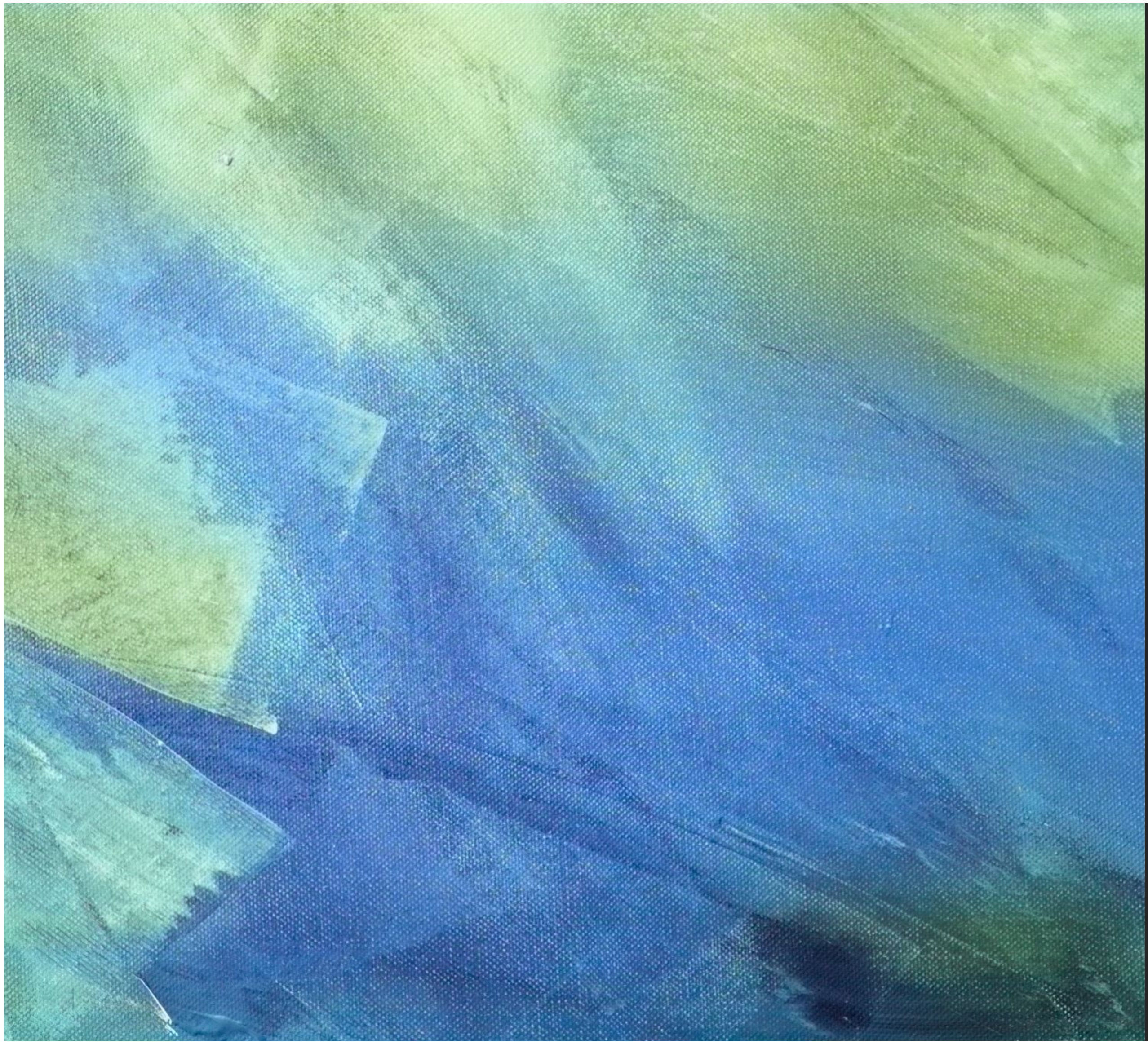


DATA TECHNIQUES FOR ENGINEERS AND SCIENTISTS

J.D. LANDGREBE

DEPT. OF CHEMICAL ENGINEERING

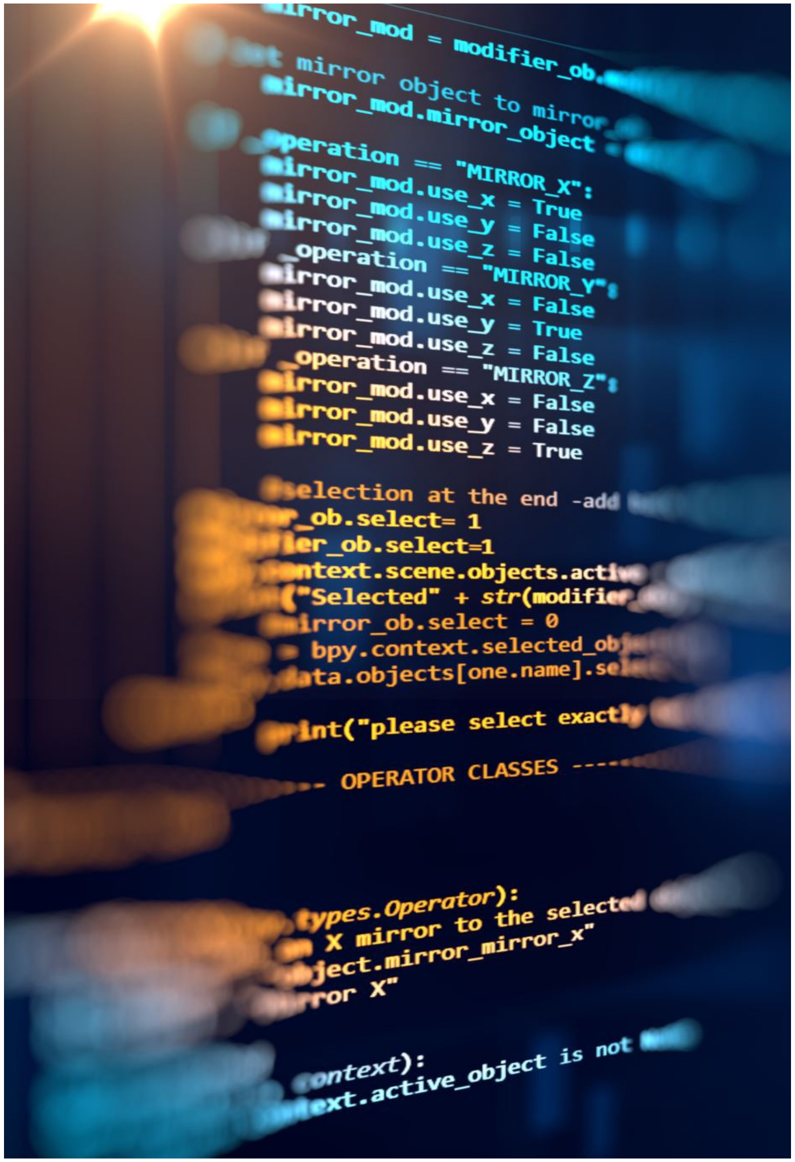
UNIVERSITY OF DELAWARE

LECTURE OUTLINE TEMP

- Baseline Software for engineering work – data analysis and modeling
- Exercise: install Anaconda/run intro to Pandas
- Understand available Python packages for general work (and understand Anaconda interaction with them)
- Nomenclature and data architecture for experiments
- What is experiment Curation?
- Data reshaping techniques Joining and summarizing an experiment –
- Exercise with JMP Tables menu – Join, Concatenate
- Parsing raw data – Exercise
- Work a case study of analyzing an experiment – TBD based on Will input

