

Documents Data Miner 2 Searching Strategies

Nan Myers
Wichita State University

DDM 2

<http://govdoc.wichita.edu/ddm2>

Documents Data Miner 2 is a library management system for U.S. government documents. A web-based data warehousing and data mining tool, DDM2 assists depository libraries in processing, cataloging, and bibliographic control of federal documents. In addition, DDM2 contains a pilot module for a public access Catalog which can be set to a depository's name and profile.

DDM2 is based on the original Documents Data Miner <<http://govdoc.wichita.edu/ddm>>, announced in 1998 as a partnership between the Government Printing Office and Wichita State University Libraries. DDM2, announced in the Fall of 2001 as a pilot project, is a collaboration between the Wichita State University Libraries and Computing Center. The development team from University Libraries are Nan Myers, Associate Professor and Government Documents Librarian, and John Williams, Head of Acquisitions. John Ellis, Manager of Internet Applications for the University Computing Center, is the programmer for both DDM and DDM2.

FEATURES OF DDM2

LOGIN/LOGOFF

Login is required to use DDM2. From the menu on the left, click on Login to set up a new account. Login then requires your e-mail address and password. If you forget your password, go to Login, input your e-mail address and click on "Get Password." Your password will be e-mailed to you electronically. Use of Logoff takes down your browser, and is not required.

HOME PAGE NAVIGATION

Modules of DDM2 may be selected from either the menu at the left or the contents area in the center. Once users are into a module, such as the SHIPPING LISTS, they should navigate from the Banner at the top of the page or use the "Back" key in the browser. The Banner offers the options to go HOME, to TOOLS, to DDM, to FDLP, to read the Introduction or Tutorial, or to e-mail us from Feedback. The footer on the home page offers most of the same choices.

TOOLS

- **Session Configuration:** Clicking on "Session Configuration" provides the union list feature of DDM2. Fill in a depository number, then scroll down to select a state, region or distance from that depository. When working within the depository's profile search screen, DDM2 will automatically return requested union list data. In DDM, establishing a session automatically provides a shortcut to that depository's profile on the frame. This is not available in DDM2. (See also: Union List Feature)
- **Exports and Downloads:** Users can build their own inhouse databases with files from the TOOLS page. The comma delimited Ascii files from the FBB are available in Ascii text or Excel. If the user sets up a "Session Configuration," then DDM2 will recognize the depository number and provide customized files available for export into Excel.

Available files (files are dated):

- Your Depository's Selected Item List
- Your Depository's Non-Selected Item List
- Your Selects and Non-Selects together
- GPO's Current List of Classes
- GPO's Current Inactive/Discontinued List
- GPO's Depository Directory (profiles)

Difference between "export" and "download":

Export – will transfer the selected useable tabular data in Ascii text directly to your hard drive.

Download – will import the same data in a zipped file into your computer. You select the point of import (Excel, etc.). Downloads are much faster.

- **Export last query into Excel:** This feature is available from the TOOLS page, but is also available at the bottom of every query return from the various modules of DDM2.
- **Set records per page:** The default on records per page is 25. Set to a higher number to allow scrolling through the query return data.
- **Communications:** "E-mail us" and "Change Password" are on the Tools page. Users may also contact us from "Feedback" on the Banner and "Contact Us" on the footer.
- **Reports:** "Additions to and deletions from the List of Classes since 6-1-2001" is available for download.
- **Agency/Sub-agency List:** Available in Internet Explorer only. Provides entrée to the sub-agencies by class. Clicking on class takes the user to the complete List of Classes entries for that sub-agency.

DATA FILES AND DATE OF LATEST UPDATE TO THE FILES

Only official GPO data from the Federal Bulletin Board files is used in DDM2. DDM2 files are updated as soon as possible after the GPO files are updated at the FBB. Generally this is the first Friday of each month for the List of Classes (listclas), Inactive or Discontinued List (inactlst], Library Directory [profiles], and Item Lister's Profiles [unionl]. Shipping lists are updated several times weekly. At present, only the Shipping List module displays the date of latest file updates to DDM2. The original DDM does reflect the latest update on the frame of its homepage.

