

Excel Archival Tool

User Guide

The Excel Archival Tool was created to programmatically convert Microsoft Excel files into open-source formats suitable for long-term archival. This is accomplished by using Visual Basic Script to convert Excel workbooks (.xls and .xlsx) into comma-separated value files (.csv), while also extracting charts/figures (in .png images), cell formulas (in .txt files), and formatting/styling information (in HTML snapshots). This tool is most useful for large Excel data sets that would otherwise require large amounts of tedious work to convert manually.

This program is available in two flavors: one with a simple graphical user interface (GUI) and one without. The guide below will primarily discuss the version equipped with a GUI. This tool requires a Windows environment to run (Windows XP, 7), and the GUI requires Internet Explorer. The non-GUI version of this tool does not require Internet Explorer.

Getting Started

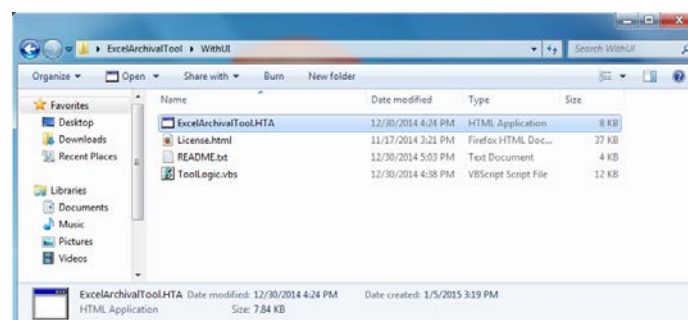
To download the tool, visit <http://z.umn.edu/ExcelTool> and click “Download ZIP” on the right side of the screen. Navigate to where you saved the download and open up the ZIP folder. Inside, you will find two folders, one named “WithUI” and the other named “WithoutUI”.

There are two files that must be inside the “WithUI” folder for the tool to run properly: ExcelArchivalTool.HTA and ToolLogic.vbs. Do not try to open the ToolLogic.vbs file; it cannot be accessed directly.

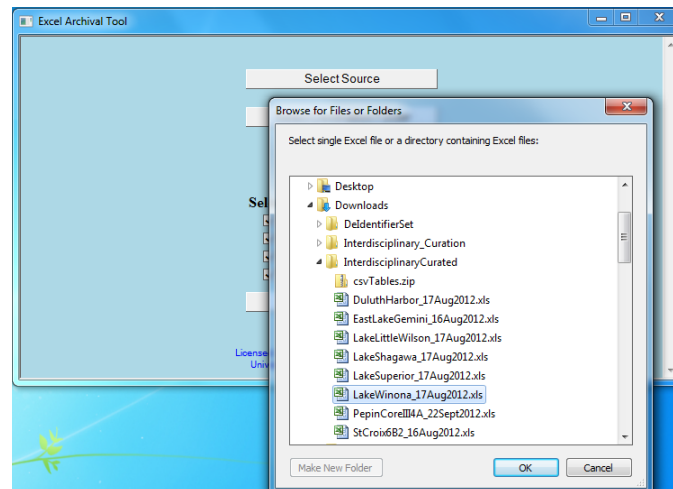
This guide will use some Excel-related vernacular to discuss different aspects of an Excel file. A workbook denotes a single .xls or .xlsx file that may contain many worksheets. For example, if you have a file named “ClinicData.xls” that has two tabs named “West Clinic Data” and “East Clinic Data”, this is considered to be one workbook with two worksheets.

Steps for use

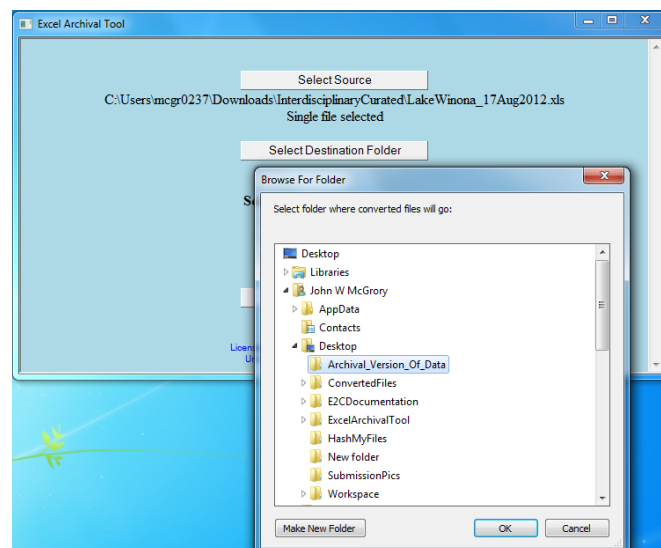
1. Close Microsoft Excel if it is currently running on your computer
2. Launch the ExcelArchivalTool.HTA file from the “WithUI” program folder



3. Select a “Source”: this can be a single Excel file or a folder containing multiple Excel files. If you select a folder, the program will recursively find and convert all Excel files within any subfolders, so be careful (e.g., if you select your Desktop, the program will search inside all of the folders on the desktop for any Excel files).



