Geophysical Research Abstracts
Vol. 21, EGU2019-15736, 2019
EGU General Assembly 2019

The LUE scientific data base for storing heterogeneous earth science data

Derek Karssenberg, Kor de Jong, and Oliver Schmitz
Department of Physical Geography, Faculty of Geosciences, Utrecht University, Utrecht, Netherlands (d.karssenberg@uu.nl)

Earth science data from observations or models may refer to natural phenomena that are spatially and temporally continuous across the area of interest, for instance groundwater, or bounded in space and time, for instance tree crowns. Continuous phenomena can be stored and manipulated in an efficient manner by current Data Cubes. However, phenomena bounded in space and time have particular properties that do not allow them to be stored in Data Cubes mainly because they require representation of multiple (sometimes mobile) objects that each exist in a small subdomain of the space-time domain of interest, while these subdomains of objects may overlap in space and time such as interleaving branches due to tree crown growth. This cannot be solved by simply adding an additional n+1 domain to a rectangular Data Cube and needs a rigorously new approach that integrates representation of fields and objects in a single data model. Our research team is following this approach by the development of a conceptual data model (de Bakker, 2017) for field-object representation which is implemented as a physical data model using HDF5 and C++, providing a Python API to communicate with other data formats and simulation models. The resulting open source LUE data model (https://lue.readthedocs.io) explicitly stores and separates domain information, i.e. where phenomena exist in the space-time domain, and property information, i.e. what attribute value the phenomenon has at a particular space-time location, for a particular object. The LUE data model aims to be a super-set of existing data models for storing earth science data. Notable extensions are support for multiple spatio-temporal objects, time domains, objects linked to multiple space and time domains, and relations between objects. Additionally, we want LUE datasets to be usable in a high-performance computing context. Therefore, LUE datasets will support parallel I/O and allow storage of as much data as is supported by the filesystem. In our presentation we illustrate the concepts of the data model with working examples from storage of individual tree crown data, remotely sensed image data, and data of travel time to food outlets for simulation of human environmental exposures.

References

de Bakker, M. P., de Jong, K., Schmitz, O., & Karssenberg, D. (2017). Design and demonstration of a data model to integrate agent-based and field-based modelling. Environmental Modelling & Software, 89, 172-189. https://doi.org/10.1016/j.envsoft.2016.11.016