

# A “data pipeline” for BIOS-SCOPE

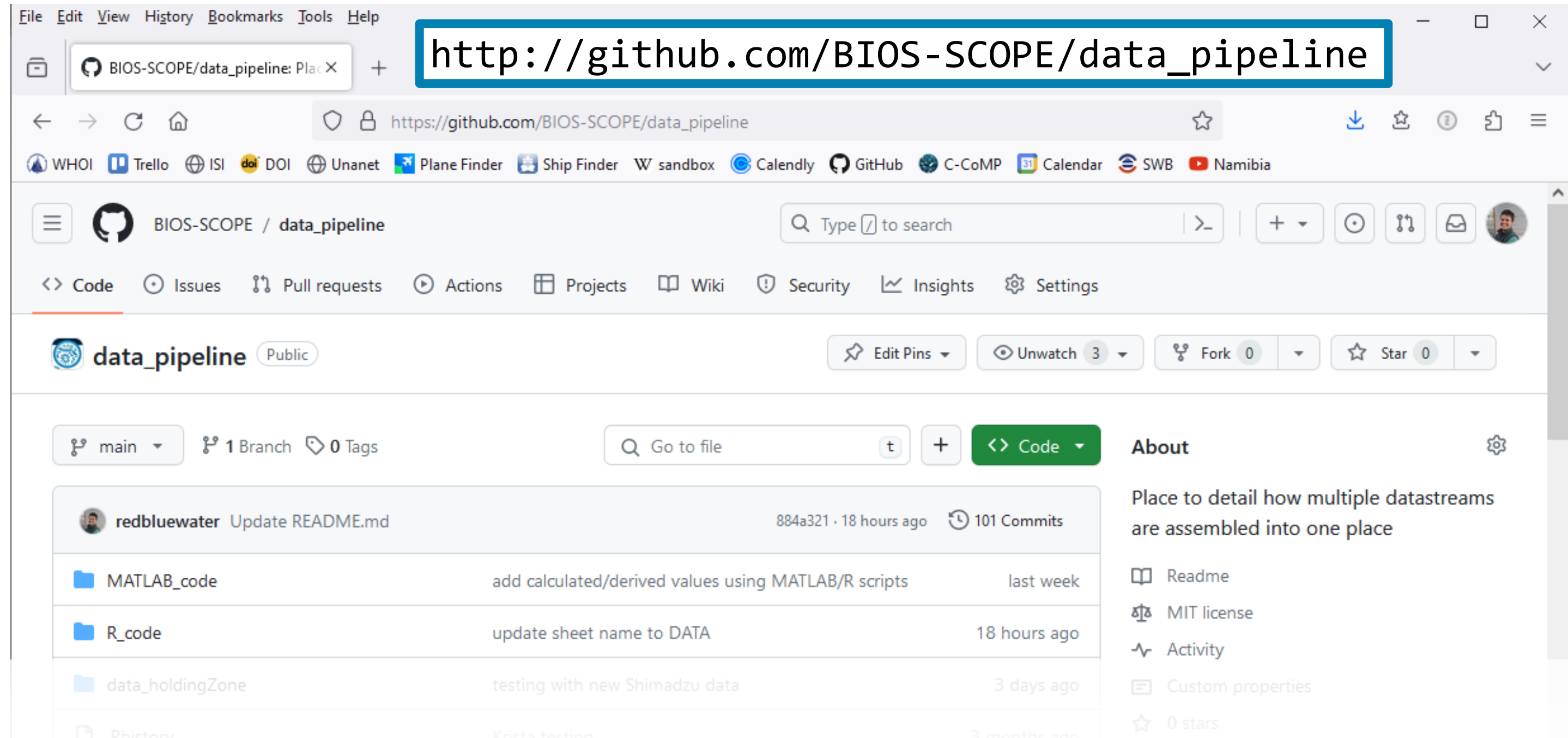
Krista Longnecker, updated 13 February 2026  
Pipe line is based on code from Shuting Liu and Ruth Curry  
With thanks to Elisa Halewood and Rachel Parsons