4. Select a “Destination Folder” where you would like the final conversion products to be stored. Although it isn’t required, we recommend that this folder be empty to avoid any unintended naming conflicts.



5. Using the checkboxes, select which aspects of the Excel file(s) you would like to have extracted.

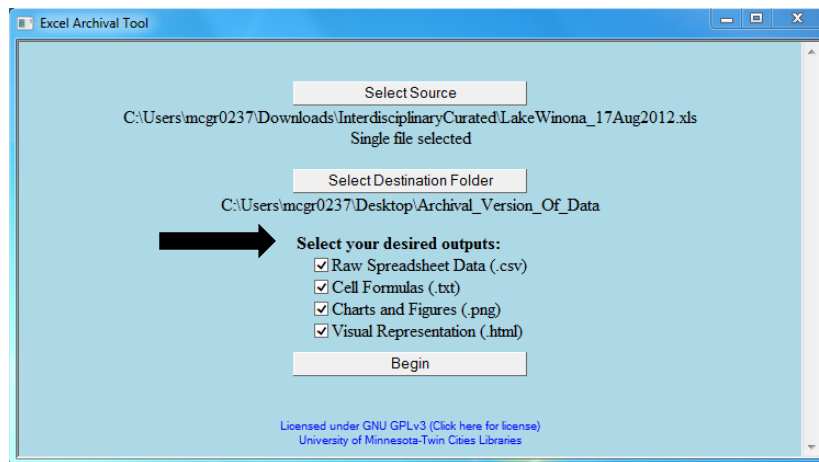
Raw spreadsheet data: This will convert all text and numerical data found in the Excel spreadsheets to a .csv file. These files will follow the naming convention,

<workbook name>_<worksheet name>.csv. If the Excel workbook contains multiple worksheets, they will each be converted as a separate .csv file.

Charts/Figures: Charts, graphs, and figures will be extracted as .png image files with the naming convention, <workbook name>/<worksheet name>_<arbitrary chart number>.png (note that this conversion step stores all of the images from a workbook within a folder named <workbook name>).

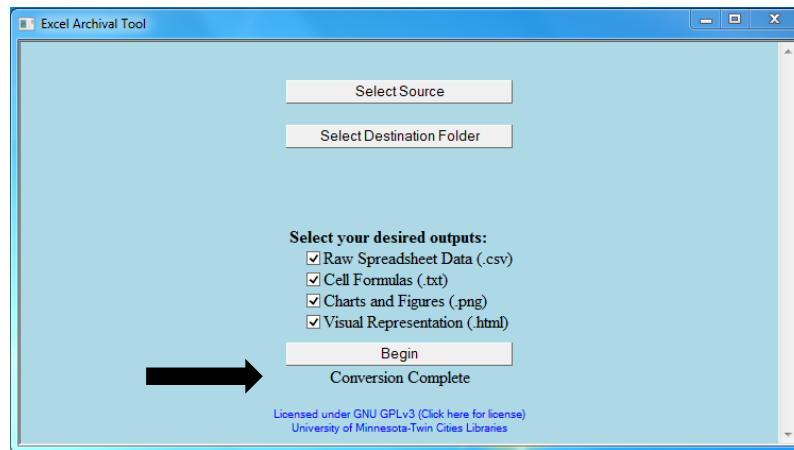
Cell Formulas: Any formulas used to derive cell values will be exported to .txt files; these files follow the naming convention, <workbook name>/<worksheet name>.txt.

Formatting/Styling: Each workbook's visual layout (e.g., cell highlighting, bolded words, etc.) is exported as an HTML snapshot. Each workbook from this conversion step is stored within a folder named <workbook name>.



*Note: Converting to an HTML snapshot also captures the graphs and figures in a workbook. So unless you need the graphs/figures in a standalone PNG file, the HTML and PNG steps will be redundant.

