At least initially, during the experimental stage, the lack of shipping list date.

Like the paper lists, the electronic versions would benefit from standardization of entry:

- Information for serials should emphasize in a standard way the enumeration and chronology of the material described.
- Information for monographs, whether in series or not, should emphasize the unique title. Most of us who work with depository material recognize that AE 2.110:104-188 indicates slip law 188 of the 104th Congress. The title "Public Law No. 104-188" would not be all that helpful if loaded into an online catalog.

Monthly Catalog:

No substitute can be found for the cards in library records; and the problem of actual use of the books in the libraries will never be solved until cards are supplied in sufficient numbers to provide for the proper entries in the various library catalogues fully covering every publication issued to them. -- William L. Post, Superintendent of Documents (Annual Report of the Public Printer for the Fiscal Year Ended June 30 1907, Washington, GPO, 1907, p. 340)

In the same report, the Superintendent of Documents quotes Richard R. Bowker, editor of the Library Journal:

[re: analytic card sets for documents] I don't think there could be a greater saving to the libraries than a system of that sort.

Depository libraries should be able to acquire and use GPO's current cataloging production in their local OPACs.

44 USC 1711, (Distribution and Sale of Public Documents) requires the Superintendent of Documents to "prepare a catalog of Government publications" and to print the catalog in pamphlet form. Shouldn't the database from which the current Monthly Catalog (indeed in pamphlet form) is printed be available to depository libraries in machine readable form?

As we have been reminded in recent GOVDOC-L communications, the Monthly Catalog database is an orphan. Created by GPO, it is sold by Library of Congress, OCLC, Marcive, Autographics and others, yet it is not maintained as a database by any single agency.

In fact, even with a super abundance of sellers and suppliers, it is probably impossible to acquire from one source, in electronic form, all the records that correspond to their equivalent in the printed Monthly Catalog.

The depository library community needs a single, authorized, maintained database. Depository libraries should have free, timely, and efficient access to the records in this database.

For many years libraries that have used GPO-MARC records in their online catalogs have had to work around problems caused by the printing requirements for the Monthly Catalog: for example, the repetitive cataloging of semiannual serials and the annual reissue of unchanged records for "periodicals." Now that the printed Monthly Catalog has been reduced to an unimportant shadow of its former self, shouldn't these disruptive and costly practices be eliminated?

The database should represent record creation and maintenance, not the printed Monthly Catalog. Ready access to this database by depositories is one of the major obstacles to permitting libraries to move towards a "more electronic" processing environment.

With current technology, current GPO cataloging in USMARC format could be made available to depositories on a daily or weekly basis via FTP.

Implementing Database Controls in LPS

The most irritating classification and item number corrections encountered by depository libraries are those that are merely typographical errors made during the shipping list creation process. Not only do these mistakes require hundreds of man/woman hours to correct in libraries across the country, they play havoc with automated systems.

Were LPS to create (or at least verify) shipping lists in a database environment, such errors could easily be eliminated at the source. Implementation of such a scheme by LPS might also provide for:

- Controlled processing for adding or creating new Item/Class categories;
- Automatic updating of announcement tools for additions, deletions and changes (e.g.; Administrative Notes - Technical Supplement); and
- Simultaneous updates to electronic files that make up the List of Classes, the Inactive List, Superseded List, and the Item List.

Role of the Depository Library Community

It is in the depository library community where the challenge (and potential) is greatest. Can we find the interest and skills among our collective personnel and resources to develop sharable tools to deal with the data that is (and can be made) readily available?

We will need libraries to collect and distribute the "value added" elements of relational database technology: macros, query design, specialized functions or programmed subroutines, table design, API's for different library systems, etc.

We will need a new vocabulary for defining and describing our data and activities.

Conclusion

In the past year the depository library community has experienced significant and increased access to basic bibliographic and control data resources that have the potential to greatly simplify local depository processing activities. With the exception of the Monthly Catalog, there are database tools in place that now permit depositories to extend processing to creation of records for local online catalogs (OPACS). GPO/LPS efforts are to be commended. If the few remaining gaps in available data and/or formats are addressed in the coming year, **Just in Time** will be a possibility for all depository libraries.

URL:

<http://www.du.edu/~ttyler/dlc97.htm>