SOFTWARE FOR ENGINEERS

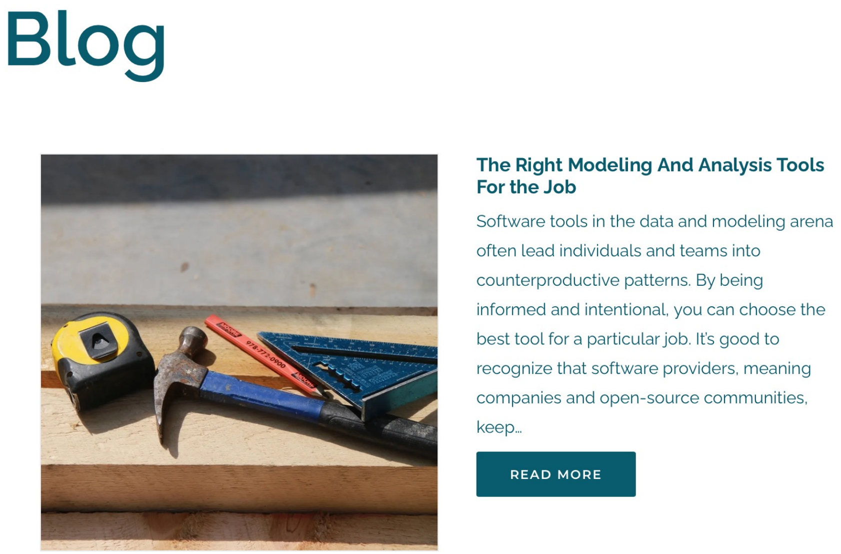
- Pick software by “use case” and use each software in its “lane” where it is best tool for the job
- “Use cases” start with “I need to…” (or “my customer needs to…”)
  - My supervisor needs to be able to scroll through our raw data to understand the results
  - I need to create a correlation matrix for my raw material’s quality data
  - I need to build a validated calculation model for reaction kinetics
- Use case focus
  - Avoids getting (yourself and others) distracted by things that software X is <u>not</u> good at – there is no perfect software tool
  - Puts a premium on making data portable across applications



```
irror_mod = modifier_ob.
t mirror object to mirror
irror_mod.mirror_object
operation == "MIRROR_X":
irror_mod.use_x = True
irror_mod.use_y = False
irror_mod.use_z = False
_operation == "MIRROR_Y":
irror_mod.use_x = False
irror_mod.use_y = True
irror_mod.use_z = False
_operation == "MIRROR_Z":
irror_mod.use_x = False
irror_mod.use_y = False
irror_mod.use_z = True

selection at the end -add
_ob.select= 1
er_ob.select=1
ntext.scene.objects.activ
("Selected" + str(modifier
irror_ob.select = 0
 bpy.context.selected_obj
ata.objects[one.name].sel

int("please select exactl

- OPERATOR CLASSES ----

types.Operator):
 X mirror to the selected
ject.mirror_mirror_x"
rror X"

context):
ext.active_object is not
```

“TOP LEVEL, GENERALIST” SOFTWARE RECOMMENDATIONS BY USE CASE

- **JMP® software** - design of experiments (DOE), exploratory analysis and data visualization with accompanying statistical analysis
- **Microsoft Excel®** - make data democratically usable by others, generate “end of the pipeline” reports and for create spreadsheet models with calculations for use by non-coders
- **Python scripts** for data reshaping and for developing coded models and data pipelines (possibly mixing in a little SQL)



Blog

The Right Modeling And Analysis Tools For the Job

Software tools in the data and modeling arena often lead individuals and teams into counterproductive patterns. By being informed and intentional, you can choose the best tool for a particular job. It's good to recognize that software providers, meaning companies and open-source communities, keep…

https://datadelveengineer.com/the-right-modeling-and-analysis-tools-for-the-job/

SOFTWARE RECOMMENDATIONS BY USE CASE

- Python scripting tools - build model and data cleaning scripts
  - JupyterLab (Anaconda)
  - VS Code script and text editor
  - Github Copilot and Chat extensions VS Code ($$)
- VBA Scripting Language – Excel automation
- Advanced visualizations – Python Matplotlib
- Github - Open sharing of projects, trainings etc.

PYTHON LIBRARIES TO EXTEND FUNCTIONALITY

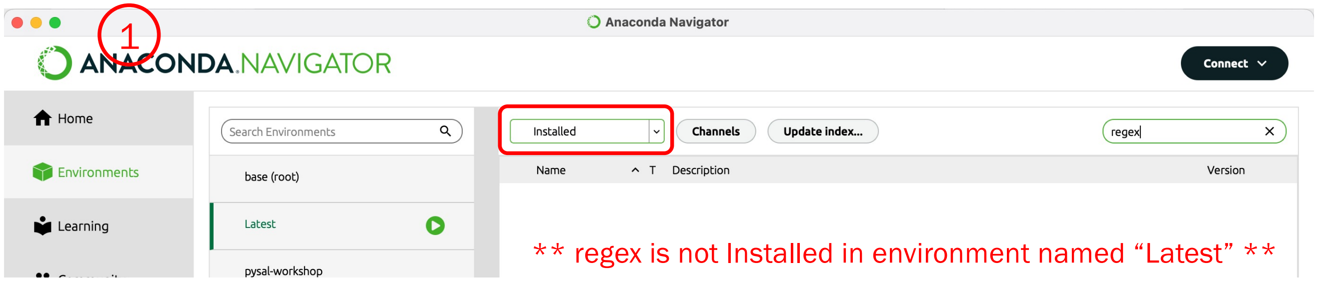
Useful Python libraries for Engineers

You get most of these with Anaconda installation!!

