

Best Practices for Data Management When Using Instrumentation

A tutorial on effective data collection, saving,
and processing methods

Created by the Office of Research Compliance and Training
As part of the Research and Data Integrity (ReaDI) Program

Supported by Columbia University Standing Committee on the Conduct of Research

Why Management of Data is Crucial

- Good data management practices keep you organized
 - Data files are easier to locate
 - Less time spent repeating experiments because a data file is unable to be located
 - Reduced time searching for data files when writing publications or thesis
- Some federally funded grants and many publications require data be made publically accessible
 - Well managed data with easily identifiable file names aid in the distribution of data
- Maintaining accurate records provides integrity to research results, defends against accusations of fraud, and supplies proof of intellectual property

Outline

- I. Best practices for Labeling the Laboratory Notebook and Naming Data Files
- II. Planning Ahead
- III. Organization and Collection of Data
- IV. Additional Considerations Using Instrumentation and Processing Data
- V. Summary

I. Best Practices for Labeling the Laboratory Notebook and Naming Data Files

- This section will introduce:
 - The importance of using a laboratory notebook
 - Notebook labeling practices
 - Data file naming practices



Importance of the Laboratory Notebook

- A laboratory notebook is an important tool because:
 - It is a written record of procedures, reagents, data, calculations, thoughts, explanations, and results of experiments
 - It is a legal document used to defend intellectual property and accusations of fraud
 - It is knowledge for future researchers
 - It is the foundation of a thesis and other publications



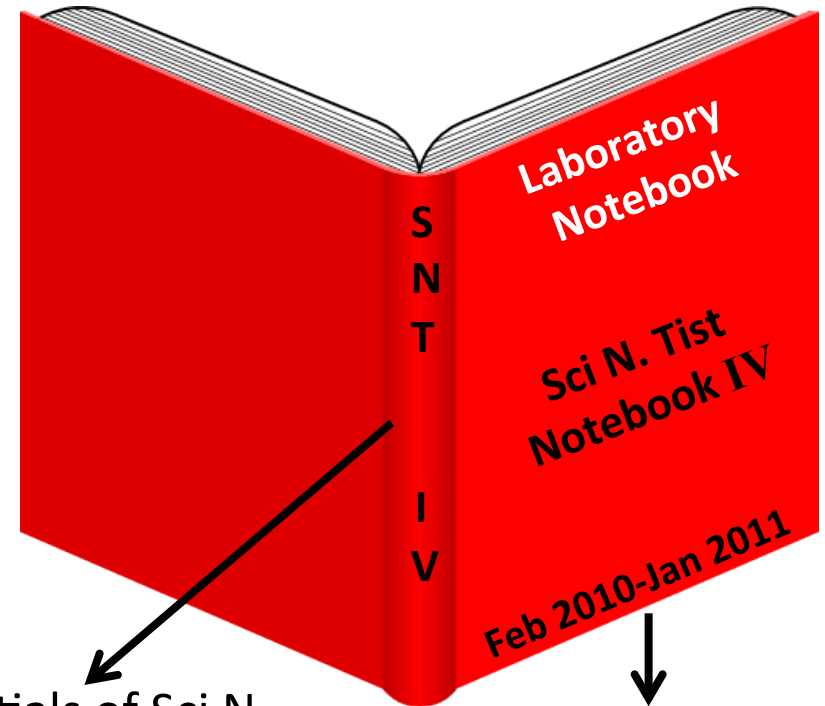
Laboratory Notebook: A Researcher's Tool for Data Organization

- Before getting started with data collection it is important to choose a labeling system for your notebook, and document it
- A notebook labeling system will be a tool used to reference data collected to a particular entry by the researcher or anybody wanting to view researcher's results
- For more information about good laboratory notebook practices refer to the *Good Laboratory Notebook Practices Tutorial*



Labeling Your Laboratory Notebook

- Check with your PI or lab manager for established conventions for your research group
- On the cover of your lab notebook include, as appropriate:
 - Your name
 - Your position
 - Your PI's name
 - The title of the project or study
 - Dates the notebook used
 - Volume number
- Label the spine if using a bound notebook



Initials of Sci N.
Tist and notebook
number

Indicates the
notebook belongs to
Sci N. Tist
and that it is the 4th
notebook used by Sci

Labeling Your Laboratory Notebook

- Include the same information on the first page of the notebook, along with:
 - Your email
 - Your phone number
 - Your lab and/or office address
 - The abbreviated name by which this notebook will be indicated in file names or other references

