

How to scape papers

Getting ready

You'll need:

1. Excel or other program that makes .xls or .csv files
2. ImageJ download for free from here <https://imagej.net/Welcome>. You'll also need to add the `Figure_Calibration.class`, which will help for giving x and y calibrations to images.
 - (a) To add the `Figure_Calibration.class`: In ImageJ go to plugins, select *add plug in*, then navigate to the `Figure_Calibration.class` file that you downloaded on your computer, click on it and follow through a few clicks to add the plugin.
 - (b) If you have some trouble getting the measurements to show up after calibrating, try switching to the pointer and clicking. You might need to set the preferences on your pointer tool to auto-measure.
 - i. A clarification on above from Tim Savas (original OPSREE lead data enterer): *After doing the figure calibration and selecting the yellow pointer tool, you start clicking inside the figure but no points appear. The reason for this is that the "rectangle" you drew for the figure calibration is still masking the figure, and until you click out of it, you can't draw points under it. It's hard to see! So to get rid of the invisible rectangle, just click the mouse once outside of its edge. Side note: After drawing all of your points onto a figure, you can press Command-M to bring up the resulting table of values. I do this in the video tutorial, and whenever I scrape, but didn't describe the key command!*

Now here's what to do

You might want to download and review this video. It's from some years ago, so if you find issues with it, please let us know.

1. **Needs updating ...** Copy the excel file `ospree_newpapers` for git repo (`ospree/data/ospree_2019update`) and make your own extension, for example, Dan would write "`ospree_2019update.dmb`". This will be the spread sheet you enter your data into and then in the future, someone will merge all of our files together into the master data.
2. Familiarize yourself with each tab:
 - (a) **meta_general**: metadata for each sheet
 - (b) **source**: list of the paper we are working with. Bibliographic information and notes on usefulness for our purposes. Note the "ToDo" column, which tells you which figure or table to focus on. You may find other figures are better, these were from our initial quick read. Also pay attention to `datasetID` column, which tells you how you should enter the identifying information for each paper.
 - (c) **study**: Details on each experiment within each paper; possibly only one line for a paper, if only one experiment is relevant. This sheet is useful for our overview of what kind of experimental manipulations were done.
 - (d) **data_detailed**: Detailed data for the experiment, with all relevant information filled out.
 - (e) Responses may be pre/post treatment, time, or other. Temporal responses, such as days to 50% budburst, are fairly common. An example of an other type of response would be percent budburst, again fairly common.
 - (f) **scratch**: For temporary formatting and manipulating data scraped from ImageJ.

- (g) The two most important tabs to fill out are **study** and **data_detailed**.
3. Read your paper and fill out the information in the “study” and “data_detailed tab.” Be sure your datasetID and study info agrees with the source tab, if not—figure out what is wrong and fix it.
 4. Find the designated figure or table as noted in the source tab of the worksheet. Also note the datasetID column in source; use this as you fill out data_detailed and study.
 5. Take a screen shot of the figure and import into ImageJ, following Tim’s instructions from lab meeting with a video of screen shots from [git/osprey/notes/howtosrape/Data Scraping Tutorial.mp4](https://github.com/osprey/notes/blob/master/howtosrape/Data%20Scraping%20Tutorial.mp4). Use the scratch tab to get data into the right format, and then copy into data_detailed. Fill out the study tab as best as possible to describe the experimental treatments used in each study within each publication.

A few more “how to’s”, trouble shooting etc

- **For dealing with lats and lons:** Tool to convert to decimal latitude and longitude: https://andrew.hedges.name/experiments/convert_lat_long/ and remember to add NEGATIVE to your longitude if it’s West. Also, you can check where things are by just typing in lat and long into Google maps.
- **For calculating the number of days between two dates:** Use this: <http://www.timeanddate.com/date/duration.html> to calculate the number of days between two dates for calculating dormancy. Note do NOT include the end date in the calculations.
- **On entering response times:** If the respvar of a study is “daystobudburst” enter a 1 into the “response” column, and fill in the recorded time in the “response.time” column.
- **On dealing with error:** If it’s figures records error and it is *clear enough* to scrape, error can be recorded. Often times the SE bars are in the way of each other or not quite discernible, in which case we’ve decided to avoid them. But if the bars are clear, record them. Values can go in “resp_error” and just SE in “error type.”
- **Do you have an ambient warming experiment?** For example, do you have outdoor chambers or such where they applied ambient+2, ambient+4 etc.? Please enter these as ‘ambient + [insert number here]’ ... if they give mean temperatures or such, you can note it but do NOT enter in forcetemp — forcetemp is generally reserved for when authors put things into ‘forcing conditions’ which is often growth chamber/greenhouse etc..