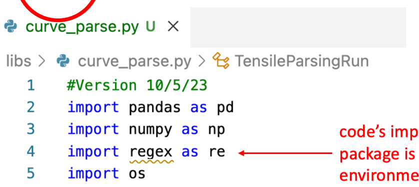
- Pandas – Data reshaping and analysis
- Numpy – Numerical calculations
- Pytest – Test/Validate your scripts
- OpenPyXL – Format outputs in Excel
- Re (Regular Expressions aka Regex) – Parse text strings

regex 2023.10.3

INSTALLING PACKAGES (AKA LIBRARIES) IN ANACONDA

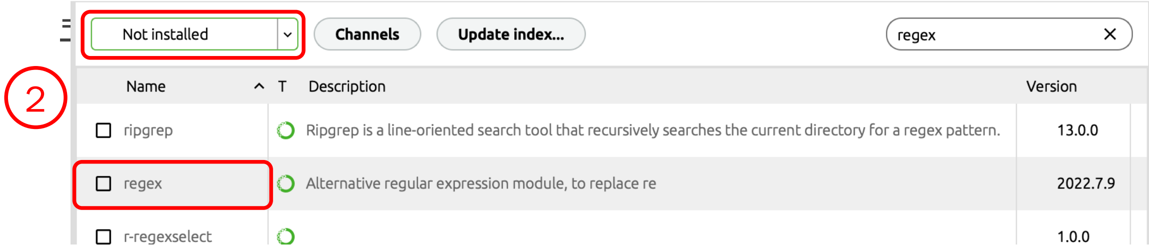


** regex is not Installed in environment named “Latest” **

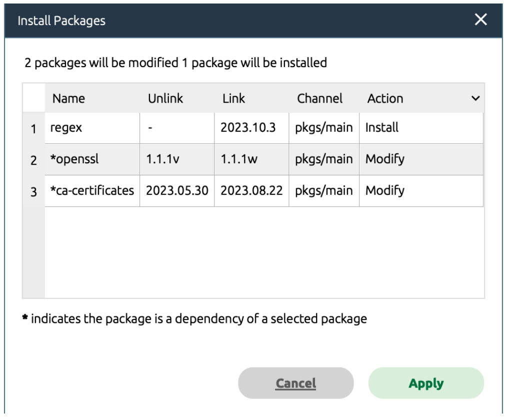


```
#Version 10/5/23
import pandas as pd
import numpy as np
import regex as re
import os
```

code’s import statement will error if package is not installed in Anaconda environment etc.



Name	T	Description	Version
□ ripgrep	○	Ripgrep is a line-oriented search tool that recursively searches the current directory for a regex pattern.	13.0.0
□ regex	○	Alternative regular expression module, to replace re	2022.7.9
□ r-regexselect	○		1.0.0



2 packages will be modified 1 package will be installed

	Name	Unlink	Link	Channel	Action
1	regex	-	2023.10.3	pkgs/main	Install
2	*openssl	1.1.1v	1.1.1w	pkgs/main	Modify
3	*ca-certificates	2023.05.30	2023.08.22	pkgs/main	Modify

* indicates the package is a dependency of a selected package

Anaconda selects related (aka “dependent”) packages needed by the selected package

OBJECTIVES

Experiments* have a common data structure (we will call it "architecture").

Recognizing this lets us master common data techniques to analyze efficiently and have data in good formats for graphing

Objectives

- Learn reusable terms for the data elements from experiments
- Learn how use common software to do needed transformations to go from raw data to analyzed summary
- Learn how to "curate" experimental data to make it easy to find and share

* Data don't know whether they were generated in a lab or virtual experiment, so this applies to computer modeling data and physical experiments

WHAT ARE EXPERIMENTS?

Experiments Cover An amazing range for engineers and scientists!