This notebook belongs to:

Sci N. Tist

Contact information:

ST00@columbia.edu

Office #: 212-555-555

Smith Hall

1000 First St. Room 300

NY, NY 01001

Dates used:

Feb 2010-Jan 2011

Notebook #:

IV





Notebook abbreviation:

SNT_IV

Best Practices for Data File Naming

- Check if there are standards that are required by your research group and/or discipline
- Files names should include:
 - Date of data collection (or data file created), listing year first sorts files chronologically
 - Notebook and page number that it refers to
 - If multiple files are being created on the same date and referring to the same sample, notebook, and page number use a numbering (or lettering) code to distinguish between data files
 - Additionally, file names may include the project identification or a brief description

Identifier:

	Date	Notebook/ Page #	Sample/ Special Notes	Series
	20101217	SNT_IV_285	Fe2O3_InLens_a	
	20101217	SNT_IV_285	Fe2O3_InLens_b	
	20101217	SNT_IV_285	Fe2O3_InLens_c	
	20101217	SNT_IV_285	Fe2O3_SE2Det	d

Interpretation: Date Taken: Dec 17th 2010 Sci N. Tist Notebook IV Page 285 Sample & detector used Different files of same sample indicated by continuation of lettering

Best Practices for Data File Naming, Continued

- Write down the name of the data file in the corresponding notebook to maintain a record of the data file being created
- Consider creating a text file, to be saved in the same folder, that defines the file naming convention
- Some conventions suggest an instrument identification
- Be consistent and document the naming convention that is chosen
- Special consideration needs to be given when naming files with PII (Personally Identifiable Information) or PHI (Personal Health Information). Refer to RASCAL trainings TC0019 or TC0087 for regulations outlined by HIPAA and human subjects research

All of the following would be considered equivalent file names:

20101217_SNT_IV_285_Fe2O3_inLens_a.tif

101217_SNT_IV285_Fe2O3_inLens_a.tif

101217_SNT4_285_Fe2O3inLens_a.tif



Tips for Naming Data Files

- Letters from A to Z and numbers 0 to 9
- Use underscores rather than spaces
- Use brief names (<30 characters)
- Avoid using special characters
 - Ex: \ ` ' • = / < > ^ : ; | # * () . ? % , “ @ ! + { } ~ - []
- The 3-letter file extension should be reserved for application-specific codes, e.g., .wrl, .mov, .tif
- Differentiate file name parts using mixed cases or underscores
- When creating a sequentially numbered series of files, use leading zeroes so that they sort properly e.g., file001.txt, file002.txt, file003.txt ... file042.txt

I. Summary

- Check with your PI or lab manager if there are standards for notebook and data file naming. Adhere to the expectations of your PI and research discipline
- Laboratory notebooks should be labeled so that it identifies the researcher and dates used, and if applicable, the research project
- Document your laboratory notebook labeling convention and be consistent
- File names of collected data should include the date of collection, notebook and page number identifier, and a short description
- Once a data file naming format is chosen, write it down and be consistent



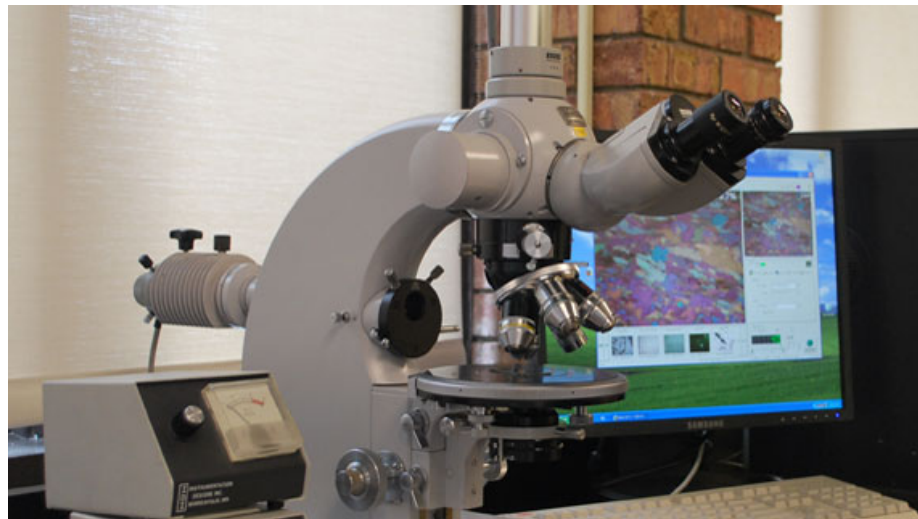
II. Planning Ahead

- This section will address
 - Important details to consider before experimentation
 - Notebook preparation
 - Tips for creating data tables
 - Information that needs to be included in the laboratory notebook



