EGU22-11766
https://doi.org/10.5194/egusphere-egu22-11766
EGU General Assembly 2022

Implementing semantic data management for bridging empirical and simulative approaches in marine biogeochemistry

Alexander Schlemmer[1,3], Julian Merder[2], Thorsten Dittmar[2], Ulrike Feudel[2], Bernd Blasius[2], Stefan Luther[1], Ulrich Parlitz[1], Jan Freund[2], and Sinikka T. Lennartz[2]

[1]Research Group Biomedical Physics, Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany (alexander.schlemmer@ds.mpg.de)

[2]Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg, Germany (sinikka.lennartz@uni-oldenburg.de)

[3]IndiScale GmbH, Göttingen, Germany (a.schlemmer@indiscale.com)

CaosDB is a flexible semantic research data management system, released as open source software. Its versatile data model and data integration toolkit allows for applications in complex and very heterogeneous scientific workflows and different scientific domains. Successful implementations include biomedical physics [1] and glaciology [2]. Here, we present a recent implementation of a use case in marine biogeochemistry which has a special focus on bridging between experimental work and numerical ocean modelling. CaosDB is used to store, index and link data during different stages of research on the marine carbon cycle: Data from experiments and field campaigns is integrated and mapped onto semantic data structures. This data is then linked to data from numerical ocean simulations. The ocean model, here with a specific focus on natural marine carbon sequestration of dissolved organic carbon (DOC), uses the georeferenced data to evaluate model performance. By simultaneously linking empirical data and the sampled model parameter space together with the model output, CaosDB enhances the efficiency of model development. In the current implementation simulated data is linked to georeferenced DOC concentration data. We plan to expand it to complex data sets including thousands of dissolved organic matter molecular formulae and metagenomes of pelagic microbial communities. The combined management of these heterogeneous data structures with semantic models allows us to perform complex searches and seamlessly connect to automated data analysis pipelines.

[1] Fitschen, T.; Schlemmer, A.; Hornung, D.; tom Wörden, H.; Parlitz, U.; Luther, S. CaosDB—Research Data Management for Complex, Changing, and Automated Research Workflows. Data 2019, 4, 83. https://doi.org/10.3390/data4020083
[2] Schlemmer, A.; tom Wörden, H.; Freitag, J.; Fitschen, T.; Kerch, J.; Schlomann, Y.; ... & Luther, S. Evaluation of the semantic research data management system CaosDB in glaciology. deRSE 2019. https://doi.org/10.5446/42478