

Creating De-duplicated Macro Input Files: One Technique
(CARLI Office, September 2, 2009)

Most macro input files are based on results from MS Access queries run against a library's database looking for records that match some specific criteria. For example, the shared macro to change the OPAC suppress flag in bibliographic records might use as its input file query results that look for bibs not currently OPAC suppressed that also have a linked MFHD with a particular location code. Another example is for the shared macro that resets a patron record's expiration date; this macro might use as its input file query results that find patron records with a specified patron group code and an existing expiration date before today's date.

Staff familiar with running MS Access queries against Voyager databases may be aware that queries such as those described above can have the same underlying Voyager record in the results more than once, depending on the data scenario. Using the examples above, the bib record could have two MFHDs linked to it that represent copy 1 and copy 2, both in the same location. Similarly, patron records can have more than one patron group/barcode linked to them, and so the query results could have two rows that actually represent the same Voyager patron record.

Macro Express "quasi-batch" macros run against Voyager databases work most predictably and reliably when the input file contains only one entry for each Voyager record to be processed. There may be other techniques that work equally as well, but this document is designed to describe one method for de-duplicating input files before a macro run is started. It assumes the query results have been saved as an Excel file to the local PC, and the technique uses Excel's filtering functionality. Each of these steps is described in detail in the remainder of this document.

This document uses an example of a list of bibliographic record IDs, but the concepts apply to any list of Voyager record IDs, including patron record institution IDs or barcode numbers.

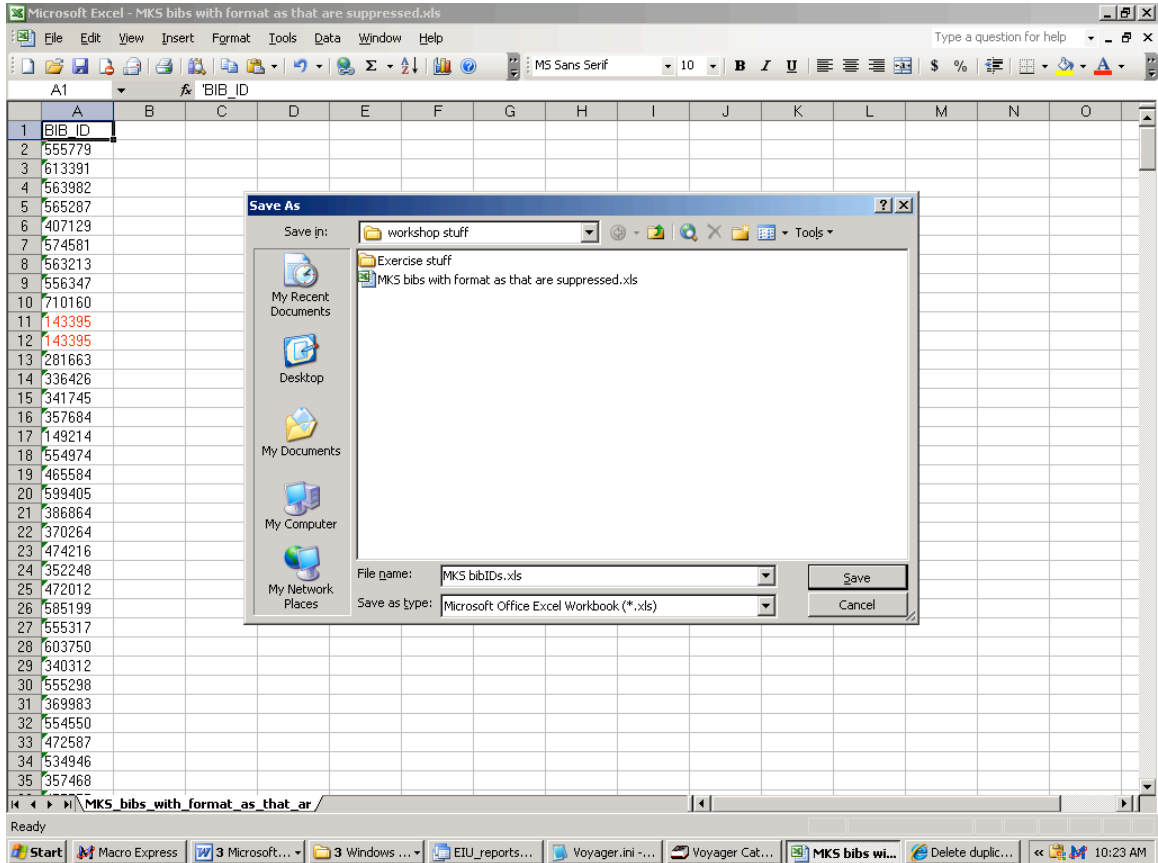
Step 1: Create a preliminary copy of record IDs to be used as the macro input file