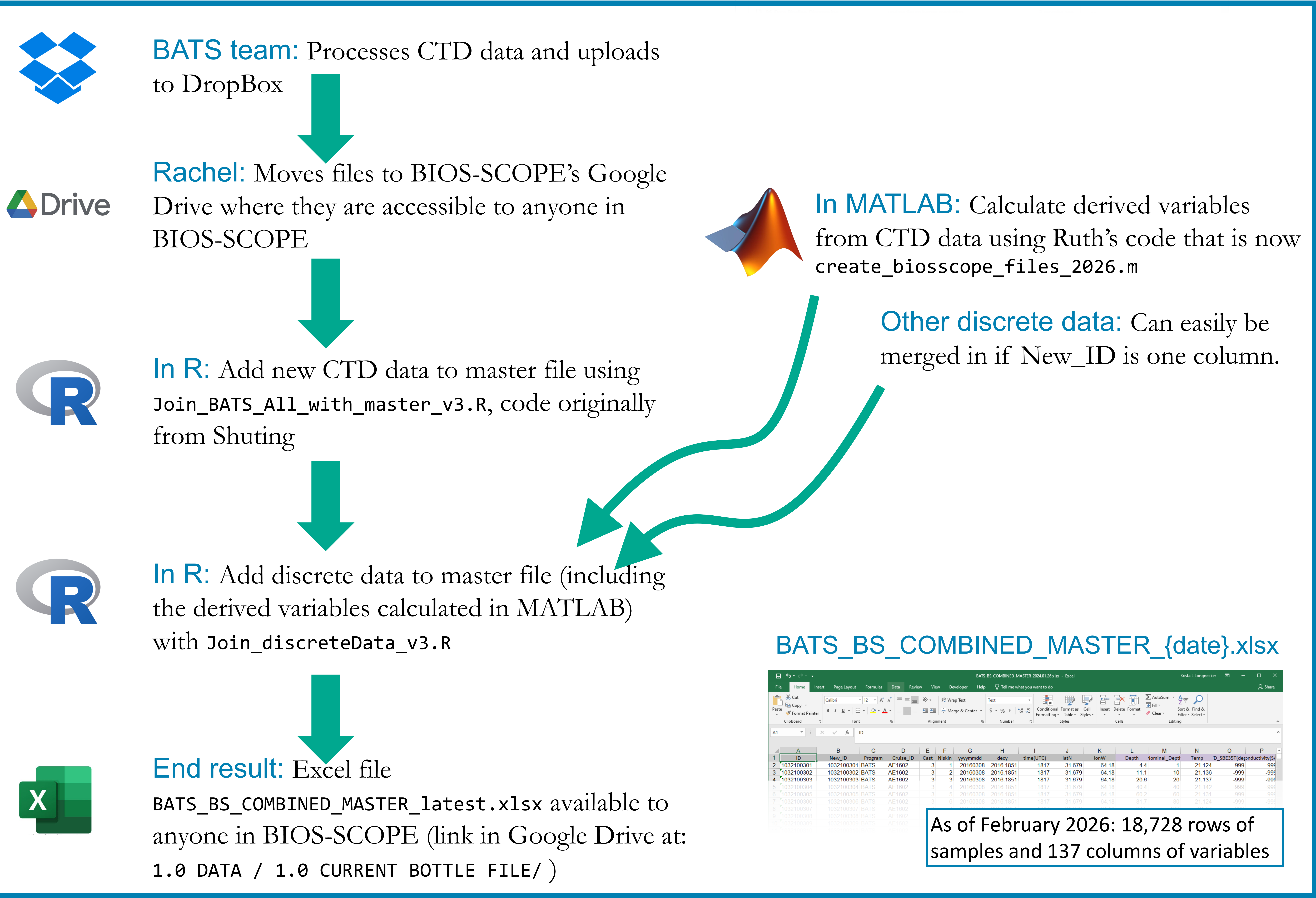
## What is the goal?

We want a single file that combines CTD data, variables calculated from CTD data, and data from discrete samples.

