The LUE scientific database: a datacube for heterogeneous earth science data

Derek Karssenberg, Kor de Jong, Oliver Schmitz, Fac. of Geosciences, Utrecht University, the Netherlands, d.karssenberg@uu.nl / k.dejong1@uu.nl / o.schmitz@uu.nl

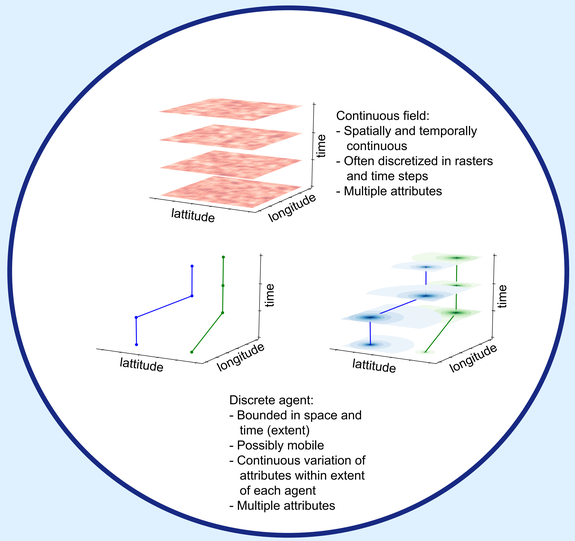
Problem

- Current data models either focus on representation of continuous fields (e.g. data cubes) or on representation of discrete agents (e.g., data models in agent-based modelling software)
- Numerical modellers need to use multiple existing data models or create ad-hoc data storage approaches

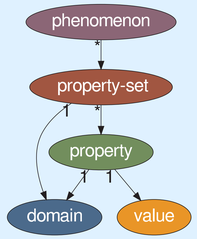
Proposed solution

- Single, universal data model for storing continuous fields and discrete agents
- API and modelling framework for numerical simulation with continuous fields and discrete agents

Integrated Solution



Conceptual data model



Phenomenon: The phenomenon to be stored (e.g., birds, reflectance, groundwater)

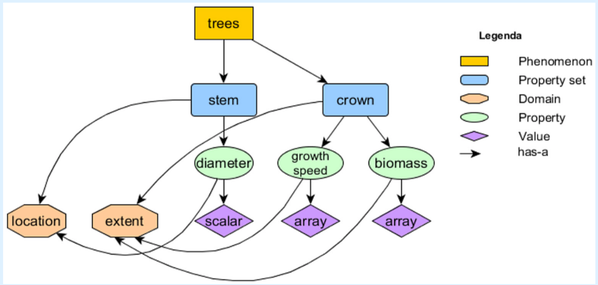
Property-set: Collection of properties sharing the same spatial and temporal domain

Domain: Information about where and when a property exists (e.g. location of birds, land surface, subsoil volume)

Property: Attribute (e.g. weight of bird, elevation of the land surface, groundwater pollution level)

Value: Magnitude of a property

Item: Identifies an individual (discrete agent)



Example: representation of a set of trees (discrete agents).

Physical Data Model

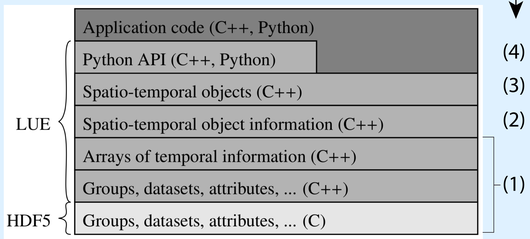
Stack of four layers of software:

(1) Universal representation of spatio-temporal data

(2) Use of (1) to represent fields and agents (spatial location/extent, temporal location/duration, attribute values)

(3) Use of (2) to represent components of the conceptual data model (Phenomenon, Property-set, Domain, Property, Value)

(4) Python API



Characteristics:
- All model data in a single, portable, file
- Supports parallel I/O
- Open source software (GitHub)

Additional information

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

de Jong, K. & Karssenberg, D. (in prep.). Design and implementation of a physical data model for simulated spatio-temporal objects. To be submitted to Environmental Modelling & Software.

http://www.pcraster.eu

https://github.com/pcraster/lue

Modelling Framework

One application of LUE is for forward simulation of integrated field-based and agent-based models. Current numerical simulation software mostly requires the modeller to define an explicit iteration over the discrete agents:

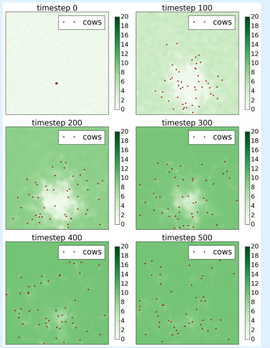
```
agents = [agent definition and instantation]
for agent in agents:
  agent.c = agent.a + agent.b
```

To avoid this low-level of implementation and to support integration simulation with continuous fields, we follow a map algebra representation of operations:

```
phenomenon.c = phenomenon.a + phenomenon.b
```

where `phenomenon` can be either a single continuous field with properties `a`, `b`, and `c` (like in standard map algebra) or multiple discrete agents, where each agent has properties `a`, `b`, and `c`.

In addition, the modelling framework provides built-in flow of control for time steps as well as Monte Carlo simulation. The prototype modelling framework runs in Python and enables reading and writing data sets from LUE.



Example simulation: dots, mobile and grazing cows (agents), green shading, biomass (continuous field).

Photography encouraged