In the screen shot below are found results of a query that looked for bib records with format coding for serials, and the bib is currently OPAC suppressed, and the location code of the linked MFHD(s) was also included in the output. Rows 11 and 12 (marked in red text for display purposes) show an instance of two entries in the file for the same bib record, because there are two MFHDs for this bib, each with a different location code.

These query results will need to be de-duplicated before they are used as an input file for the macro run that will change the OPAC suppress value in the bib records.

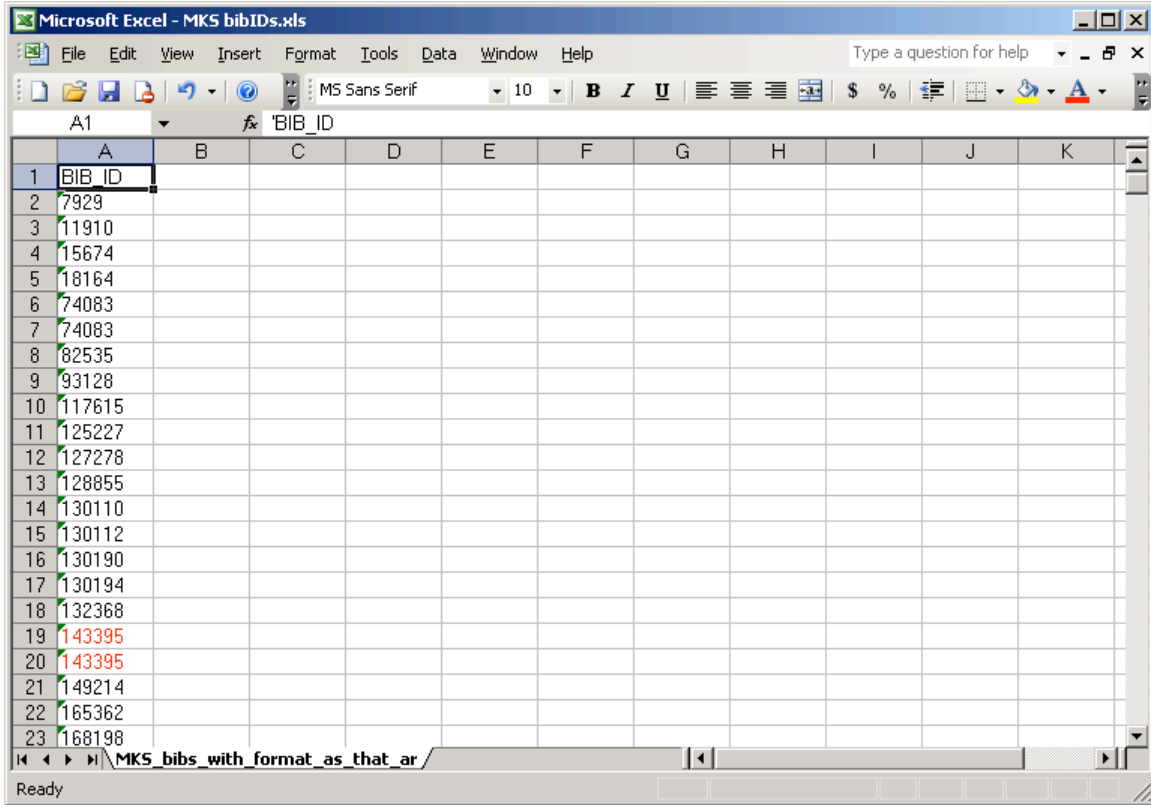
	A	B	C	D	E	F
1	BIB_ID	BIB_FORM	SUPPRES	TITLE_BRIEF	LOCATION_CODE	
2	555779	as	Y	... Annual report /	stacks	
3	613391	as	Y	... Annual report /	stacks	
4	563982	as	Y	... Annual report /	stacks	
5	565287	as	Y	... Annual report /	stacks	
6	407129	as	Y	AARP bulletin : a publication of the American Association of Retirec	per currnt	
7	574581	as	Y	AARP the magazine.	per currnt	
8	563213	as	Y	AIMS newsletter.	per currnt	
9	556347	as	Y	ASM handbook.	ref	
10	710160	as	Y	Adjunct advocate : the magazine for adjunct college educators.	per currnt	
11	143395	as	Y	Admission requirements of U.S. and Canadian dental schools.	stacks	
12	143395	as	Y	Admission requirements of U.S. and Canadian dental schools.	ref desk	
13	281663	as	Y	American arachnology : the newsletter of the American Arachnologi	per currnt	
14	336426	as	Y	Annual report - Becton, Dickinson and Company, inc.	stacks	
15	341745	as	Y	Annual report - Fleming Companies, inc.	stacks	
16	357684	as	Y	Annual report - Hon Industries.	stacks	
17	149214	as	Y	Annual report - May Department Stores Company.	stacks	
18	554974	as	Y	Annual report - Royal Appliance Mfg. Co.	stacks	
19	465584	as	Y	Annual report /	stacks	
20	599405	as	Y	Annual report /	stacks	
21	386864	as	Y	Annual report /	stacks	
22	370264	as	Y	Annual report /	stacks	
23	474216	as	Y	Annual report /	stacks	

Make a copy of this Excel file containing only the record ID column (BIB_ID in this example), and save it with a different name than the original file. This will ensure that the original file is retained as is, in case any problems with the de-duplicated file are encountered.

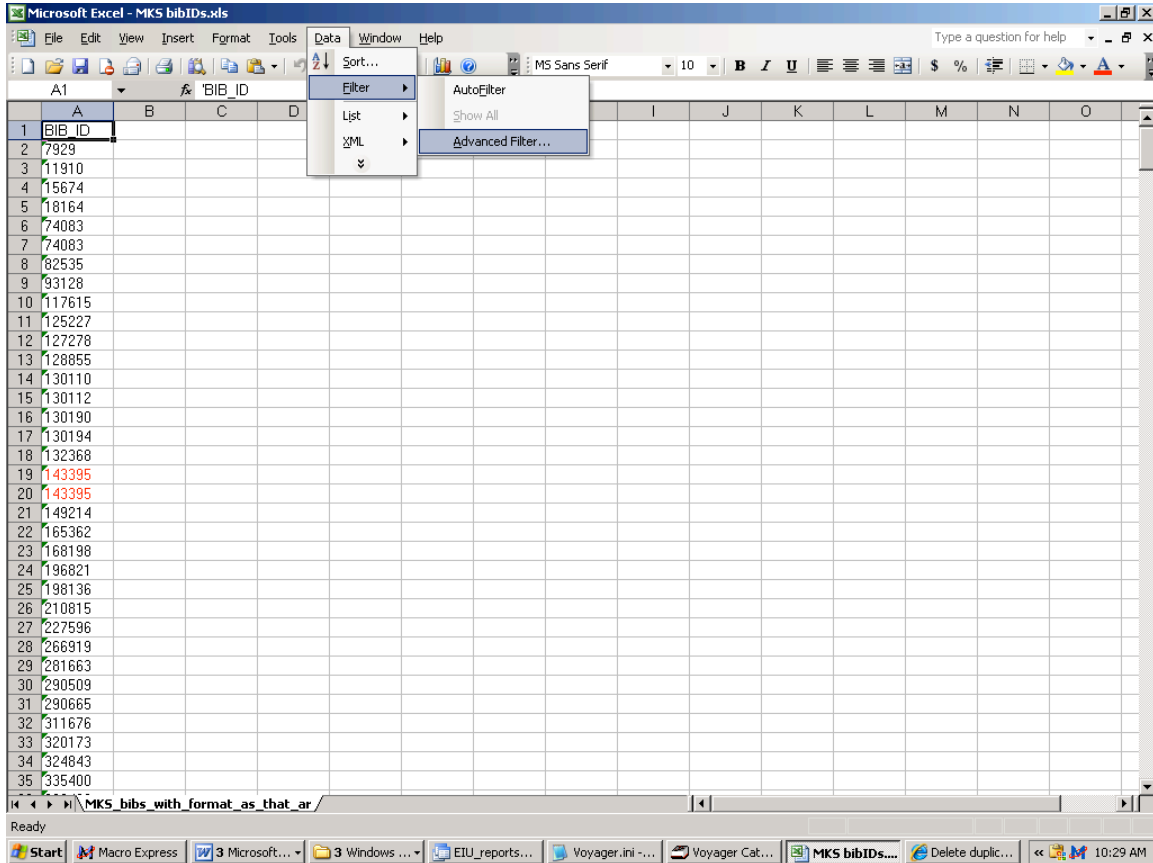


Step 2: Use Excel's Filter functionality to remove duplicated record IDs.

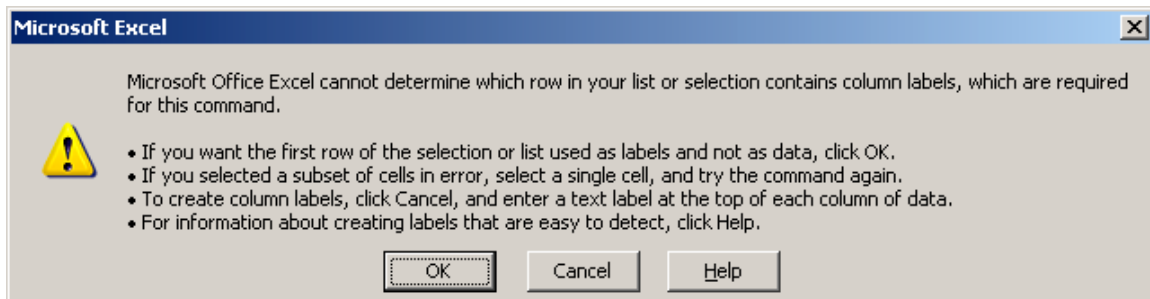
Open the new file of record IDs only. It may be useful to sort the file by the record ID number for verification purposes, as is seen in the screen shot below:



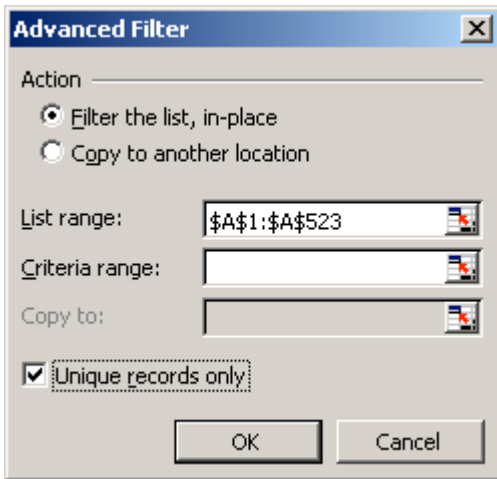
Go to the menu bar's Data → Filter → Advanced Filter option:



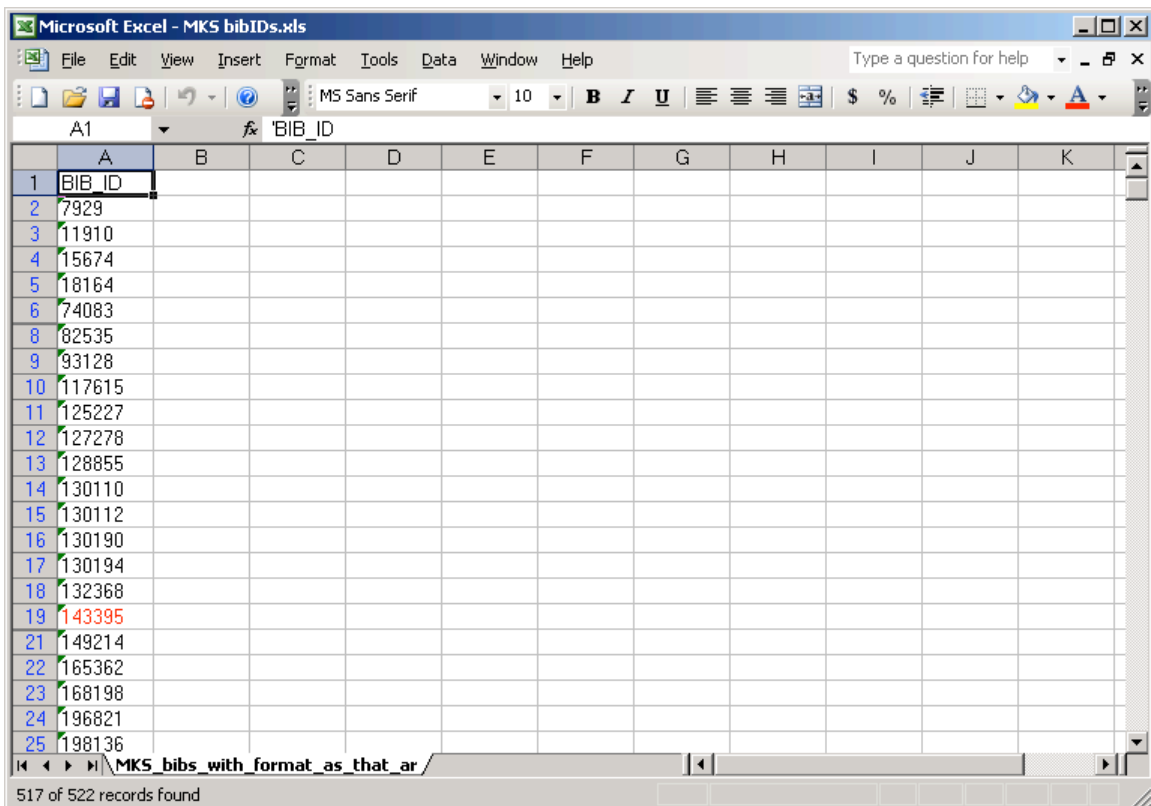
If you get a warning message about column labels, and your file contains the column header with the default Voyager field name (BIB_ID in this example), click OK to accept the Voyager text as the column header.



In the Advanced Filter dialog box, accept the defaults of “Filter the list, in place” and the List range, which by default is the entire set of records. Check the option for “Unique records only” and click OK.



The results display. NOTE: the file’s row numbers show that the duplicate entries are really just hidden from view, rather than deleted from the file (i.e., row 20 is not displaying, but it is actually still in this file).



