

Proceedings of the 7th Annual Federal Depository Library Conference

April 20 - 23, 1998

Documents Data Miner: A Resource for Collection Development and Management

Cathy Hartman, University of North Texas
Denton, TX

Introduction

Let me begin by saying "thank you" to Nan Myers and John Ellis for making this resource available to the depository library community. A few months ago, Nan contacted me to ask if I would work with her to test and evaluate the Documents Data Miner as a collection development resource and then present practical uses of the Data Miner as a member of this panel. After one viewing of the site, I immediately agreed to join her team. From my first searches of the databases, it was readily apparent that the Data Miner would be an exceptional tool for many aspects of our collection management and development work.

For several weeks, Nan and I had weekly, long phone conversations during which she very patiently walked me through all the features of the site. With this basic knowledge, I spent the next few weeks using the site.

Nan and John have given you an overview of the Documents Data Miner, including the features, search techniques, and the technical aspects of the relational databases. To illustrate practical applications of the Documents Data Miner, I selected a few of my favorite uses to demonstrate for you, including uses for your library, for cooperative collection development in your city, state, or region, and for other miscellaneous uses. Then, I will close with a few suggestions for enhancement of this wonderful resource.

Collection Development and Management

At the University of North Texas (UNT) depository library, we serve a population of over 25,000 students in addition to the Dallas/Fort Worth region. We are classified as a large depository library with depository holdings of more than one million volumes in the documents collection, and this fall we will be celebrating our 50th anniversary as a depository library. Within this context, there are many ways we can use the Data Miner to assist with collection management and development.

Collection Management and Development within Library

A long-standing goal in our department is to better understand the usage patterns of our collection. We recently implemented a new integrated library system that allows us to build circulation reports by call number ranges. We simply supply our head of circulation with a range of SuDocs stems for the reports. Reports organized by SuDocs stem, i.e., by publishing agency, facilitate using the data for stack management and item selection. Also, if your library scans barcodes of items used in-house, the usage data will be more comprehensive, covering non-circulating titles, as well. Usage statistics will assist our decision making in two areas:

1. With limited stack space, parts of our collection must be sent to a temperature and humidity controlled remote storage unit. We, of course, prefer to send seldom-used materials. Usage statistics facilitate informed decision-making about relocation of materials. Also, selective depositories need solid usage data for targeting of specific areas for weeding.
2. With all collections, accurate usage statistics are invaluable in the annual item selection review process. Documents Data Miner allows us to build the circulation report lists of currently received SuDocs stems quickly and easily:
 - a. From the main page of Documents Data Miner, select "Depository Selection and Directory."
 - b. Type in depository number. UNT's number is 0608A. Submit.
 - c. Data Miner found the specified depository. To see UNT's item selections, select depository number 0608A.
 - d. The resulting screen allows for searching items that UNT selects by agency, by SuDocs stem, and by item number. To search by agency, select the down arrow key on the text box.
 - e. Displayed is the number of items selected by UNT from each agency.
 - f. Select, for example, the Government Printing Office, and submit or double click.
 - g. Displayed is a list of the 29 items UNT selects from GPO. This data can be used and manipulated in so many ways because it can be saved and opened in spreadsheet software. Simply open the file menu, and
 - h. Select "Save frame as."
 - i. Open your spreadsheet software (example uses Excel) and click the "open file" button.
 - j. The application asks that a file be selected to open. Be certain to change the type of file to "all files" since files from the Data Miner will save with a .asp extension. Find your file and open it.
 - k. The headings and the table are saved, in color. Because this data is now in spreadsheet format, it can easily be manipulated in many ways, such as deleting columns, merging tables, or adding a column for local notes. To build our list for usage statistics, we need a simple text list of call number stems.
 - l. Simply block the column with SuDocs stems, then select "edit/copy" or use the key strokes, <ctrl> C.
 - m. Open word processing software, then select "edit/paste" or use the key strokes, <ctrl> V.

- n. The column of the table has been copied. Using the table pull-down menu option, "convert table to text," a simple text list of SuDocs numbers remains, ready to forward to the head of circulation for our report.

Cooperative Collection Development: City/State/Region

The North Central Texas area has several Federal depository libraries, and to support cooperative efforts, we formed the North Texas Documents Group. The group includes librarians from about 15 depository libraries that meet twice each year to discuss issues of interest to all. Represented are a variety of sizes of libraries, both academic and public. Cooperative collection development has long been discussed informally in the group, however we found it difficult to coordinate a plan. In 1991, a group of 4 or 5 libraries using one of the first electronic List of Classes, commissioned a printout with each library's selections tagged. Other libraries, experiencing a cycle of lean budgets, could not afford to participate, and of course, the printout was quickly out of date with the next item selection cycle. The Documents Data Miner solves all of our technical problems by allowing each library easy, free access to the up-to-date selections of every library in the group. To monitor the selections of the other libraries in the group when making retention or selection decisions is simple.

