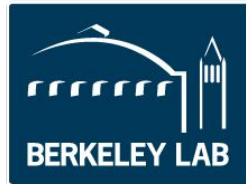


# Semantic Annotations and Data Integration in Ecology

Anne E Thessen, Mark Schildhauer, Meghan A Balk, Chris Mungall

[annethessen@gmail.com](mailto:annethessen@gmail.com)

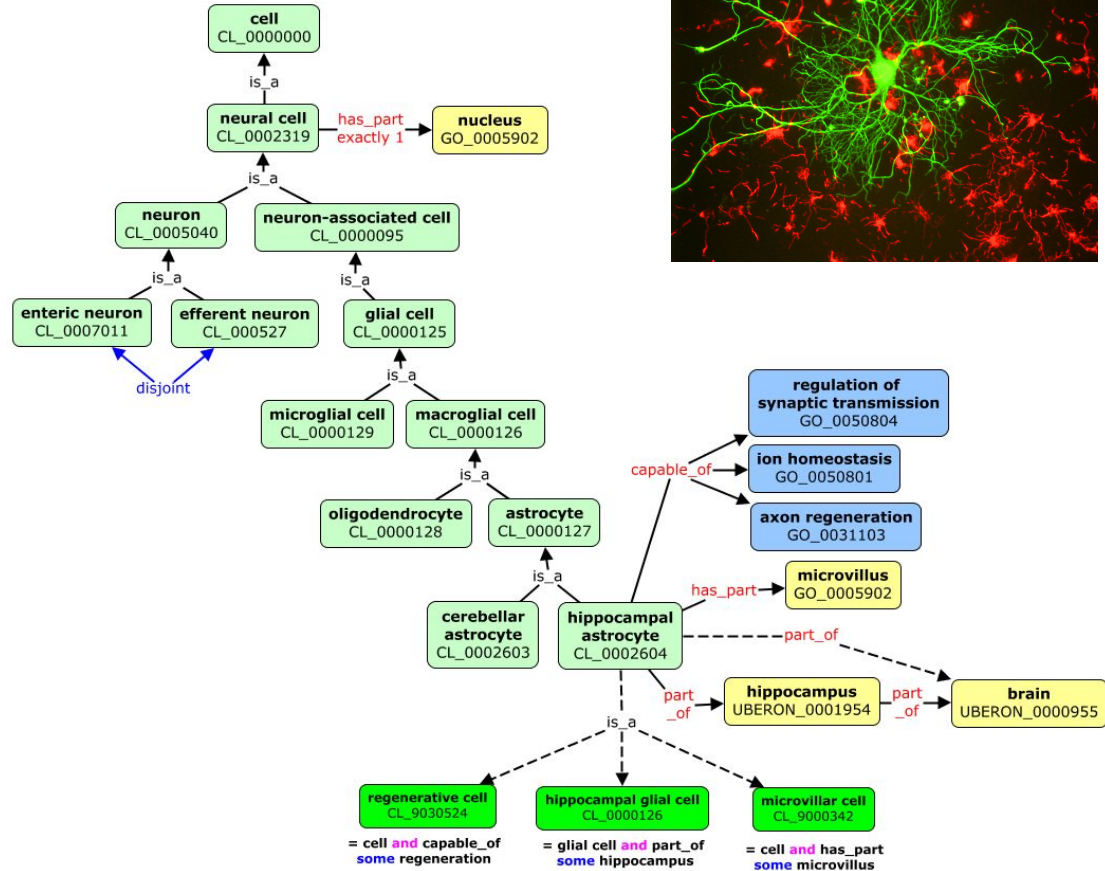
[@diatomsRcool](https://twitter.com/diatomsRcool) 



# What is semantic technology?

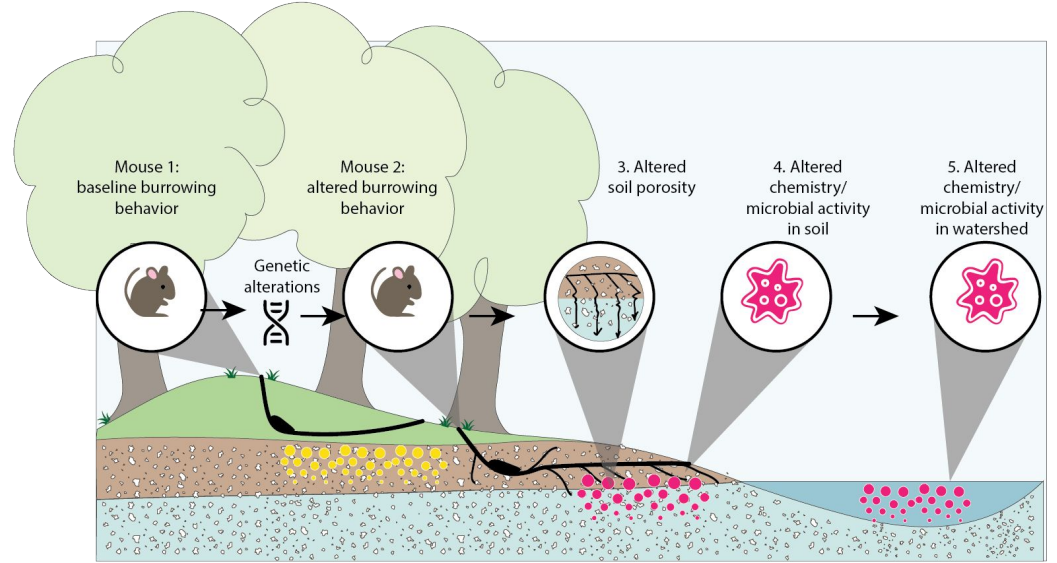
## KEY FEATURES:

- Terms are defined
- Semantics - relationships between terms are defined, allowing logical inference and sophisticated data queries
- Terms are arranged in a hierarchy
- Expressed in a knowledge representation language such as RDFS, OBO, or OWL



# The Vision for Ecology

- Why is the water quality poor?
- The answer requires multiple, heterogeneous data sets from different disciplines
- Generate hypotheses across discipline, taxon, and scale
- Link data and knowledge in unprecedented ways



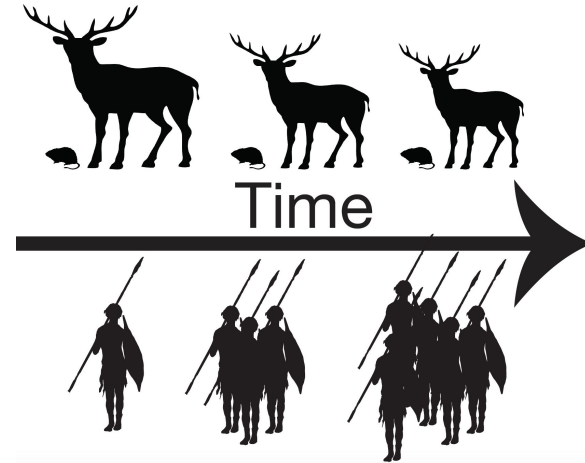
# Using Semantic Technology to Integrate Data in Functional Trait Resource for Environmental Studies (FuTRES)

## Goals

- Collect, store, integrate, and share data at the individual or specimen level and higher levels without loss of information.
- Generate interoperable, accessible data
- Engage the community of functional trait researchers to liberate legacy data.

## Approach

- Expand existing vertebrate phenotype ontologies (UBERON and OBA) to encompass paleo- and neontological functional traits.
  - Create modules for trait terms
- Create and promote standardized templates for vertebrate trait data.
- Provide tools for data submission (GEOME) and discovery (FuTRES portal).



### Case Study:

Evaluating the impact of human hunting pressure on prey size as a function of population size versus other regional and global pressures such as climate change.



# Using Semantic Technology to Integrate Data in the Arctic Data Center

## Goals

- Make Arctic data more FAIR: easier to Find, easier to Reuse, and easier to Integrate
- Use Semantics to provide logics-based, standards-based ways to organize, describe, and inter-relate metadata terms in EML

## Approach

- Link metadata descriptors in EML to ontology terms: *semantic annotation*
- Environment Ontology (EnvO)
  - <http://environmentontology.org>
- Ecosystem Measurement Ontology (ECSO)
  - <https://purl.dataone.org/odo/ECSO>

## What is the Arctic Data Center?

- Primary data and software repository for the Arctic Section of NSF's Office of Polar Programs. <http://arcticdata.io>; over 640,000 datasets, many ecological measurements. Still growing...
- All data described in EML, Ecological Metadata Language, and stored as XML documents
- Descriptors, especially of Measurements ("Variables", "Attributes") often not well described
- **Semantic annotation**: Metadata linked to terms in ontologies enable-- finding synonyms, differentiating homonyms, and exploring hierarchies



*litter*



*litter*



More info:

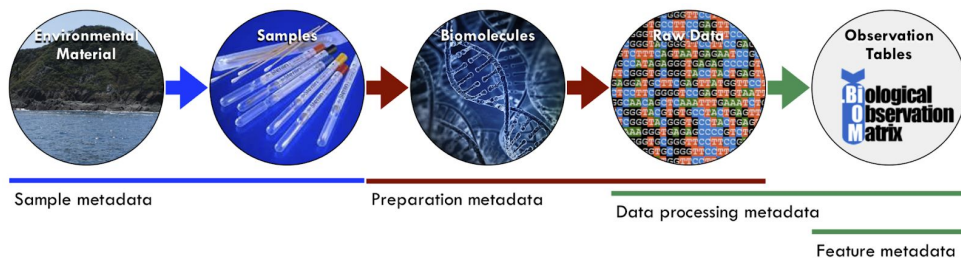
<https://arcticdata.io/semantic-annotations/>

<https://eml.ecoinformatics.org/semantic-annotation-primer.html>

# Using Semantic Technology to Integrate Data in the National Microbiome Data Collaborative

## Goal

- Make multi-omics microbiome data FAIR
  - Environments
  - Metagenomes
  - Metatranscriptomes
  - Metabolomics
  - Metaproteomics
- Leverage existing ontologies and standards
- Enable discovery in microbiome science



## Approach

- Create schema in LinkML
  - Integrates with traditional infrastructure (JSON/MongoDB)
  - Ontology-enabled
- Describe environments using ENVO and MixS
- Create mappings to other vocabularies



**nmDC**

National Microbiome  
Data Collaborative

# Conclusions

- Ontologies transform human knowledge into a computable format
- Knowledge graphs, supported by ontologies, integrate heterogeneous data into a queryable structure
- Annotating data with ontology terms can make the data more FAIR
- FuTRES, the Arctic Data Center, and the National Microbiome Data Collaborative are already using ontologies to make data easier to find, integrate, and reuse

annethessen@gmail.com  
<https://tislabs.org/>



@diatomsRcool 