UNION LIST FEATURE

Documents Data Miner was originally designed as a collection development tool. DDM2 retains all the union listing capabilities of DDM. By selecting TOOLS and then SESSION CONFIGURATION, a user is at the screen to enter their depository number and to filter their depository profile by state, by region, or by specific distance (radius) from themselves. The default is the state of the Home depository, so if a depository number is entered, the state is the union list feature whether selected or not.

The Union List feature is a powerful tool for collection development, whether building or down-sizing. The user can look at data in their own profile and then click on the Item Number in the display to determine which other depositories in a certain geographic area also select specific items.

PROFILE STATUS OPTIONS

Perhaps a library wants to see what they DO NOT select, or what has been dropped from their profile. This is particularly useful during the annual update cycle in June and July. By going to the STATUS feature in their depository profile screen, the user may query for five options:

- Active
- Dropped
- Unselected
- Active + Dropped
- Active + Dropped + Unselected

MODULES IN DDM2 THAT ALSO APPEAR IN DOCUMENTS DATA MINER

Four modules in DDM2 originally appeared in Documents Data Miner, which is still available. Nothing has changed in the use of these modules. These are:

- LIST OF CLASSES
- INACTIVE LIST
- DEPOSITORY SELECTION AND DIRECTORY
- TOOLS

MODULES THAT ARE UNIQUE TO DOCUMENTS DATA MINER 2

DDM2 offers six additional modules:

- SUPERSEDED LIST
- SHIPPING LISTS
- SHELF LISTS
- MARC LOCATOR
- URL LOCATOR
- CATALOG

MODULES OF DOCUMENTS DATA MINER 2

LIST OF CLASSES

From the LIST OF CLASSES, a user can:

- Search the current LIST OF CLASSES by field,
- Search the INACTIVE/DISCONTINUED LIST by field,
- Or, merge the searches by choosing "all" at the "Status" box.

The search grid offers the following options:

- **Agency:** Search "all" or use the pop-up box for a list of agencies with the sum of active item number stems for each agency.
- **Item Number:** Enter a full item number, which requires exact spacing and punctuation, or enter a partial item number.
- **SuDoc Stem:** Enter a complete or partial SuDoc stem. The search is on a string, so no truncation symbol is required. Example: Search for C, C 1, or C 1.54. Spacing and punctuation must be exact, but the ending colon is not required. Wildcard searches are possible. Example: D%23%.
- **Title:** Enter an exact title, or words from a title. Automatic left/right truncation is built in.
- **Format:** A drop-down box allows the following choices – Any Format, Paper, Microfiche, CD-ROM disks, Electronic, and Electronic Library. Electronic Library (EL) refers to a title which is available online. These are the formats supplied by GPO. We added "Unknown" since over 1,000 active records have no format data supplied.
- **Status:** Search for either active, inactive/discontinued, or all.

Results Screens (or Query Returns): The results screen (called the "Complete Class List") will state what you requested and offer a list of results. The array in the LIST OF CLASSES is always in SuDoc order.

Additional features:

- **Click on SuDoc stem** for a list of all SuDoc stems assigned to that item number. As part of our de-selection process, we have to determine what other SuDoc stems we will lose if we de-select an Item Number.
- **Click on Item Number** for the Union List feature. If you do not have Union List parameters set at this point, the DDM2 will take you to the TOOLS page to set those up.
- **Status of a record** – what is "Inactive" and what is "Undefined"? The LIST OF CLASSES and the INACTIVE/DISCONTINUED

LIST are maintained as two separate databases by the GPO. There are times when an item number falls off the LOC and yet has not been added to the INACTIVE LIST. Data Miner automatically tags these records as if they had been discontinued. But to maintain referential integrity, the display in DDM2 at the GPO Status box reflects either:

- **"Inact"** - at GPO Status: This item number appears in GPO's INACTIVE AND DISCONTINUED ITEMS text file edition from the FBB (Nov. 7, 2001).
 - **"Undef."** - at GPO Status: Is neither in the GPO's INACTIVE/ DISCONTINUED LIST nor in the LIST OF CLASSES.
- **The "And" Function:** Documents Data Miner 2 allows the user to "build" selections. If you want to narrow a search, you may. There is no limit to the number of fields you combine.