Important Details to Consider Before Experimentation

- Before beginning data collection, it is essential to know how the instrument saves and exports data
- Unsure how the instrument saves and exports data?
 - Ask a fellow group member (graduate student, post-doc, staff scientist, etc.)
 - Ask the person who maintains the instrument
 - Practice with an unimportant sample (blank)



Questions Related to Instrumentation

- What file format should data be exported as?
 - Know which file formats are compatible with other operating systems, analyses programs, and data archiving
 - Avoid using file extensions that are unique to a particular software used for data collection. If data are available only in closed or proprietary formats, are there programs available for file migration to a more useful format?
 - Most compatible formats include:
 - ASCII, PDF, .csv, FLAC, TIFF, JPEG2000, MPEG-4, XML, RDF, .txt, .r
- What descriptors should be added? What information will you need to find, identify, or understand the file in the future?
 - Units and other relevant information
 - Independent & dependent variable column headers
 - The ability to add notes or descriptors

Notebook Preparation

- Prior to running samples or starting the experiment, make sure you know the answers to the following, and that the answers can be found in the experiment documentation
 - How many samples will be looked at?
 - How is each sample different?
 - What descriptors will you use to distinguish each sample?
 - If the same sample is being probed multiple times, what are the varying parameters?
 - What is the minimum amount of information that must be recorded to reproduce the experiment?
- Create a table in your notebook with descriptive headers
 - Create a template (can use Word or Excel) that can be re-used each time that particular instrumentation is being used



General Tips for Creating A Table

- Use descriptive labels, including units, in header **row**, for each **column**
- The **column** furthest to the left should be the independent variable (the item that is changing, i.e. your sample)
- Dependent variables should follow in the next **columns**
- Any derived variables (from calculations) follow the dependent variables. Calculations and formulas need to be documented
- Derived data should reflect the same number of significant figures as the measured values
- If creating a template, the use of formatting (ex: **color**, **boldface**, or *italicized* fonts) can help visually organize data

Independent Variable ↓	Dependent Variables			Derived Variable ↓
Sample	pH (Trial 1)	pH (Trial 2)	pH (Trial 3)	Average pH
Unfunctionalized WO ₃ nanoparticles in H ₂ O	3.58	3.62	3.64	3.61
Citric acid modified-TiO ₂ nanoparticles in H ₂ O	5.38	5.44	5.45	5.42



Relevant Information to be Included in Notebook

- Any parameters that are *REQUIRED* to reproduce the experiment need to be recorded
 - E.g.: Substrates, energy output/input, solutions used, background samples, etc.
- Any observations regarding the instrumentation process that would be considered “unusual”
 - E.g.: noise background due to construction, elevated temperatures due to broken AC, etc.
- Any software that is unique to that instrument or a code that has been written for particular types of experiments
 - E.g.: LabVIEW program specifically designed for experiment
- Initial observations while data is being collected
- During data collection, any pre-planned tables should be filled in after each file is created

Example Notebook Entry

2010 December 17

Scanning Electron Microscopy (SEM) images of Fe₂O₃ nanoparticles. Synthesized 2010 December 15 as referred to on page 282 of this notebook

Instrument Parameters:
Accelerating voltage: 2.0 kV
Used both In-Lens and SE2 detectors
Substrate used: Si-wafer <0.001 Ω

Special notes: construction on floor created some minor difficulty with focusing images

- ✓ **Thorough description of sample**
- ✓ **Instrument parameters used**
- ✓ **Special notes**
- ✓ **Table that organizes data collected**
- ✓ **Descriptive file name with reference to notebook location**
- ✓ **Where saved data files are located**

Image File Name	Detector Used	Scale Bar
20101217_SNT_IV_285_Fe2O3_InLens_a	In Lens	1 um
20101217_SNT_IV_285_Fe2O3_InLens_b	In Lens	200 nm
20101217_SNT_IV_285_Fe2O3_InLens_c	In Lens	2 um
20101217_SNT_IV_285_Fe2O3_SE2_d	SE2	200 nm

Data file location in Research Group Archive:
SciNTist_Research → Instrumentation →
SEM → 20101217_SNT_IV_285