Step 3: Create a text file containing just the de-duplicated record IDs

Shared macro input files must be a simple text file with one record ID per line, and no blank lines. In the file created in step 2, copy the entire record ID column, or a subset of record IDs, if multiple input files are needed.

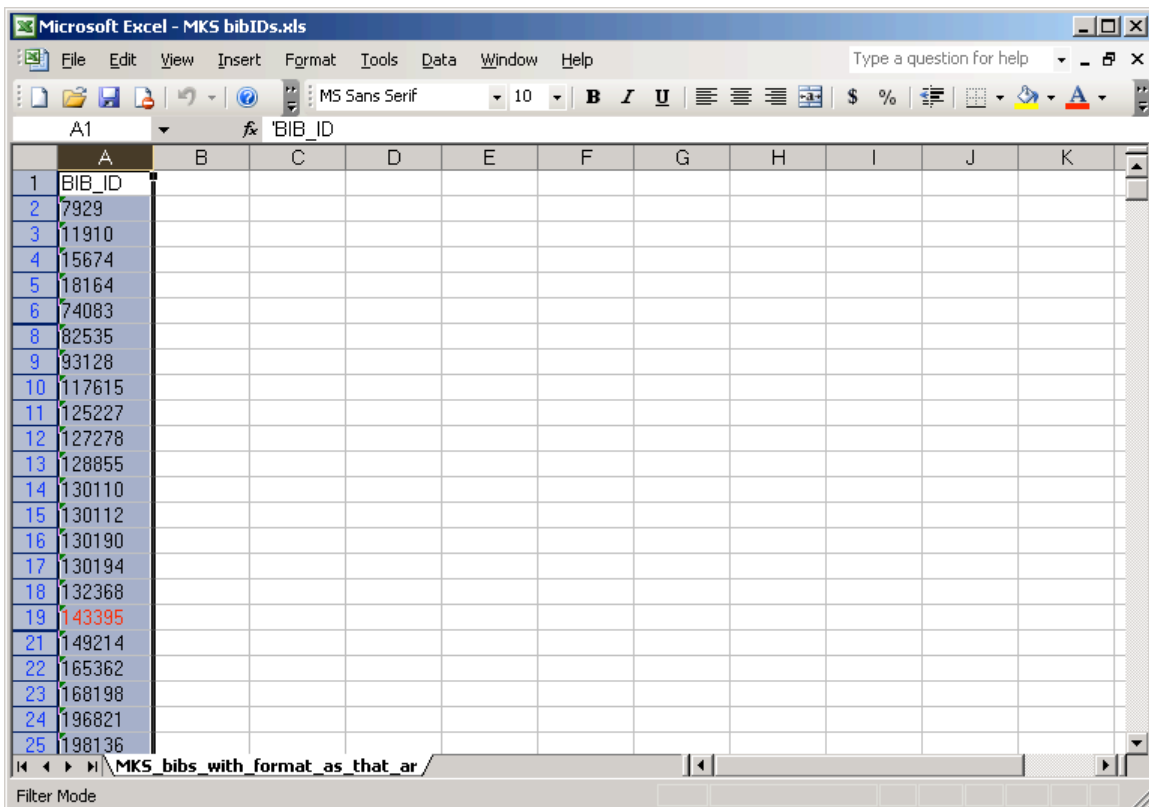
If the de-duplicated Excel file contains more record IDs than can usually be processed in a single work day, it is recommended that multiple input files be created. The rule of thumb regarding the number of records that shared macros can process in a work day is:

Bib and Holding (MFHD) records = 1500 per day / file

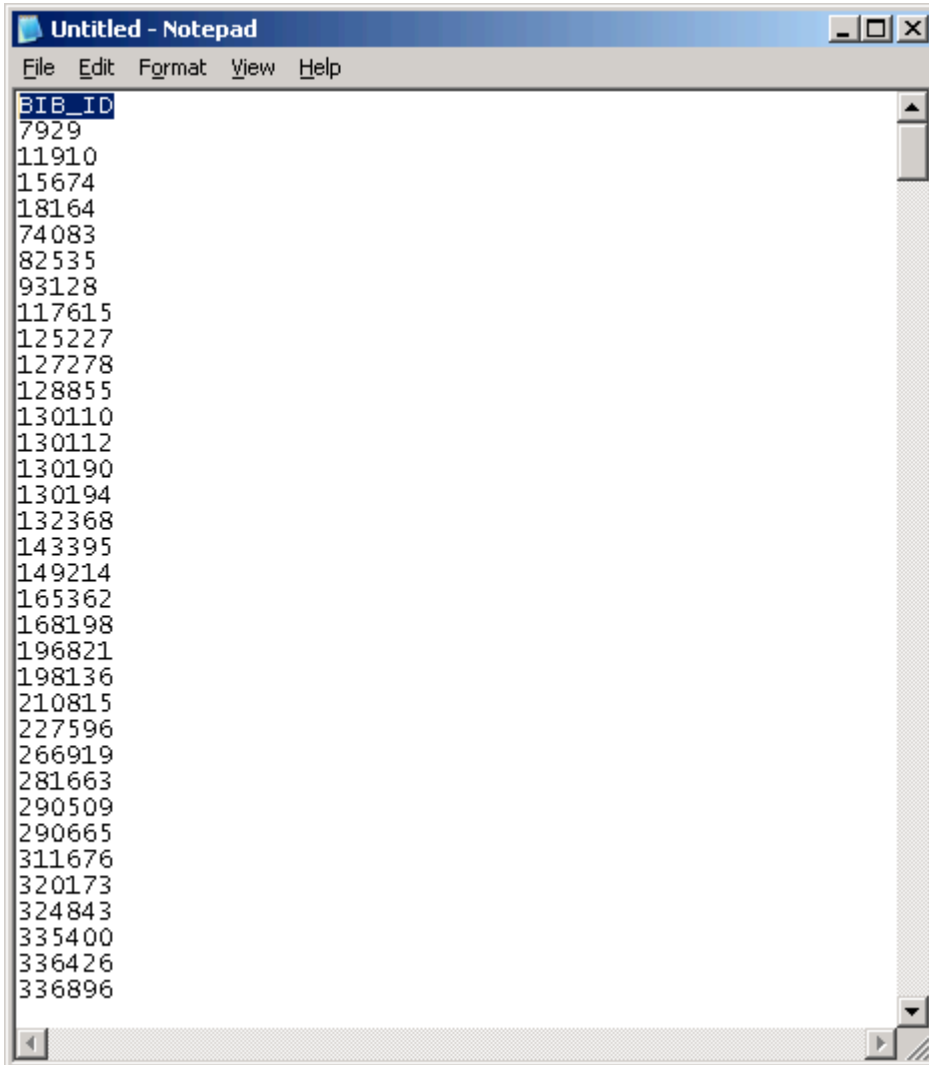
Item and Patron records = 2000 per day / file