- a. Using the Documents Data Miner, Select "Union List Profile" in top frame. I find the frame at the top very useful making switching from one feature to another fast and easy.
- b. Type "100" (or any number of miles) in the text box indicating desired radius in miles around your library and your depository library number (0608A) in the depository number text box and submit.
- c. Select "List of Classes" from the frame at top.
- d. Select down arrow key by the "Agency" text box and highlight an agency name. For this example, I will select EPA.
- e. Now, select the down arrow key by the "Formats" text box and highlight a format. As Electronic Resources Coordinator for the UNT Libraries, I am especially interested in seeing what electronic products other libraries in my area have chosen, so I will highlight CD-ROM. Submit.
- f. Displayed is a list of all CD-ROMs offered by EPA. Select the item number of any CD-ROM,
- g. to see libraries within 100 miles that select it.
- h. Should this EPA CD-ROM be an item that we are considering for deselecting, I may want to query the libraries on this list to see if someone else will be retaining it. If the list includes only one or two libraries, I can use e-mail links on the table, but if the list is longer, I may prefer to build a list of e-mail addresses using the same technique demonstrated earlier. I can easily save/block/copy a list of e-mail addresses to paste into my e-mail address box to query several librarians with one message.

Now that we have all of our technical impediments removed by Documents Data Miner, the North Texas Documents Group no longer has a reason (excuse?) to delay formulating a cooperative collection development plan. Because Documents Data Miner allows us to monitor selections by agency and by media, it will facilitate assignment of specific agencies for in-depth collection and retention.

Also, we frequently make referrals to other libraries in the group. Another benefit of the Data Miner is the enabling of easy referral of patrons to the appropriate library in the area.

Miscellaneous uses

In the few months since Nan asked me to review the Data Miner site, several miscellaneous tasks offered the opportunity to use the databases. I selected two examples to illustrate tasks made easier by Data Miner.

First, as the Content Partner with GPO to provide permanent public access to the electronic records of the Advisory Commission on InterGovernmental Relations (ACIR), we receive requests from other states and other countries for ACIR documents currently available only in paper. One morning, I received a phone call from an elderly gentleman. He was calling from a distant state and wanted to know if I had received his fax. He told me that he had called several places trying to find a series of ACIR documents and someone in North Carolina had referred him to us. His fax was sent to a fax machine somewhere on campus (the unknown recipient is probably still wondering what it is), so he resent it to the correct fax number. The fax included clippings from an "unorthodox" publication with an amazing amount of misinformation about the ACIR.

With considerable effort, we found the documents he wanted. Then, I used Documents Data Miner to locate a depository library in his area (there was one in his city), and to e-mail the documents librarian. I sent the list of call numbers and titles for the desired documents. The librarian immediately responded that they did not have them, but would borrow from their regional and that I should refer the gentleman to them. I faxed the information to him and he called a few days later, called me "little lady," and told me how often librarians had helped him find information and what a great group we are. The Documents Data Miner facilitated the exchange between libraries. It was a quick, easy, and effective means of finding and communicating with another depository library.

Second, before the recent conference of the Texas Library Association, I used the Data Miner to build a list of e-mail addresses of all the depository libraries in Texas. I used the list to quickly let everyone know about a meeting that was added too late to be in the conference program. With more than 50 Federal depository libraries in Texas, typing in each e-mail address or sending each a separate message was an unacceptable alternative.

Enhancements

Nan and John have devoted so much time to this project already that "ideas for other enhancements" may meet with the response of, "Okay, but only if you are willing to help." I do have two ideas for enhancement, and I am willing to help with at least one of them.

The first idea for enhancement is for the feature, Depository Selection and Directory. When the depository library number has been entered and the number selected on the next screen, the search screen appears.

A search by media/format in this section would assist with planning for space in MF and CD-ROM cabinets. By selecting an agency and a format, a library could view all items

selected from that agency in a specific format and with the frequency notes included in the table, make projections for purchasing cabinets or other storage equipment. Trends for moving to electronic publishing could be observed for an agency as the percentage of electronic items grows and other media decrease.

Second, we suggest enhancement of the Inactive/Discontinued feature of the Documents Data Miner. The Inactive/ Discontinued database currently contains items from 1950 to current, but these items cannot be matched to individual library selections from that time period. By including previous selections for libraries for Inactive/ Discontinued items, this feature has immense potential to assist with collection management, referrals, and cooperative collection retention projects. If inactive/discontinued items from 1942 (in November of that year, Item Numbers were first used) to 1950 are added to the database to complete the coverage, then depositories that wish to participate could add their own retrospective item selections. Nan and John would be needed to coordinate with libraries that wish to participate and to outline the technical requirements of the data that we would provide.

This enlarged inactive/discontinued database would, in my opinion, be a major national resource. Collection management for the older sections of our collections would be added to the list of Data Miner uses. Also, because most of us are batch-loading records in our catalogs for depository library titles, our library symbols are not set in OCLC. An enhanced inactive/discontinued database would clearly facilitate the process of referrals and interlibrary loans.

In closing, again, thank you to Nan, John, and others at Wichita State University involved in this project that began as a local collection development tool and developed into the Documents Data Miner.