II. Summary

- Be prepared prior to data collection
- Create a table in notebook to aid in organization of the experiment and data collection
- The notebook needs to contain all the necessary information so that the experiment could be repeated by yourself or other researchers
- Include the data file name in the notebook entry along with the file location



III. Organization of Collected Data

- This section will evaluate the importance of:
 - Archiving data
 - An electronic filing system
 - Use of folders and sub-folders
 - Use of shared computers and portable storage devices
 - Transferring data from a portable storage device to data archive



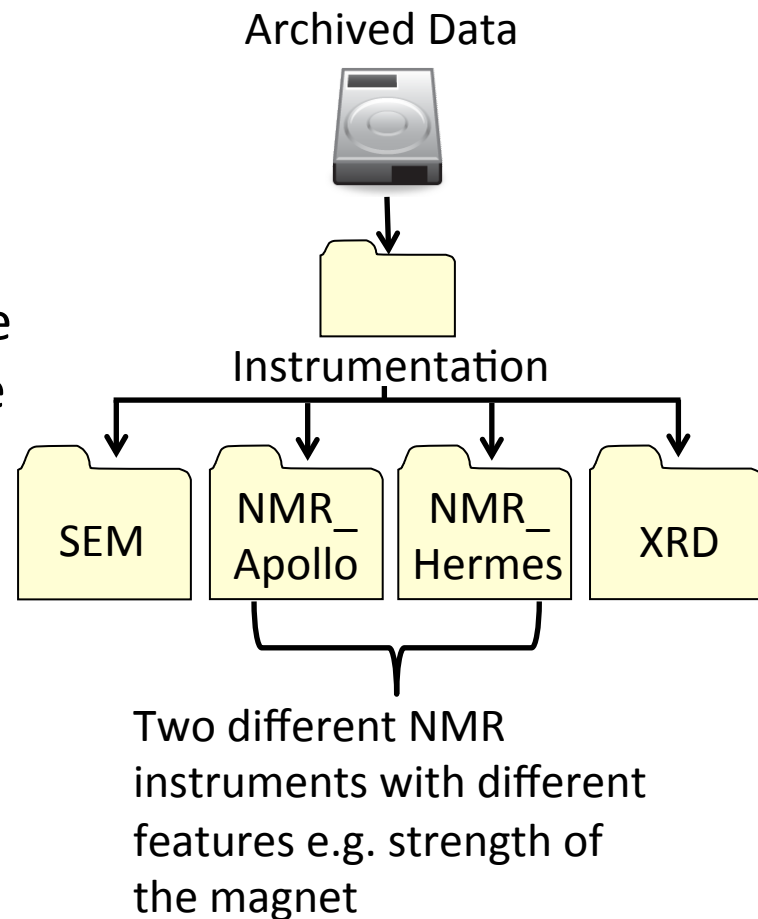
Archiving Data

- It is essential that a method for archiving data is established prior to any data collection
- It is the responsibility of the PI to maintain archived data; check with your PI on the procedures for storing and saving data
- Archived data should be secured as required by University policies and procedures, and state, and/or federal regulations
- Archived data should be in one location that is accessible by all group members and/or collaborators (e.g.: research group drive, online repository, etc.) with at least two back-ups
 - In addition to the file on the local machine, there should be one copy backed up on a separate local system, and one copy held at a geographically distant location
- For more information regarding data management and storage, please visit (Insert RDM tutorial link by Joel and Amy)



Electronic Filing System for Data Organization

- Before data collection, it is important to determine an electronic filing system that best suits your research project
- The use of a filing system helps organize collected data
- Within the archive, it may be useful to create an “Instruments” folder; at a minimum, a file for EACH type of instrument used to collect data should be created
- If more than one of the same type of instrument is used, but in a different facility, with different capabilities, etc., use underscores to help identify each one
 - It is *extremely* important to document where each data set originates, for reproducibility of experiments