Make a production or lab-scale batch of several product formulations

Run a packing line using pre-set conditions for a set amount of time

Crash a sensor-equipped car into a wall

Diaper a baby with several product designs

Drive a F1 car around a track with several various fuel mixtures

Ask a man to shave his face with 3 different types of razors on consecutive days

Make batches of cookies with different types of chocolate

Wash test, fabric swatches with different detergent formulations

Use different catalysts to run a chemical conversion with the same starting materials

Use PCR to replicate and sequence the DNA from multiple virus samples

Mop floors with mops using different cleaning solutions

Surgically implant artificial joints using different polymeric coatings

WHAT DO THESE HAVE IN COMMON?

Experiments (or ”**studies**”…we will use synonymously) are a building block for designing and confirming products.

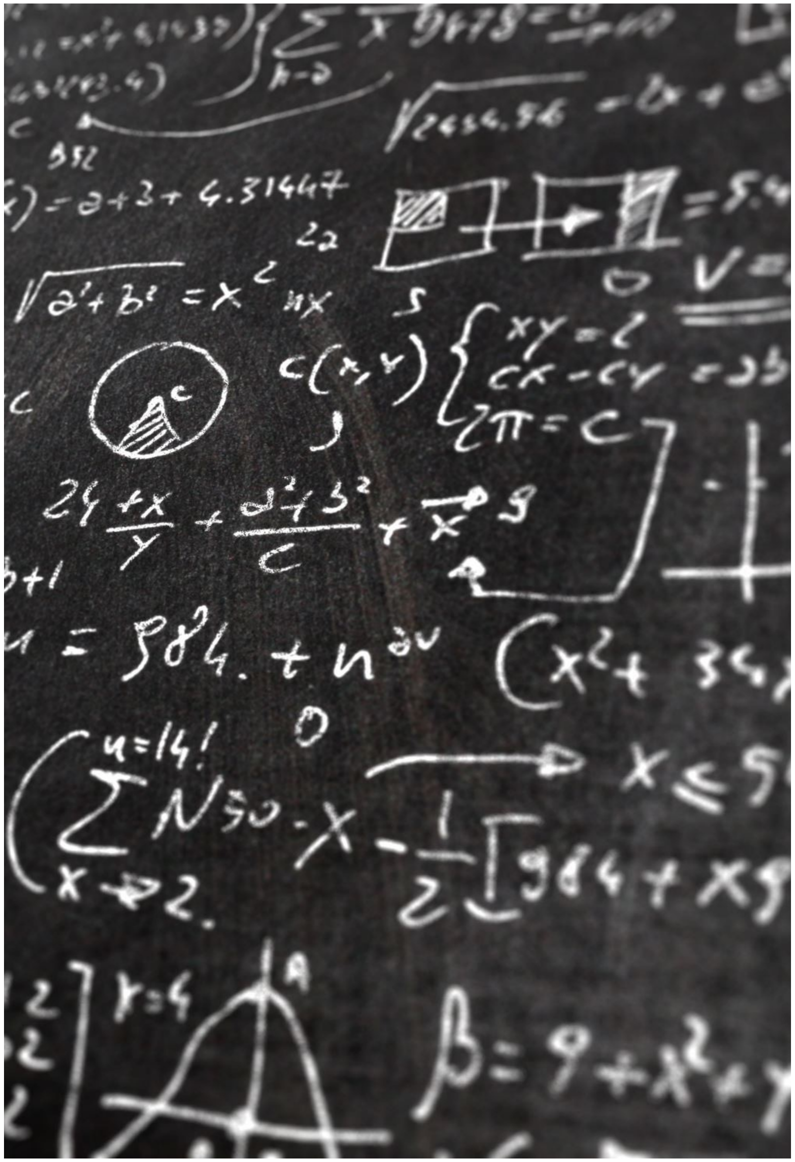
Experiments have “**runs**” consisting of one or more conditions having controlled inputs such as the starting materials and test conditions

Each run can be described by a list of the “**run variable**” values which are the pre-set inputs and important ambient conditions recorded during the run.

