

Microscopy Data Management Primer

When you acquire an image and hit the “save” button, you face several seemingly inconsequential, but very important, questions. *What do I name it? Where do I save it?* Your answers to these and other questions will impact how easily you and others can access and use your data days, months, and years from now. Poorly managed data may be unusable by future-you, your lab, and other researchers—in other words, it may be useless. In contrast, well managed data may help you complete your project more easily, inform future research by your lab, and be a beneficial resource to the wider scientific community.

What are my data?

Your images... but also their metadata!

Assuming you took your **images** carefully and correctly, they are an accurate representation of what you observed through the microscope. To maintain their integrity, they need to be saved the correct file format, stored in a safe location, and never directly modified (see following sections for more details).

In addition, you will have **metadata** or “data about your data.” They are information about the instrumentation, parameters, and specimen preparation that were used to produce your data (images). While they may not seem important, they are critical for enabling someone (including future-you) to understand and reproduce your results. Some of these metadata are embedded in your images; others must be recorded by hand. They must be reported whenever you present or publish your data. We will discuss this more during your training.

In which file format should I save my image?

Use a file format that preserves your image exactly as you collected it.

In possible, you should save your image in the **microscope’s proprietary file format** (e.g. Nikon .nd2, Leica .lif). These file formats will contain the maximum amount of metadata meaning that you do not need to record as much by hand. Some systems do not have a proprietary format or your research advisor may have a reason why they don’t want to use them. In these cases, you should use TIFF (.tiff or .ome.tiff) or uncompressed/lossless JPEG-2000 (.jp2) files because these will save your data exactly as they were acquired.

Avoid “lossy” file types such as normal jpg/jpeg, gif, and png. They will compress your images, generating smaller files but also losing information. One advantage of these files is that they can be easily shared (email/text/chat) and opened on most devices, while the three file types in the previous paragraph cannot. If you want an image to share or put in a presentation, first save your image in one of the three good formats, then save a copy in one of the lossy/compressed formats.

What should I name my image file?

Names should be consistent, orderly, and informative.

Bad	Good
slide1b feb14 2016	slide1b_2016-02-14image1
Slide1B-feb14-2016 image2	slide1b_2016-02-14image2
1c 2-14-16 1 ...	slide1c_2016-02-14image1 ...

Organize images in folders to reduce the amount of information you need to put in image names. How you do it depends on what makes the most sense; just pick a scheme and stick with it. For example...

- Name folders by experiment/project (“exp10-protocolTest”)
- Name folders by date (“2016-01-01_more_cell_movies”)
- Make sub-folders (e.g. one folder for each slide in an experiment)

Do not use names to record notes about the experiment, e.g. “control2 re-do USE THIS ONE” or “treated3-weird-blob.” Instead, use **readme** files to provide context for your files (see below).

Good naming conventions

No special characters (/, %, etc.)

No spaces. Use ‘_’, ‘-’, or useCamelText.

Format dates: 2016-01-31 so they’ll sort correctly. Don’t use slashes (/).

Pad numbered files with 0s (file01, file02...file10) so they’ll sort correctly.

Where should I save it?

Save it somewhere safe, backed-up, and easily accessible to you and your lab.

Your lab likely already has a standard procedure for where to save data. Here are some additional tips:

- Do *not* keep your files on the facility computers. Data may be erased at any time and without warning.
- Store data on the biology server or in the cloud (e.g. OneDrive, OSF, Teams) because these locations are backed-up and easily accessible. Do *not* use photo-sharing sites. Only use local physical drives (lab computer, USB drives, etc.) as a last resort and, if you do, be sure they are backed-up to a location in another building.
- Be sure that your research advisor (and the rest of your lab, if applicable) can access your files. Remember that, if you are conducting research in a lab, your data is ultimately the property of your lab and university.
- Do *not* modify your original files. These are your data. If you need to modify them, e.g. during analysis or to prepare a figure, make a copy and work on the copy instead.
- Do *not* delete your data.

Is there anything else I should save?

Recommended: make a Readme file.

This is a simple text file (e.g. use Notepad) that you save alongside your images where you can keep notes about how your images were acquired and how they’re organized. This isn’t a replacement for your lab notebook; it’s a way for you and others to quickly identify and understand the images stored in a specific location. Examples of things to record:

- Name / date
- Explanation of the file naming and organization
- Where to find the experiment in your lab notebook
- What you’re imaging
- Notes about the images
- **Metadata** that aren’t already embedded in your image files

Why a basic text file? Text files will still be readable many years from now, take up very little space, and don’t require any special software to open.

Why not just write this information in your notebook? The whole point is to have the file “travel” with the data. It will point people back to your lab notebook and help people who are reading your notebook find the corresponding files.

What do I do with the metadata I collected?

Report them!

When you present your research, you must also tell your audience how you performed your experiments. This information not only helps your audience understand your results, but also allows them to verify that your interpretations are correct and ultimately to reproduce or build upon your findings. In addition to reporting how you prepared your specimens, you must report how you imaged them and how you used the images to arrive at your conclusions. There are many important imaging parameters that must be reported and, luckily, most of them are contained in the images' metadata. The level of detail you report usually depends on how you're presenting your research. Oral presentations contain only the most essential details while a manuscript should contain everything needed to fully reproduce your imaging. Posters typically include an intermediate level of detail. We will discuss what details need to be reported during your training.