All of this code is available online at GitHub



## Step-by-step:

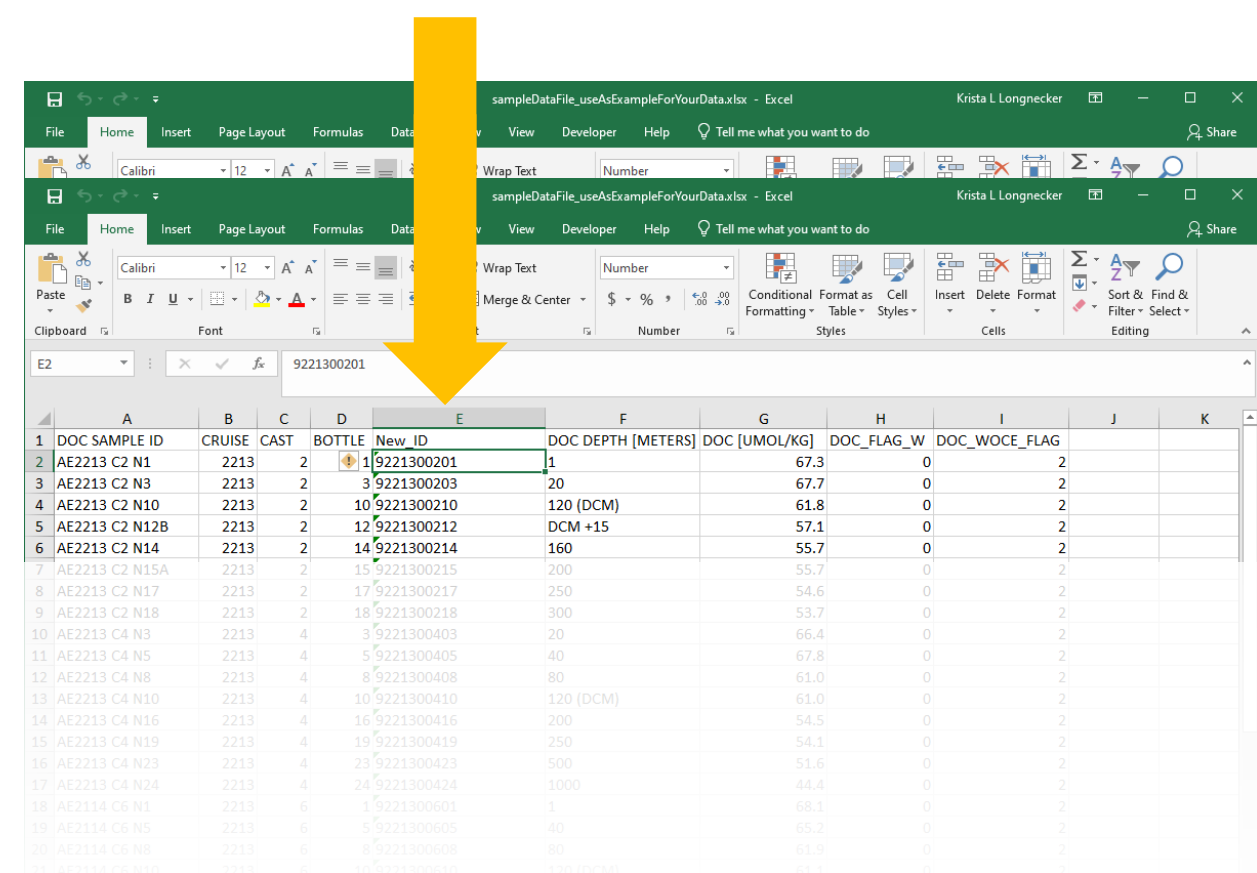


## Another way of describing this pipeline:

at sea	Generate CTD data	Collect water for discrete samples
on land	BATS team: CTD data processing	Process discrete samples (in lab)
	BATs team: move data to DropBox	End result: data file for each set of discrete samples
	Transfer data to BIOS-SCOPE Google Drive	
	Assigned to: Rachel Parsons (or Craig Carlson)	
	Get R script from <a href="https://github.com/BIOS-SCOPE/data_pipeline/R_code">github.com/BIOS-SCOPE/data_pipeline/R_code</a>	
	<code>Join_BATS_All_with_master_v3.R</code>	
	Status: generates CSV file with new samples	
	Action required: copy the CSV file as new rows in the existing bottle file	
	Plan: add new rows in a way that automatically saves a new version of the bottle file	
	Get MATLAB scripts from <a href="https://github.com/BIOS-SCOPE/data_pipeline/MATLAB_code">github.com/BIOS-SCOPE/data_pipeline/MATLAB_code</a>	
	<code>Create_biosscope_files_2026.m</code>	
	Status: generates CSV file with the calculated variables	
	Action required: the derived variables are copy/pasted into the bottle file	
	Plan: edit the bottle file to insert the calculated variables automatically	
	Get R script from <a href="https://github.com/BIOS-SCOPE/data_pipeline/R_code">github.com/BIOS-SCOPE/data_pipeline/R_code</a>	
	<code>Join_discreteData_v3.R</code>	
	Status: generates a new bottle file with data from discrete samples	
	Action required: need to update the log in the Excel file with user/date information	
	Plan: need to see an individual data file	

### How to add your data:

If you have new data to add to the master bottle file, you must have a column with **New\_ID** in your data file.  
The new data files can be Excel files (xlsx) or a CSV file.



### Acknowledgements:

This work is supported by funding from the Simons Foundation International's BIOS-SCOPE program.