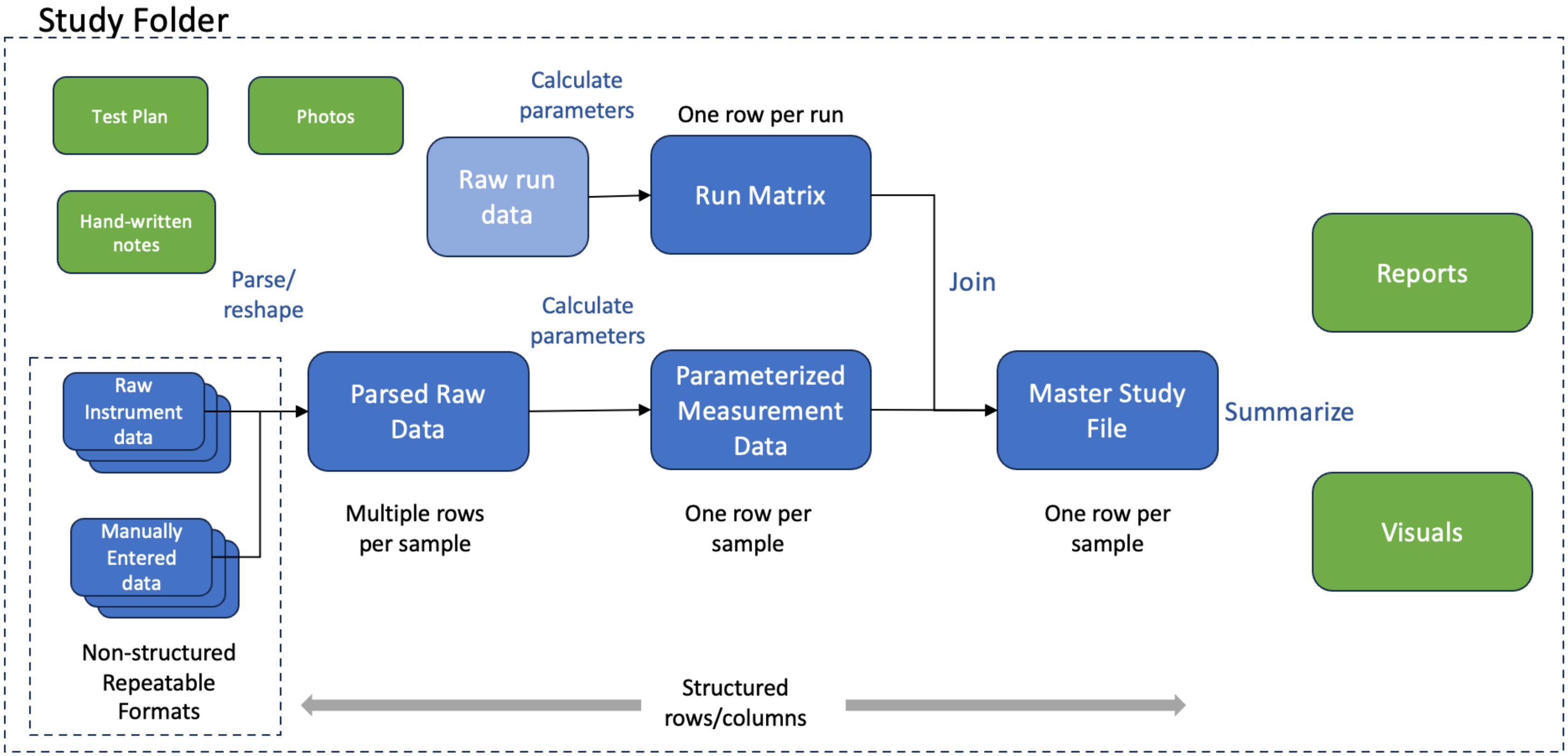
Experiments generate “**raw data**” consisting of the original measurements to assess what happened (detailed speed versus time data for F1 car circuit around a track)

The raw data gets converted to “**parameterized data**” which is quantities reflecting the outcome (average speed for a complete track circuit)

The run variable values can be assembled into a “**run matrix**” with a row of values for each run



EXPERIMENT DATA "ARCHITECTURE"



WHY DOES “NOMENCLATURE” MATTER?

1. It is reusable ”lens” that you can use in any context
2. Identifies efficient, generic ways of converting the Run Matrix and Raw Data into analyzed data and conclusions
   - Parsers to convert raw data to structured tables
   - “Join” and “Concatenate” to bring data together into a “Master Study File”