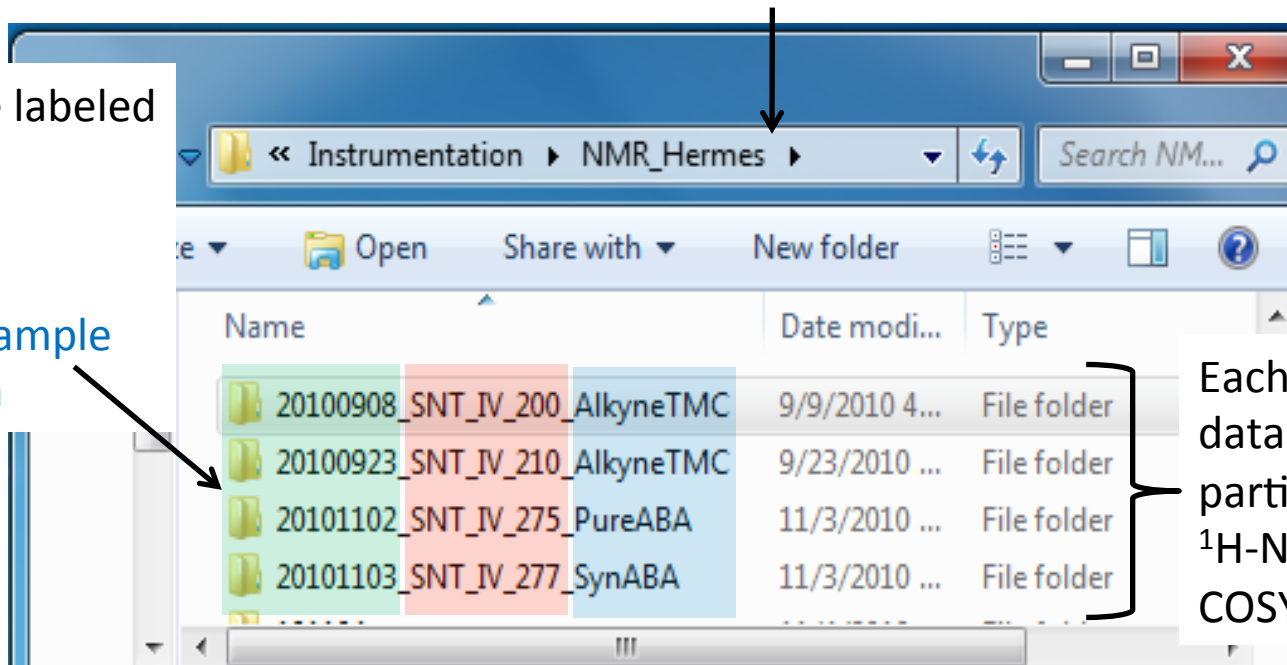


Organizing Sub-folders

- Within each instrument folder, create a new folder in which all the data collected can be organized
- Use of sub-folders helps keep data organized for easier referral
- Name sub-folders using date, descriptions, notebook identifiers, etc.
 - Remain consistent with the names of sub-folders

Within the NMR_Hermes folder, sub-folders have been created

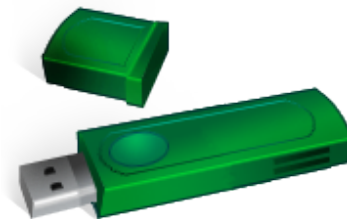
Folders are labeled with **date**, **notebook reference**, and **brief sample description**



Each folder contains data related to this particular sample, *e.g.*: ^1H -NMR, ^{13}C -NMR, COSY, etc.

Collecting Data Using Portable Data Storage

- Many instruments are connected to a computer that is shared with many other researchers from different groups and disciplines
- If privileges are allowed, create a folder on the shared computer with your name to use as temporary storage while collecting data
- **Do not** use shared computers for permanent (long-term) data storage
- Whenever possible, save data to a secure server that employs regular backup protocols, with quality control check
- A portable flash drive can be used to remove data from shared computers to a permanent storage location
- Use same folder and sub-folder organization on portable flash drive
- Use password protected flash drive if saving sensitive data



Transferring Data to a Permanent Location

- Flash drives should never be used as permanent storage
- Using the same folder and sub-folder organization facilitates transfer of data
 - Click and drag (or copy and paste) data files into correct folders on data archive
- Never delete data from portable device until data has been transferred and can be confirmed by either opening data file and/or using a checksum

The image shows two Windows Explorer windows. The left window displays the path 'Removable Disk (D:) > Instrumentation > NMR_Hermes' and a list of files. The right window displays the path 'Research > Instrumentation > NMR_Hermes' and a list of the same files. An arrow labeled 'Copy and Paste' points from the selected file in the left window to the right window. Another arrow labeled 'Same sub-folder identification' points from the 'NMR_Hermes' folder in the left window to the 'NMR_Hermes' folder in the right window.

