Refining Your Search: Using the CSV File

Every search of the Choir Library of St. Mary's in Lübeck (CLSM) database returns a formatted view of the pertinent data. That data is also provided as a "comma-separated values" (CSV) file. The purpose of this article is to describe what a CSV file is, and why and how it might be used.

What: A CSV file is a text file that stores tabular data in a structured format. In the CSV file provided by CLSM, each row contains all the fields related to each unique work/source/edition combination. The fields are always in the same order, left to right, and each field is separated from the next by a comma. The first row of the file contains headers, also separated by commas, with the names of the fields in the rows below. A CSV file can be opened and used in many computer applications, most notably any database or spreadsheet application.

It is important to note that the CSV file is a "flat file" representation of what may be a multi-dimensional match in the CLSM database. For instance, a given work may appear in multiple sources, and/or may have multiple editions, and/or may have multiple incipits, etc. As an example, there are only four works by the composer Agostino Agazzari in the CLSM database; but a search on that composer returns six records, because while two of those works are found in only one source, the other two are each found in two sources. In the CSV records corresponding to those works with two sources, the data in each row corresponding to the work will be the same; the data corresponding to the sources will be different. When counting items in the CSV file, it is crucial to take that duplication of data into account.

Why: A CSV file provides a way to refine a search. For example, let's say the user wishes to find works in the database for one or two voices, within a certain range of dates. The existing Works Search dialog does not allow selecting for two different values in the Voices field, nor for a range of dates in the Date field. If the user selects IS NOT NULL in both those fields, the resulting search will contain records with any non-null value in both the Voices and Date fields. The user can then load the resulting CSV file into a spreadsheet or database program and then use the tools available in the spreadsheet or database program to further select only those records where Voices = 1 or 2 and Date is within the desired range. Apart from refining a search, the exported CSV file also provides a way to save the results of a search and use those results multiple times without having to re-execute the search; and it also provides a way to move search results data to other platforms or users.

How: Start by going to the CLSM Works Search page and executing a search. For the purposes of this explanation, I have executed a search for works by the composer Samuel Capricornus. The Search Results page states that "46 records [are] found." It

also states that the user is "viewing page 1 of 2" with "25 results per page"; and it states that the records are in ascending "order by siglum." I would like the CSV to contain **all** the results for Capricornus in order by title; so, in the drop-down list showing "25 results per page" I select "All results"; and in the drop-down list showing "order by siglum," I select "order by title." The search results are immediately re-displayed according to those parameters.

Now I click on "export as .csv." A CSV file with the name "clsm-export-*yyyy-mm-dd hh_mm_ss[am/pm]*.csv" is created and placed in the computer's default download folder.

If Microsoft Excel is installed on your computer, in most cases clicking on a CSV file to open it will cause it to open in Excel. Make sure that it gets saved as an Excel file by choosing File – Save as, and saving it as an Excel file in the folder of your choice and with a meaningful name such as "Capricornus.xlsx."

Test Case: As an example, let's say that you wish to create a list of all works in the database that have modern editions and are associated with Easter. In the Works Search dialog, there are six values in the Feast drop-down menu associated with Easter, and you can only select one of them. So, you would have to do six searches. Using the CSV allows you to do one larger search and then refine it. In this example, we will assume that, on your computer, the CSV will open in Microsoft Excel. Here are the steps:

1. In CLSM, select all works that are associated with a feast and have a modern edition by selecting Feast IS NOT NULL and Edition Abbreviation IS NOT NULL. Result: 472 records found.
2. By default, the Search Results page shows only the first 25 of these records; and will output to the CSV file only the records that it shows. So, change "25 Results per Page" to "All Results" and in the Order dropdown select "Order by Composer." Result: Now there is only one Search Results page with all 472 records shown, in order by composer.
3. Select "Export as .csv". Result: A file will be created with the name clsm-export-*yyyy-mm-dd hh_mm_ss[am/pm]*.csv; on many computers, it will be placed in the download folder.
4. Click on the file to open it. The file will open in Microsoft Excel. When the file first opens, all 472 records will be shown. (If Excel wants to remove leading zeros, click "Don't Convert.") Save it as an Excel workbook with your chosen name and folder.
5. By default, all the columns will have the same width, so that much of the data is obscured. You can set the column widths to be as wide as the data they contain using these steps:

      a. Select all the data. The easiest way to do this is to press CTRL and A simultaneously. The background color of the data should change, indicating that the data has been selected.

      b. If it's not already selected, click on "Home" at the top of the screen, and then, in the ribbon above the data, click on the "Format" dropdown in the Cells division. Select "Autofit Column Width." Result: all the columns will now be wide enough that the data they contain will be fully visible.

6. Our concern is to narrow these 472 records down to those where Feast has the word "Easter" in it. Here are the steps:

      a. Each column contains data from one of the fields in the database; the name of the field is in the topmost row of the column.[1] Scroll right until you find the column with the name "sFeast."

      b. Click on "sFeast." Then, in the ribbon above, find the filter symbol, which looks like a small funnel. Click on it; then click Filter. A dropdown symbol will appear to the right of the field name "sFeast."

      c. Click on the dropdown symbol. A window will open showing all the values of sFeast. In the search box, enter Easter, then click OK.

      d. Now only the 41 rows where Feast contains "Easter" and the row contains a modern edition will be shown. Save this result. At this point you might want to print the first three columns and return to the database itself for further information about any of these works in a more friendly format.

There are many different versions of Excel; also, it is possible to open the CSV in Google sheets, in various database programs, and, on Macs, in Numbers. Given that, the steps above will be similar, but the procedure may differ. For any further help, feel free to contact the CLSM associate editor, John Sheridan, at john.s.sheridan@gmail.com.

---

[1] The field names are the native names in the database, which start with a lower-case letter or letters that indicate the table the field is in, followed by the name of the field itself, capitalized. So, the Feast field has the name sFeast, as Feast is in the "Source" table, which is indicated by the initial "s."