INACTIVE OR DISCONTINUED

Since this data can be searched from the LIST OF CLASSES module, it is not used as often as it was when first developed. It does offer search limited only to data about inactive or discontinued items. From this screen, you may search by Item Number, SuDoc Stem, or Title.

This module does provide some added value:

- **Inactive Date** – Available if the item was made inactive after we began loading files in DDM in October 1997.
- **Notes Field** – The annotations for the NOTES were mined from the BDL, where they were manually entered from Shipping Lists, Technical Supplements Additions & Changes, and other LPS sources. The Inactive/Discontinued List data was originally mined from the BDL prior to a data file being available at the FBB. Official GPO data has now overlaid the initial data from the BDL; however, the Notes were retained.

DEPOSITORY SELECTION AND DIRECTORY

This module combines and merges profile data with List of Classes fields, creating the Union List function. It also contains the depository directory information and e-mail functions.

Use this point of entry:

- To search any depository profile
- To obtain complete depository directory information for all depositories

- To e-mail another depository
- To click on URLs for depository homepages

The search parameters at "Depository Selection" are designed to allow varied searches:

- Enter the depository number
- Enter an institution or library name — or a partial name,
- Search for depositories in a certain city,
- Request depositories for an entire state,
- Or, search by type of library, such as Community College Libraries, Academic Law Libraries, or State Libraries. There is a pop-up table for states and types of libraries.

Search Demonstrated: Search for a depository number, such as 0204A for Wichita State University. Then, click on submit. This presents a screen that allows four functions:

1. Click on the E-Mail Address to send a message...to the depository librarian at that institution. (Since this data is supplied to the GPO by individual depositories, you may occasionally encounter an outdated e-mail or a blank box if a library has not kept the GPO informed of up-to-date information. Corrections should be sent to the GPO.)

2. Click on the Home URL to go to that depository's homepage.

3. Click on the Depository Number to search the profile. This brings up a search grid for that library's Item Lister profile of depository selections.

4. Click on the Depository Name for directory information.

Directory Information: All the fields available in the GPO database are displayed: names, addresses, phone numbers, e-mail and URL addresses, depository type, library type and size, designation code, year designated as a depository, and congressional district. We have added:

- Date of Last Update
- Selected GPO Item Count
- Selected GPO Percent
- Longitude and Latitude
- DDM Inactive Item Count*

*DDM Inactive Item Count: This is our tracking mechanism for how many inactive items we have tagged for your depository since July 1998. Items

become "inactive" either when a depository de-selects it or the GPO removes it from profiles when the item becomes inactive/discontinued.

Depository Profile Search: A profile may be searched by:

- **Agency:** Search any agency or use the pop-up box for a list of agencies with the sum of active item numbers for that depository library.
- **Item Number:** Search as you would the LIST OF CLASSES module.
- **SuDoc Stem:** Search as you would the LIST OF CLASSES module.
- **Title:** Search as you would the LIST OF CLASSES module.
- **Formats:** Search as you would the LIST OF CLASSES module.
- **Status:** See Profile Status Options on p. 3.

Shelf List Feature in the Depository Profile Search: The query return screen of item data in a profile includes a hotlink to the Shelf List module. Clicking on Shelf List provides a list of all pieces from that Item Number/SuDoc Number which appeared on shipping lists from the GPO from 1997 to present.

Shelf List records provide hotlinks to the shipping lists on which the pieces appear.

SUPERSEDED LIST

This module represents a searchable publication of the GPO – the **2002 SUPERSEDED LIST: U.S. Documents That May Be Discarded By Depository Libraries, Annotated for Retention by Regional Depositories**. It is searchable by:

- **Agency Name:** Full or partial searches are possible, as are truncated and wild card searches using %. For example, %aviation% brings up the FAA and the Aviation Medicine Office.
- **Item Number:** Full or partial searches are possible, as are truncated and wild card searches using %. For example, 00%% brings up all item numbers beginning with the double 00.
- **SuDoc Number:** The SuDoc entry requires exact spacing and punctuation. However, truncated and wild card searches are also possible.
- **Title:** Exact title or words in a title. Use % for wild card searches.

The query return provides: The agency name, SuDoc, Item Number, Title, Instructions, and Regional Note.

SHIPPING LIST SERVICES

Searchable Shipping Lists: DDM2 offers the only searchable depository shipping list utility available to the GPO and depository libraries. Shipping lists may be searched by:

- Shipping List Number
- Title
- Fiscal Year and Month
- Shipping Year and Month
- Item Number

- SuDoc Number
- Category: All or filter for Paper, Microfiche, Electronic, Separates.
- Depository Filter: This filter eliminates shipping lists with item numbers not selected by the depository.

PDF Links: Shipping lists for FY2002 and FY2003 are hotlinked to the pdf versions of the official lists from the FDLP Desktop. PDF versions are created for paper, electronic, and separates Depository Shipping Lists. No pdf versions are available for microfiche lists.

MARC Records: The Shipping List module is linked to the MARC LOCATOR module. Retrieval of an individual shipping list will attach MARC records to the Title, SuDoc and Item Number data. MARC records can be viewed and downloaded into a library's catalog or saved to a disk. Shipping lists offer:

- Individual MARC record download, or
- Bulk download of either all monograph records or all serial records affiliated with a specific shipping list.

The Shipping List module currently warehouses all shipping lists available at the Federal Bulletin Board

(7,258 GPO shipping lists as of 10-17-02)

GPO MARC RECORDS – the "MARC LOCATOR"

- Warehouses all MARC records created by GPO Cataloging Division from monthly files posted at the Federal Bulletin Board (as of 10-17-02 over 60,000 records). In addition, beginning in October 2002, the MARC LOCATOR will begin housing all GPOMARC records 1976 to present. Currently, 191,584 MARC records are available from DDM2 dating from 1991 to the present.
- Records are searchable using:
 - OCLC number
 - Item or SuDoc numbers
 - Agency (from 1xx fields)
 - Title
 - Title Key Words
 - Subject (from 6xx fields)
 - Formats
- Query return provides title, item number, SuDoc number, hotlinked PURLS, OCLC number, access to the MARC view of the OCLC record, the GPO timestamp, and the option to download the record. If the search is done on "agency," the agency name also appears.

URL LOCATOR

- A subset of the GPO MARC Record Locator.
 - Restricted to records with the 856 field for hotlinking to Web resources.
 - Warehouses 25,475 records with PURLS.
 - Searchable in the same multiple fields as the MARC Locator records.
 - Query return provides the same data as the MARC Locator records.

SHELF LISTS

- Ties the individual pieces on the shipping lists to the MARC records and offers the only existing automated shelf-listing of multi-part titles and the general publications classes of the SuDoc class system.
- Currently holds data elements for 139,503 individually shipped pieces.

This module is also linked to the Depository Profiles. Searches in a specific depository profile produce a query return screen with a column for "Shelf List." Clicking on Shelf List provides a list of all pieces from that Item Number/SuDoc Number which appeared on shipping lists from the GPO from 1997 to present. Shelf List records provide hotlinks to the shipping lists on which the pieces appear.

DDM2 CATALOG

The DDM2 Catalog is designed as a public access catalog to GPO MARC records, offering both PUBLIC and MARC (staff) views of the records. This module is still under development.

- The Catalog is designed to serve as an individual library's catalog. The Depository Number and filter for profile are added by setting up a Session Configuration from TOOLS.
- Query returns may be arrayed in three different ways by clicking on
 - Year
 - Title
 - Call Number
- From the index of the query returns, patrons may View each record. The Public View includes:
 - Title
 - Author
 - Publication
 - Description
 - Subject Headings
 - Hotlinks from PURLs
 - Call Number
 - OCLC Number
 - MARC Revision Date – Date record last updated by GPO Cataloging.

- The staff view is of the MARC record. It also includes header data:
 - OCLC number
 - Whether the record is for a monograph or serial
 - MARC Revision Date – Date record last updated by GPO Cataloging.
 - DDM Revision Date – Date loaded into DDM2.
 - The word "Leader" is hotlinked to the Leader explanation in MARC 21 Format.

- Subject headings can be cut and pasted into a box at the bottom of the record. Clicking on "Search" provides an index of all records with the same subject heading.

For additional information or feedback, please contact:

Nan Myers

Associate Professor and Government Documents Librarian, Wichita State University, Wichita KS

nan.myers@wichita.edu (316) 978-5130 1-800-572-8368

October 18, 2002