

EGU21-15037

<https://doi.org/10.5194/egusphere-egu21-15037>

EGU General Assembly 2021

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



How to turn kilos of mud into megabytes of data? 10 years of efforts in curating lake sediment cores and their associated results

Fabien Arnaud¹, Cécile Pignol¹, Bruno Galabertier¹, Xavier Crosta², Isabelle Billy², Elodie Godinho³, Karim Bernardet³, Pierre Sabatier¹, Anne-Lise Develle¹, Rosalie Bruel^{1,4}, Julien Penguen⁵, Pascal Calvat⁵, Pierre Stéphan⁶, and Mathias Rouan⁶

¹Environment Dynamics and Territories of the Mountain (EDYTEM), Université Savoie Mont-Blanc, CNRS, Chambéry, France (fabien.arnaud@univ-savoie.fr)

²Environnements et paléoenvironnements océaniques et continentaux (EPOC), Université de Bordeaux, CNRS, Bordeaux, France

³Division Technique de l'Institut des Sciences de l'Univers (DT-INSU), CNRS, La Seyne-sur-Mer, France

⁴Rubenstein Ecosystem Science Laboratory, University of Vermont, Burlington, VT, USA

⁵Observatoire Aquitain des Sciences de l'Univers (OASU), Université de Bordeaux, CNRS, Bordeaux, France

⁶Littoral, Environnement, Télédétection, Géomatique (LETG), Université de Bretagne Occidentale, Brest, France

Here we present a series of connected efforts aiming at curating sediment cores and their related data. Far to be isolated, these efforts were conducted within national structured projects and led to the development of digital solutions and good practices in-line with international standards and practices.

Our efforts aimed at ensuring FAIR-compatible practices (Plomp, 2020; Wilkinson et al., 2016) throughout the life cycle of sediment cores, from fieldwork to published data. We adopted a step-by-step, bottom-up strategy to formalize a dataflow, mirroring our workflow. We hence created a fieldwork mobile application (CoreBook) to gather information during coring operations and inject them toward the French national virtual core repository “Cyber-Carothèque Nationale” (CCN). At this stage, the allocation of an international persistent unique identifier was crucial and we naturally chose the IGSN.

Beyond the traceability of samples, the curation of analysis data remains challenging. Most international repository (e.g. NOAA palaeo-data, PANGAEA) have taken the problem from the top by offering facilities to display published dataset with persistent unique identifier (DOI). Yet, those data are only a fraction of the gross amount of acquired data. Moreover, those repositories have very low requirements when it comes to the preservation and display of metadata, in particular analytical parameters, but also fieldwork data which are essential for data reusability. Finally, these repositories do not permit to get a synoptic view on the several strata of analyses that have been conducted on the same core through different research programs and publications. A partial solution is proposed by the eLTER metadata standard DEIMS, which offers a discovery interface of rich metadata. In order to bridge the gap between generalist data repositories and samples display systems (such as CCN, but also IMLGS, to cite an international system), we developed a

data repository and visualizer dedicated to the re-use of lake sediment cores, samples and sampling locations (ROZA Retro-Observatory of the Zone Atelier). This system is still a prototype but opens yet interesting perspectives.

Finally, the digital evolution of science allows the worldwide diffusion of data processing freewares. In that framework, we developed "Serac" an open-source R package to establish radionuclide-based age models following the most common sedimentation hypotheses (serac.). By implementing within this R package the input of a rich metadata file that gathers links to IGSN and other quality metadata, we are linking fieldwork metadata, the physical storage of the core and the analytical metadata. Indeed, Serac also stores data processing procedure in a standardized way.. We hence think that the development of such softwares could help in the spreading of good practices in data curation and favour the use of unique identifiers.

By tackling all aspects of data creation and curation throughout a lake sediment core life cycle, we are now able to propose a theoretical model of data curation for this particular type of sample that could serve as the sole for further developments of integrated data curation systems.