

EGU21-9401

<https://doi.org/10.5194/egusphere-egu21-9401>

EGU General Assembly 2021

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



Harmonizing heterogeneous multi-proxy data from Arctic lake sediment records

Gregor Pfalz^{1,2,3}, Bernhard Diekmann^{1,2}, Johann-Christoph Freytag^{3,4}, and Boris K. Biskaborn^{1,2}

¹Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Potsdam, Germany (gregor.pfalz@awi.de)

²Institute of Geosciences, University of Potsdam, Potsdam, Germany

³Einstein Center Digital Future, Berlin, Germany

⁴Department of Computer Science, Humboldt-Universität zu Berlin, Berlin, Germany

Lake systems play a central role in broadening our knowledge about future trends in the Arctic, as their sediments store information on interactions between climate change, lake ontogeny, external abiotic sediment input, and biodiversity changes. In order to make reliable statements about future lake trajectories, we need sound multi-proxy data from different lakes across the Arctic. Various studies using data from repositories already showed the effectiveness of multi-proxy, multi-site investigations (e.g., Kaufman et al., 2020; PAGES 2k Consortium, 2017). However, there are still datasets from past coring expeditions to Arctic lake systems that are neither included in any of these repositories nor subject to any particular standard. When working with such data from heterogeneous sources, we face the challenge of dealing with data of different format, type, and structure. It is therefore necessary to transform such data into a uniform format to ensure semantic and syntactic comparability. In this talk, we present an interdisciplinary approach by transforming research data from different lake sediment cores into a coherent framework. Our approach adapts methods from the database field, such as developing entity-relationship (ER) diagrams, to understand the conceptual structure of the data independently of the source. Based on this knowledge, we developed a conceptual data model that allows scientists to integrate heterogeneous data into a common database. During the talk, we present further steps to prepare datasets for multi-site statistical investigation. To test our approach, we compiled and transformed a collection of published and unpublished paleolimnological data of Arctic lake systems into our proposed format. Additionally, we show our results from conducting a comparative analysis on a set of acquired data, hereby focusing on comparing total organic carbon and bromine content. We conclude that our harmonized dataset enables numerical inter-proxy and inter-lake comparison despite strong initial heterogeneity.

[1] D. S. Kaufman et al., "A global database of Holocene paleotemperature records," *Sci. Data*, vol. 7, no. 115, pp. 1–34, 2020.

[2] PAGES 2k Consortium, "A global multiproxy database for temperature reconstructions of the Common Era," *Sci. Data*, vol. 4, no. 170088, pp. 1–33, 2017.

