

EGU21-3595, updated on 01 Dec 2022

<https://doi.org/10.5194/egusphere-egu21-3595>

EGU General Assembly 2021

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



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

Tierney Latham<sup>1</sup>, Catherine Beck<sup>1</sup>, Bruce Wegter<sup>1</sup>, and Ahra Wu<sup>2</sup>

<sup>1</sup>Hamilton College, Geosciences, Clinton, NY, United States of America

<sup>2</sup>Hamilton College, Library & Technology Services, Clinton, NY, United States of America

Increases in technology have rapidly advanced the capabilities and ubiquity of scientific instrumentation. Coupled with the demand for increased transparency and reproducibility in science, these advances have necessitated new systems of data management and archival practices. Laboratories are working to update their methods of data curation in line with these evolving best-practices, moving data from often disorderly private domains to publicly available, collaborative platforms. At the Hamilton Isotope Laboratory (HIL) of Hamilton College, the isotope ratio mass spectrometer (IRMS) is utilized across STEM disciplines for a combination of student, faculty, and course-related research, including both internal and external users. With over 200 sets of analytical runs processed in the past five years, documenting instrument usage and archiving the data produced is crucial to maintaining a state-of-the-art facility. However, previous to this project, the HIL faced significant barriers to proper data curation, storage, and accessibility including: a) data files were produced with variable format and nomenclature; b) data files were difficult to interpret without explanation from the lab technician; c) key metadata tying results to respective researchers and projects were missing; d) accessibility to data was limited due to storage on an individual computer; and e) data curation was an intellectual responsibility and burden for the lab technician. Additionally, as the HIL is housed within an undergraduate institution, the high rate of turnover for lab groups created additional barriers to the preservation of long-term, institutional knowledge, as students worked with the HIL for a year or less. These factors necessitate the establishment of new data management practices to ensure accessibility and longevity of scientific data and metadata. In this project, 283 Excel files of previously recorded data generated by the HIL IRMS were modified and cleaned to prepare data for submission to EarthChem, a public repository for geochemical data. Existing Excel files were manually manipulated, several original R code scripts were generated and employed, and procedures were established to backtrace projects and collect key metadata. Most critically, a new internal system of data collection was established with standardized nomenclature and framework. For future usage of the IRMS, data will be exported directly into a template compatible with EarthChem, thereby removing barriers for principal investigators (PIs) and research groups to archive their data in the public domain upon completion of their projects and publications.