It is also recommended that a “master copy” of each input file be created and maintained until at least the project is completed, and a working copy of each input file be used for the actual macro run. This section uses this model in the examples.

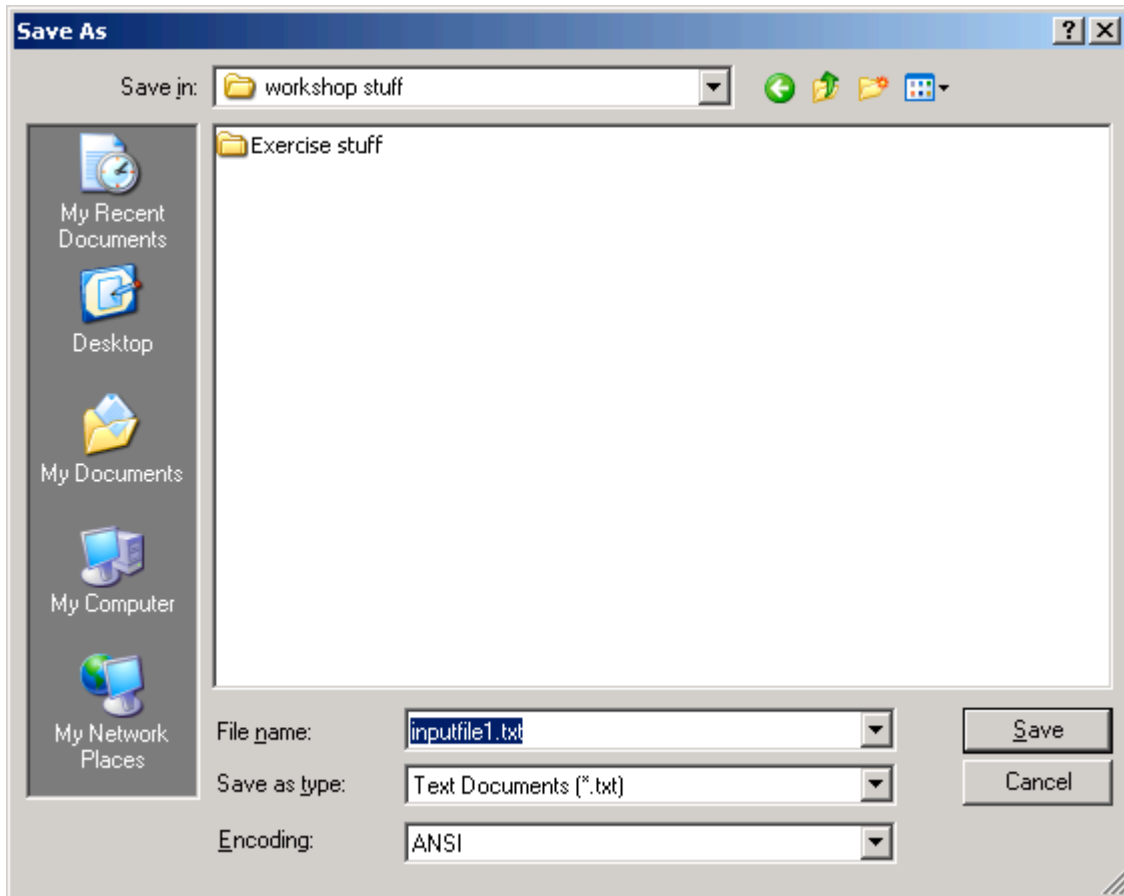
Continuing with the example Excel file above, the filtered file contains 517 records, so only a single input file is needed for this particular macro run.



