

Data Management Activity

CURE created by: Paula A. Trillo, Gettysburg University; Elissa Sorojsrisom, Columbia University; Carly Jordan, The George Washington University; Janice Krumm, Widener University

STANDARDIZED NAME PROTOCOLS

Keeping track of data is critical to being an effective researcher. Although it seems trivial while you are working with the data, it is common to re-visit or share datasets weeks, months, even years after they have been compiled. You will likely not remember all the pertinent details of the data at this point. However, if you use consistent, clear naming and keep track of metadata then you should have no problem coming back, or sharing your data with collaborators. This handout will show you how to manage your data.

H

information into your metadata file as well.

		the your spreadsheets the following information, letters and numbers only, no special characters or spaces.	
١.		Your institution code (your professor will provide this):	
		Your unique group name (6 characters):	
		Dataset type (occurrence, species name, etc.):	
		Current date, formatted as year-month-day (YYYY-MM-DD):	
2.	Now, put it all together, separating the parts with underscores.		
	(This \	ole: wu_blue17_PierisOccurrence_2020-03-21.xlsx would be a spreadsheet saved on March 21, 2020, by the blue17 research group at er University, and it contains occurrence data about <i>Pieris</i> specimens.)	
3.	. What would you name a spreadsheet of multimedia data that you downloaded today from iDigBio?		
4.	workin	You will save a new version of the file every time you make significant changes or finish your working session. Older versions of the file will be kept and archived . When you save a new version, update the file name to the current date. If you need to save more than one version per	

day, you can add a letter (a, b, c, etc.) to the end of the file name to distinguish between the versions. Every time you create a new version, remember to enter the updated file name

MAINTAINING A METADATA FILE

Every dataset needs to have an associated metadata file. Metadata is information that will help you (and others) understand how to read your data, how you processed and cleaned the data, and how you improved and added to the data. You will create a chronological narrative that describes every step of the process. It will help you tremendously in writing the methods section of a paper or presentation. It also helps if you set aside your research for a time, and then need to revisit the results and understand what those column titles on your spreadsheet mean.

Each collaborative group will build a metadata document for each dataset. You will be building your metadata file in GoogleDocs, so every member of your group will be able to access the file at all times. The metadata file name will be similar to your spreadsheet file names, the dataset type will be metadata.

Example: wu_blue17_PierisMetadata

Starting Your Metadata File

- 1. When you start your Google Doc metadata file, name it with your institution code, group name and dataset type. Example: wu_blue17_PierisMetadata
- 2. In your first entry, you should include:
 - a. The current date
 - b. The names of all group members, and the initials each person will use to identify their metadata entries
 - c. All information you know about the dataset (download dates, origin of dataset, etc.)
- 3. Begin each subsequent entry with the current date and the initials of the person making the entry. Then record the changes you made to the dataset and the reasons for each modification.

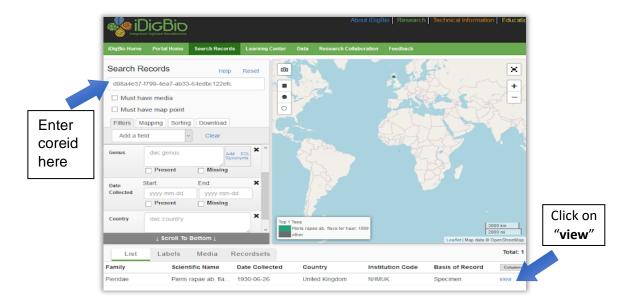


HOW TO DOWNLOAD, NAME, AND SAVE YOUR IMAGES

This process is a little complicated because you will need the image file name to connect back to the data on your spreadsheet. We will be doing this by using the specimen **coreid**. This is generally the first column in an iDigBio download.

If you downloaded your data from iDigBio, the folder contained 7 files. You will be using the "occurrence_raw.csv" file to download and appropriately record your file names. Alternatively, your instructor may provide your data file.

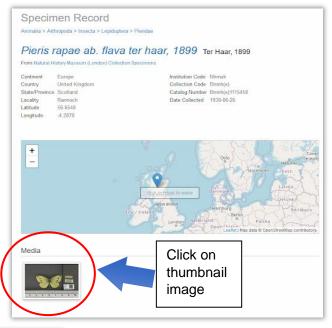
- 1. On the "occurance_raw.csv" downloaded spreadsheet or provided "occurrence" spreadsheet, find the column labelled "**coreid**". It usually is found in column A. This column contains the unique identifier for specimens in the iDigBio data portal. The coreid entries will be a long string of letters, numbers and dashes, such as "d98a4e37-f799-4ea7-ab33-64edbc122efc"
- 2. Add a new column to the right of the "coreid" column. Name the new column "imagefilename"
- 3. Open the <u>iDigBio search portal</u>.
- 4. Copy the coreid for the first of your assigned specimens. Paste it into the top search field, called "search all fields."



- 5. The specimen record associated with the coreid will appear at the bottom of the screen. Click on "view" at the far right of the specimen record information.
- 6. Scroll down through the specimen record. You can find information such as where and when the specimen was collected, and the museum or collection where it is currently located. Find the section labelled "**Media**"



- 7. Under "Media", you should find thumbnail images of any digital images of the specimen that are available. For some organisms, a single specimen record may include multiple organisms that were collected at the same time, in the same place, so don't be surprised if there are images of more than one specimen.
- 8. Click on the thumbnail image. You will now see a larger image.
- Below the image you will see the website where the image is stored, and a "Download File" link. Click "Download File". This will open the full size image.





- 10. Right **click on the image** to open a menu, then select "**Save Image As...**" to open the "Save Image" window. Choose the folder where you wish to save your file.
- 11. When you name an image, include the coreid in the file name, you will always be able to connect the image back to its associated data. You will name the images using a format similar to our datasheet naming convention:

Institution code_group name_coreid.jpg Example: wu_blue17_000d5e58-1b43-4277-b93a-a159d3f376b5.jpg

- a. If multiple image files have the same coreid (this will be very rare), then add on "image1", "image2", etc. to the end of the file name to distinguish them.
 - Example: wu_blue17_000d5e58-1b43-4277-b93a-a159d3f376b5_image1.jpg
- 2. Next, record the image file name in the "occurrence" spreadsheet. Add it to the "imagefilename" column. Make sure it is in the correct row. The coreid codes will match.
- 3. Finally, you will upload all of your renamed image files to the online folder for your working group.