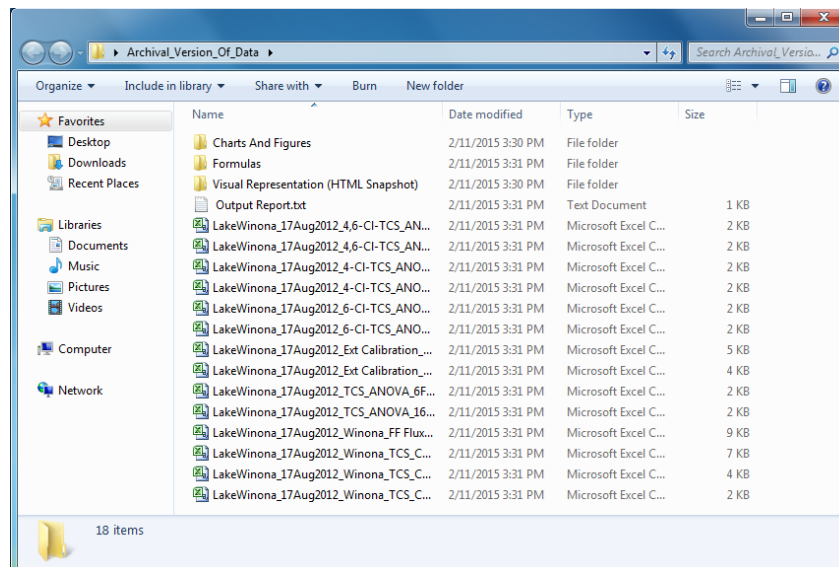
6. Click Begin. The interface will read "Processing <file/folder name>..."
7. When the conversion process is finished, your Destination Folder will open, and the interface will read "Conversion Complete".



8. Repeat steps 3-6 if you have any more files/folders to convert. When you are done, simply close the interface box.

Output

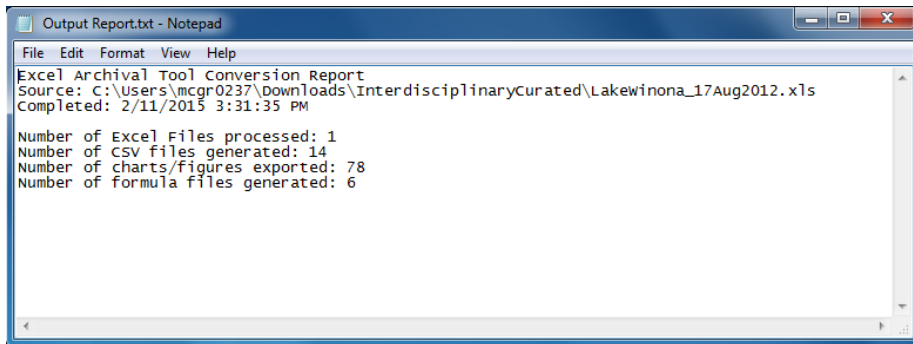
Your Destination Folder will now be populated with the output that you selected in Step 5. Any .csv files will be in the top level of this folder, while any secondary products (charts/figures, cell formulas, or HTML snapshots) will be in appropriate subfolders.



In addition to the output you selected in Step 5, the tool will also generate a report and place it in the top level of the Destination Folder. This report can serve as a record of what was converted and the final products that were created. The report contains the following:

- Name of the file/folder that was converted
- Date and time of the conversion
- Number of Excel workbooks that were processed
- Number of .csv files generated

- Number of charts/figures exported as .png image files
- Number of cell formula .txt files generated



```
Output Report.txt - Notepad
File Edit Format View Help
Excel Archival Tool Conversion Report
Source: C:\Users\mcgr0237\Downloads\InterdisciplinaryCurated\Lakewinona_17Aug2012.xls
Completed: 2/11/2015 3:31:35 PM

Number of Excel Files processed: 1
Number of CSV files generated: 14
Number of charts/figures exported: 78
Number of formula files generated: 6
```

License

This tool is released under a GNU GPLv3 license. This license allows users to use the software for any purpose, to change the software to fit your needs, to share the software freely, and to share any changes you have made to the software. A copy of the full license text is included in the Excel Archival Tool folder and can also be found at <http://www.gnu.org/licenses/gpl.html>.

Source code for this tool is available at <http://www.github.com/mcgrory/ExcelArchivalTool>.

If you have any suggestions for new features, bug fixes, or the like, please email datarepo@umn.edu.