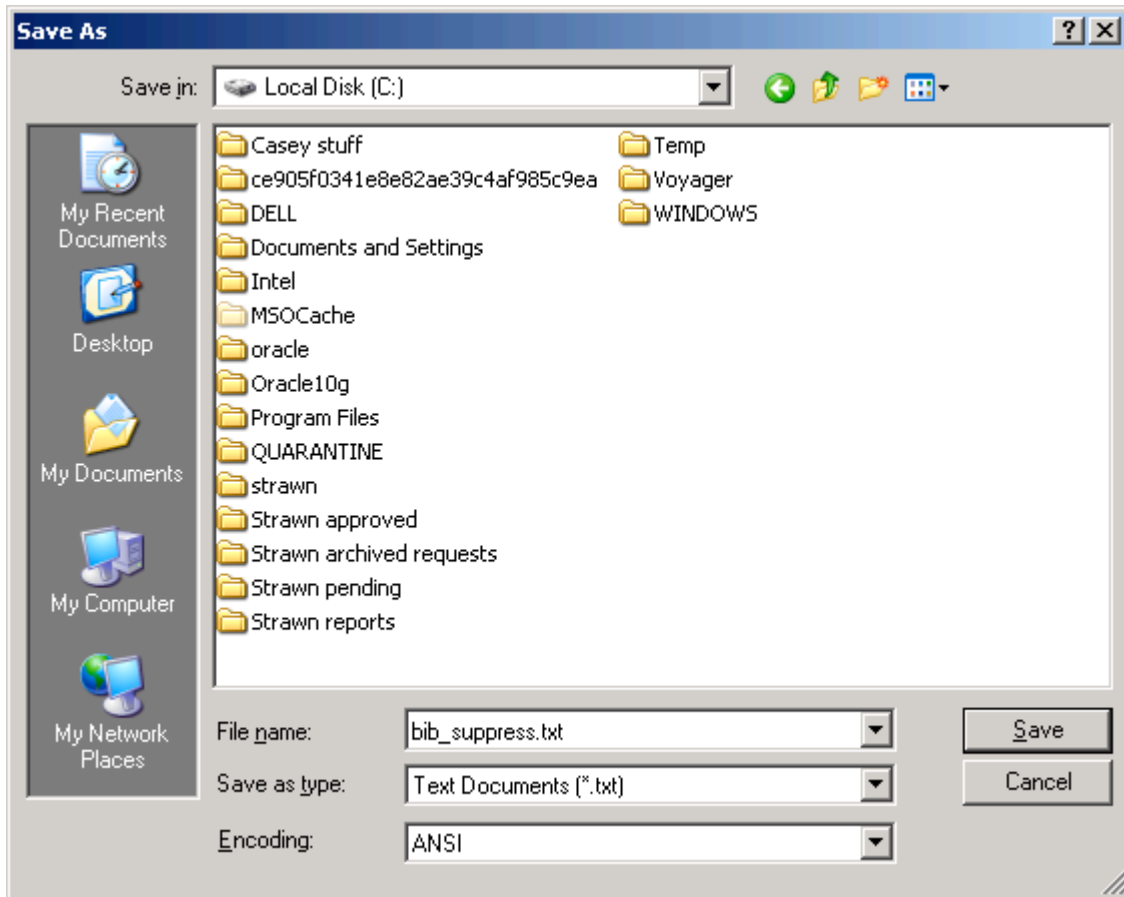
Open a new Notepad file and paste the copied record IDs. Go to the beginning of the new Notepad file and delete the column header (BIB_ID in this example):



Delete any remaining blank lines and save the text file with a distinct and descriptive name. This file will be considered the master copy of the input file.



The shared macro that changes the bib record's OPAC suppress flag expects the input file to be found on the PC's C:\ drive, with the filename of "bib_suppress.txt". Make a working copy of this master file with the required filename by using Notepad's Save As function:



You should now have a master and working copy of a de-duplicated input file for your macro run.

Getting Help

Contact the CARLI Office (support@carli.illinois.edu) if you have any questions about the content of this document or any of the shared macros on our website.