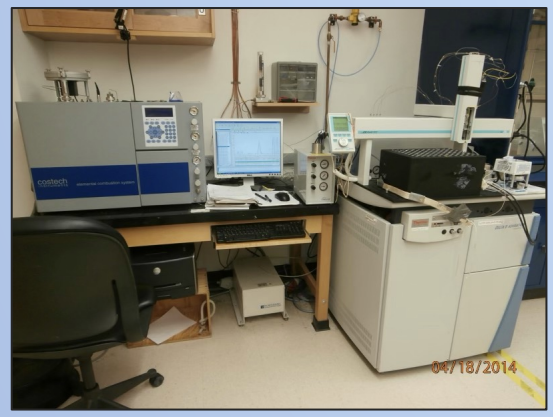
Advancing Data Curation and Archiving: an Application of Coding to Lab Management in the Geosciences

Tierney Latham[1], Catherine Beck[2], Bruce Wegter [3], Ahra Wu[4]

[1]Hamilton College, Geosciences, tlatham@Hamilton.edu, [2]Hamilton College, Geosciences, ccbeck@Hamilton.edu, [3]Hamilton College, Sciences, bwegter@Hamilton.edu, [4]Hamilton College, Library & Information Technology Services, axwu@Hamilton.edu

Hamilton Isotope Laboratory

- Lab group specializing in stable isotope analyses
- Hamilton College: a liberal arts, 4-year undergraduate institution of about 1,800 students located in Clinton, NY, USA
- **The problem:** faced significant barriers to proper data curation, storage, and accessibility



Courtesy of Bruce Wegter

Methods and Results

The solution: An undergraduate geoscience student was recruited to standardize previously produced data and evaluate current data storage procedures, with a goal of uploading data to **EarthChem**, a public repository for geochemical data.

Manual Excel Cleaning

- Edited 283 files of IRMS data by hand

** Established standard nomenclature and formatting with less barriers to coding*

Coding in Rstudio

- Identified and removed duplicate files
- Converted individual files to a new template
- Converted folders of files to a new template

** Files conform to a standard, organized template compatible with EarthChem submission requirements*

Re-evaluation of lab practices

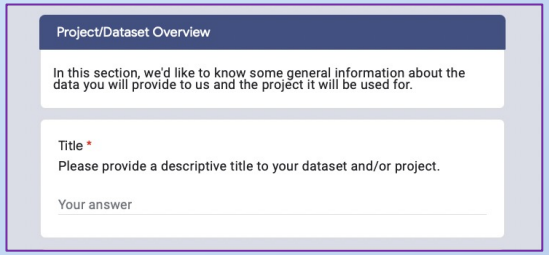
- Proposed Metadata survey
- Proposed new workflow

** Further removes barriers to repository submission*

	δ13C Reg	y = 0.9953x + 2.9258 R² = 0.9961

	E	F	G
	Errors	δ13C Regression	δ13C Regression R^2
	No	y = 0.9953x + 2.9258	0.9961
	No	y = 0.9953x + 2.9258	0.9961
	No	y = 0.9953x + 2.9258	0.9961
	No	y = 0.9953x + 2.9258	0.9961
	No	y = 0.9953x + 2.9258	0.9961

	A	B	C	D	E
1	Sample Name	IGSN	Latitude	Longitude	Elevation
2	C10-1				
3	C10-2				
4	C20-1				
5	C20-2				



Takeaways

1. Updating to a new system of data collection and curation is a challenging and laborious endeavor.
2. Coding can be beneficial to data curation, particularly with previously produced data.

References & Acknowledgments

EarthChem, https://www.earthchem.org
Rstudio, http://www.rstudio.com

Thank you to the LITS Isotope Database Team of Lisa McFall, Gisella Stalloch, and John Drew.

For more information: