Six soil data challenges and four ways semantic web technologies can help

Victoria Janes Bassett[1], Jessica Davies[1], Graham Dean[2], Ross Towe[2], Vatsala Nundloll[2], Gordon Blair[2]

[1] Pentland Centre for Sustainability in Business, Lancaster Environment Centre, University of Lancaster, UK.

[2] School of Computing and Communications, Lancaster University, UK.

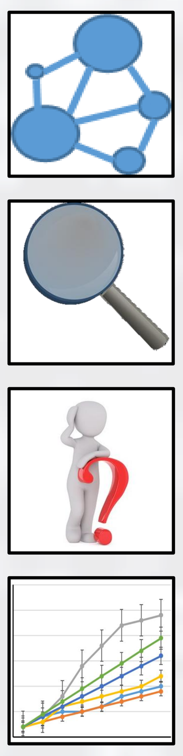
Lancaster University

Modelling Data Needs

Soil modelling often requires large amounts of data from a diverse range of sources including (but not limited to) soil chemistry, land-use, atmospheric deposition, meteorological and climatological. This data is used not just for model input, but also in model calibration and validation. Such sources of data are often sparse and hard to find, making searching for model data a time consuming process that often misses out key sources. Here we discuss the potential for semantic web technology to address the data needs of soil, and other environmental modellers.

Semantic Web Technology

The semantic web aims to provide support for an interconnected web of data. The data are linked with metadata, which through a querying process is interrogated. Here we will discuss four high-level themes that underpin the semantic web and how these can be applied to soil model user needs:

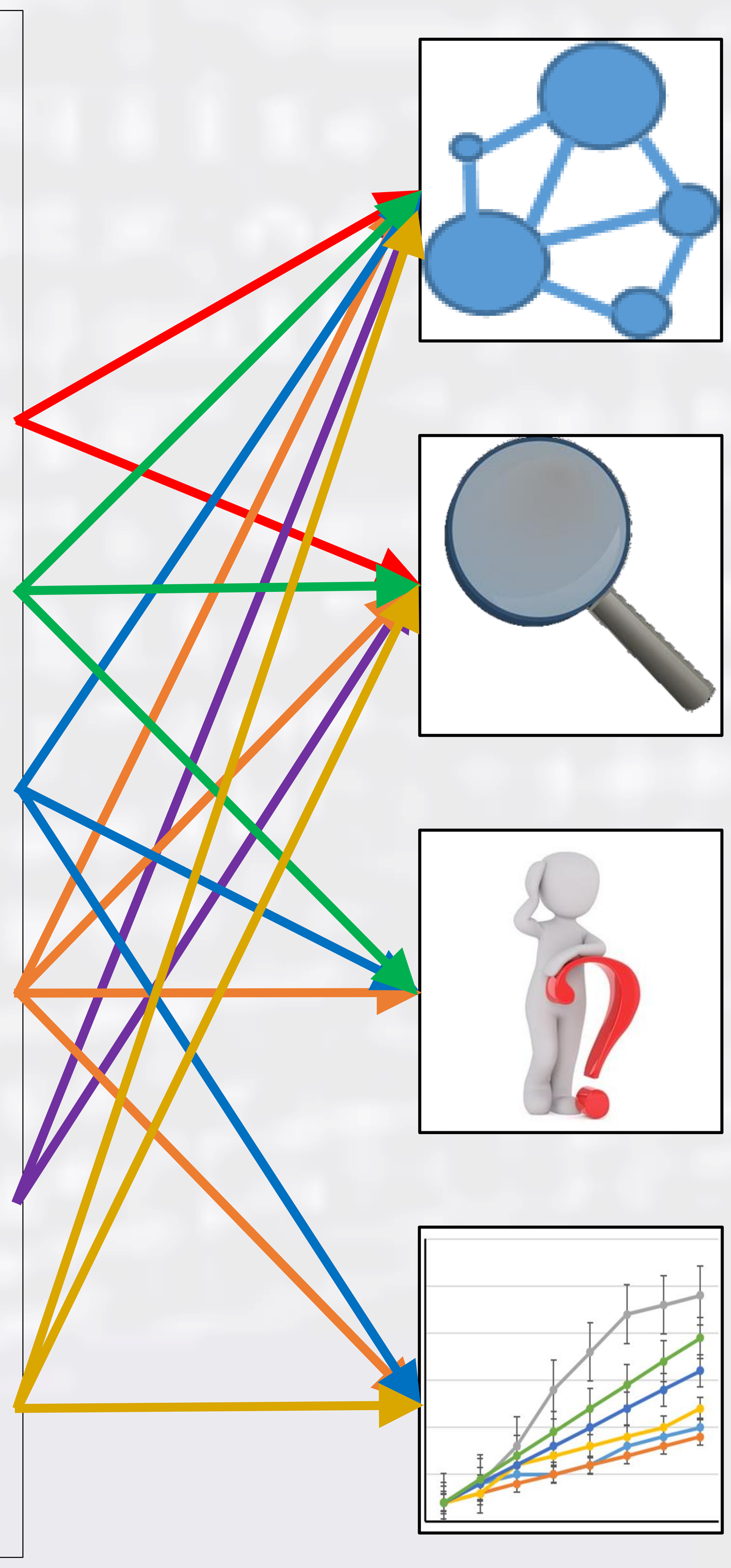


- **Semantic integration of data and models:** Semantic integration allows us to make relationships between different datasets and models explicit, making them suitable for machine processing.
- **Data and model discovery:** Making resources such as suitable datasets for specific spatial locations discoverable and usable directly on the web.
- **Reasoning:** Across diverse heterogeneous datasets with differing spatial and temporal characteristics.
- **Uncertainty management:** Different types of uncertainty need to be addressed in an explicit manner.

Data Needs

Through discussion with soil system modellers, we identified six classes of problems relating to data discovery and utilisation. Many of these issues are general, and will be relevant to a wide range of environmental modelling disciplines.

- **Ease of measurement:** if not automated this can be labour intensive, time consuming and costly, posing a barrier to data collection. **Example: Soil Chemistry**
- **Temporal scale of interest:** data sets over long timescales are often difficult to find due to lack of consistent support for long term data collection. Archiving issues can also create a barrier. **Example: Change in soil C**
- **Diversity in methods of observation**: where multiple methods exist to measure a variable there are issues of inconsistency within datasets. **Example: Soil P**
- **Level of regulation:** with no national/international regulation of data collection there is often no data collation facility. Regulation can also set standardised practices of collection method and units. **Counter example: CEH National River Flow Archive**
- **Frequency of use:** infrequently used data provides little demand (and incentive) to make data easily available, creating a positive feedback; data that is hard to find is used less. **Counter example: CEH National River Flow Archive**
- **Extreme value error:** issues with the level of accuracy associated with extreme values. Linked to high uncertainty due to fewer observations in extremities. **Example: High flow river sediment data**

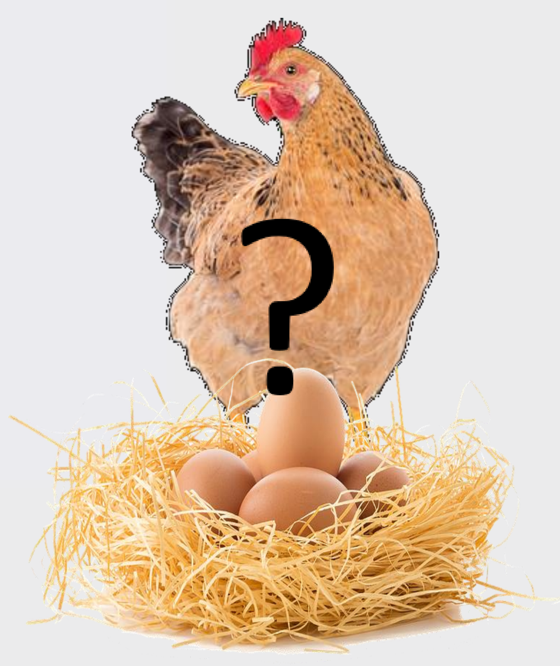


What needs to happen before semantic web technologies can be applied?

There are several steps that soil modellers need to take before the semantic web can be put in to practice with existing datasets. These include:

- Describe context of data that exists (develop metadata)
- Develop ontologies of data (vocabulary)
- Upload data to semantic database (such as Virtuoso or GraphDB)
- Develop ways of finding data repositories (service discovery)

Metadata provides information about other data, making it accessible. Service discovery means the data can be found. Each is useless without the other. So where do we start?



Contact:
Dr Victoria Janes Bassett
v.janes@lancaster.ac.uk