Name	Date modified	Type
20100908_SNT_IV_200_AlkyneTMC	9/9/2010 4...	Folder
20100923_SNT_IV_210_AlkyneTMC	9/23/2010 ...	Folder
20101102_SNT_IV_275_PureABA	11/3/2010 ...	Folder
20101103_SNT_IV_277_SynABA	11/3/2010 ...	Folder

Name	Date modified	Type
20100908_SNT_IV_200_AlkyneTMC	1/28/2014 ...	Folder
20100923_SNT_IV_210_AlkyneTMC	1/28/2014 ...	Folder
20101102_SNT_IV_275_PureABA	1/28/2014 ...	Folder

CC BY NC COLUMBIA | READI PROGRAM

III. Summary

- A data archive system with multiple data back-ups should be established by the PI of the research group
- Creation of folders for different types of instrumentation keeps data files organized
- Use of sub-folders allows for easier referral to collected data files
- Portable flash drives are not meant for permanent data storage
- Data collected on portable devices should be transferred to a data archive ASAP
- Special requirements govern the proper handling of human subjects data



IV. Additional Considerations

Using Instrumentation and Processing Data

- This section will provide recommendations for the following scenarios:
 - Use of instrumentation with user modified software and/or programming
 - Use of instruments that have limited ability describing data being collected
 - Processing multiple data sets
 - Instrumentation ethics



Instruments That Allow Additional Remarks

- Some instruments have multiple ways of allowing additional comments within the program
- If you choose to utilize these options, then:
 - Be aware that these comments are often not exportable to another file format, and that the original proprietary program may be required to view these comments
 - Be sure to record ALL comments about the data file in your notebook as well
 - Do not use this option to replace descriptive comments in your notebook
 - Do not use this option to replace descriptive data file names

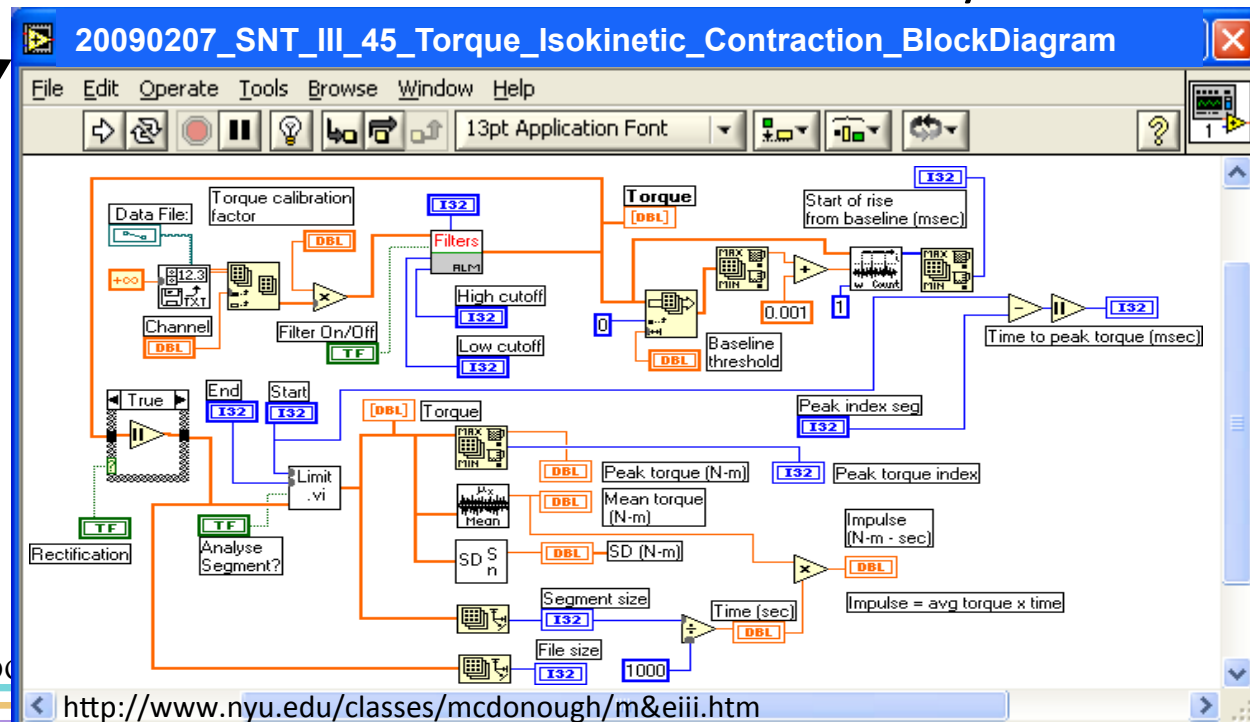
Commands	Parameters
Cyclic voltammetry potentiostatic	
Remarks	Cyclic voltammetry potentiostatic
End status Autolab	
Signal sampler	Time, WE(1).Potential, WE(1).Current
Options	1 Options
Instrument	MAC80007#1
Instrument description	
Autolab control	
Set potential	0.000
Set cell	On
Wait time (s)	5
Optimize current range	5
CV staircase	[0.000, 1.000, -1.000, 0.000, 2, 0.1000000]
Set cell	Off
<>	

