

EGU21-2521, updated on 01 Dec 2022
<https://doi.org/10.5194/egusphere-egu21-2521>
EGU General Assembly 2021
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.



Automation of (meta-)data workflows from field to data repository

Linda Baldewein, Ulrike Kleeberg, and Lars Möller

Helmholtz-Zentrum Geesthacht, Geesthacht, Germany (linda.baldewein@hzg.de)

In Earth and environmental sciences data analyzed from field samples are a significant portion of all research data, oftentimes collected under significant costs and non-reproducibly. If important metadata is not immediately secured and stored in the field, the quality and re-usability of the resulting data will be diminished.

At the Helmholtz Coastal Data Center (HCDC) a metadata and data workflow for biogeochemical data has been developed over the last couple of years to ensure the quality and richness of metadata and enable that the final data product will be FAIR. It automates and standardizes the data transfer from the campaign planning stage, through sample collection in the field, analysis and quality control to the storage into databases and the publication in repositories.

Prior to any sampling campaign, the scientists are equipped with a customized app on a tablet that enables them to record relevant metadata information, such as the date and time of sampling, the involved scientists and the type of sample collected. Each sample and station already receives a unique identifier at this stage. The location is directly retrieved from a high-accuracy GNSS receiver connected to the tablet. This metadata is transmitted via mobile data transfer to the institution's cloud storage.

After the campaign, the metadata is quality checked by the field scientists and the data curator and stored in a relational database. Once the samples are analyzed in the lab, the data is imported into the database and connected to the corresponding metadata using a template. Data DOIs are registered for finalized datasets in close collaboration with the World Data Center PANGAEA. The data sets are discoverable through their DOIs as well as through the HCDC data portal and the API of the metadata catalogue service.

This workflow is well established within the institute, but is still in the process of being refined and becoming more sophisticated and FAIRer. For example, an automated assignment of International Geo Sample Numbers (IGSN) for all samples is currently being planned.