- The user is able to include any identifiers about the sample
- The “Data View” option in this program allows the remarks to be seen; useful for quick analysis
- The remarks are not exported with the data file

Unique Software and Programming

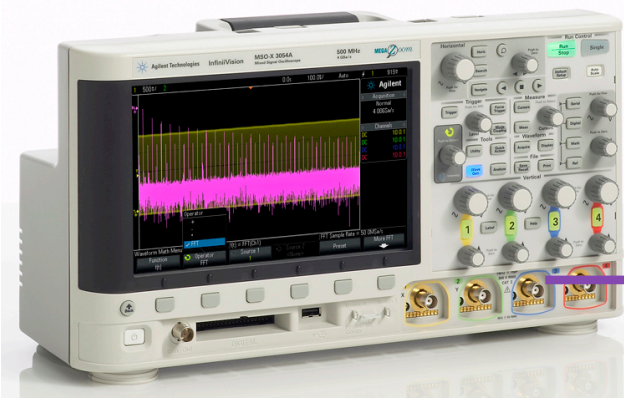
- When using software like LabVIEW to write a specific program, it is important to:
 - Record in notebook the EXACT methods and codes used, including hand-drawings and/or printed-out images affixed to laboratory notebook page
 - Describe the different elements used, including model numbers if applicable
 - Save the file so that it can be referred to the notebook entry

This LabVIEW was created on Feb. 7, 2009 to measure torque parameters and can be referred to page 45 in Sci N. Tist's notebook #3



Instruments That Assign Default File Names

- Some instruments do not allow users to give file names, but generate a default name
- For these types of instruments, it is useful to include an additional column in your table to record the default name given, then rename files later

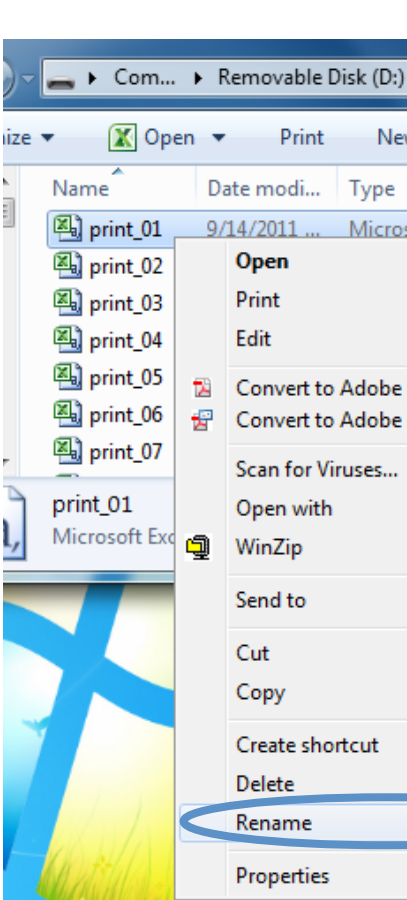


Default file name given to exported data from the instrument

Sample	Excitation wavelength	Pulse Energy (mJ)	Saved As	Renamed As
10 nm Ni Nanoparticles SNT_V_32	400 nm	0.100	Print_01	
20 nm Ni Nanoparticles SNT_V_15	550 nm	0.050	Print_02	
75 nm Ni Nanoparticles SNT_V_40	600 nm	0.100	Print_03	

Renaming Default File Names

- Default names give no description of the data they contain
- Rename data files ASAP using notebook table as guidance
 - Right-click file and select rename option or file renaming tools for multiple files
- Use a file name that is descriptive, so the files are easier to find for later referral



Sample	Excitation wavelength	Pulse Energy (mJ)	Saved As	Renamed As
10 nm Ni Nanoparticles SNT_V_32	400 nm	0.100	Print_01	20110914_SNT_V_32 _10nmNiNPs_400nm _Excite

Use of additional column for renamed filename

- maintains organization
- maintains a written record in notebook of data file being created
- simplifies reference when searching for electronic data files



Processing Data

- Some data sets require further manipulations after collection (e.g., baseline correction, background subtraction, etc.). These transformations cannot take the place of raw data
- Never alter the raw data set
 - Indicate that it is the raw data in the file name
 - Save each data manipulation with its own unique file name
- If working with multiple data sets in one spreadsheet or database, use descriptive titles that refer to the raw data file

File name of data and brief description

Explains how data was manipulated

Independent variable and instrument background

Describes the instrument parameters, date, and dependent variables

	A	B	C	D	E	F
1		Measured Photovoltage at 350 nm excitation for diffent SnO ₂ Samples, taken 20100401				
2		Voltage (mV)				
		Dark Background Noise Signal	20100401_SNT_IV_34 100 nm SnO ₂ Nanorods Synthesized 20100329	20100401_SNT_IV_34 100 nm SnO ₂ Nanoparticles Synthesized 20100315	20100401_SNT_IV_34 100 nm SnO ₂ Nanorods Background Subt (C#-B#)	20100401_SNT_IV_34 100 nm SnO ₂ Nanoparticles Background Subt (D#-B#)
3	Time (sec)					
4	0.00E+00	1.0547E-04	2.8320E-03	1.2793E-03	2.7266E-03	1.1738E-03
5	5.00E-10	1.4453E-04	2.9297E-03	1.4258E-03	2.7852E-03	1.2812E-03
6	1.00E-09	1.2109E-04	3.1250E-03	1.9531E-03	3.0039E-03	1.8320E-03

Instrumentation Best Practices

- A written record (either paper or digital) should be used to record who used the instrument and when. Be sure to fill out the log accurately
- In the logbook, record any malfunctions, maintenance performed, changed settings, etc.
- Most instruments are equipped with security controls, which identifies person and timestamp. These should NEVER be disabled
- Each user/group should have unique login and password. NEVER share your password, especially with unauthorized users
- Login and logbook information provides “audit trail” of instrument usage

Log of Xenon Lamp Usage and Water Raman Peak Intensity (2000-2010)

Date	Start Time	End Time	Hours Used	Total	Water Range
5/1/06	12:30pm	5:15pm	4:45	4:45	1.00 × 10 ⁶
8/14/06	4:15pm	5:20pm	1:05	5:50	1.00 × 10 ⁶
8/16/06	11:40am	1:00pm	1:20	6:10	1.16 × 10 ⁶
8/16/06	3:00pm	5:25pm	2:25	8:35	
11/1/08	12:40 PM	2:20 PM	1:40	10:15	1.00 × 10 ⁶
11/1/08	4:30 PM	5:40 PM	1:10	11:25	5.71 × 10 ⁵
11/1/08	6:00 PM	7:00 PM	1:00	12:25	3.5 × 10 ⁵
11/1/08	2:10 PM	3:15 PM	1:05	13:30	2.09 × 10 ⁶
11/1/08	7:00 PM	9:40 PM	2:40	16:10	
11/11/09	12:20pm	2:20pm	2:00	18:10	
10/13/09	7:15pm	9:30	2:15	20:25	1.00 × 10 ⁶
10/13/09	5:15pm	8:45	3:30	23:55	
11/14/09	1:30 PM				

Summary of Best Practices for Data Management Using Instrumentation

1. Always check with your PI about expected standards to be followed by all group members
2. Use documented notebook and data file naming conventions to keep data organized
3. Prepare notebook prior to the beginning of experimentation
4. Use a table to organize data as it is collected
5. Organize data using electronic filing system with descriptive folder and sub-folder names
6. Use portable storage only as TEMPORARY means to transfer data to a permanent location
7. Understand the capabilities of instruments regarding exporting data files
8. When processing data, never alter raw data sets
9. Use clear descriptors if combining multiple data sets to one spreadsheet
10. Use instrument logbooks and do not share login information

Additional Resources and Further Reading:

- Tutorial on *Good Laboratory Notebook Practices*
- Tutorial on *Data Management*
- RASCAL trainings TC0019 and/or TC0087 for PHI, PII
- Laboratory notebook and data file naming best practices:
 - <http://datalib.edina.ac.uk/mantra/libtraining/Session2GroupExercise1.pdf>
 - <http://uresearch.miami.edu/?p=103&s=25>
 - https://www.training.nih.gov/assets/Lab_Notebook_508%28new%29.pdf