

Supplementary Figures

Figure 1: Semantic extensions to Metacat.

Semantic extensions (highlighted in blue) to the Metacat data and metadata repository support improved precision and recall in searches for scientific data sets.

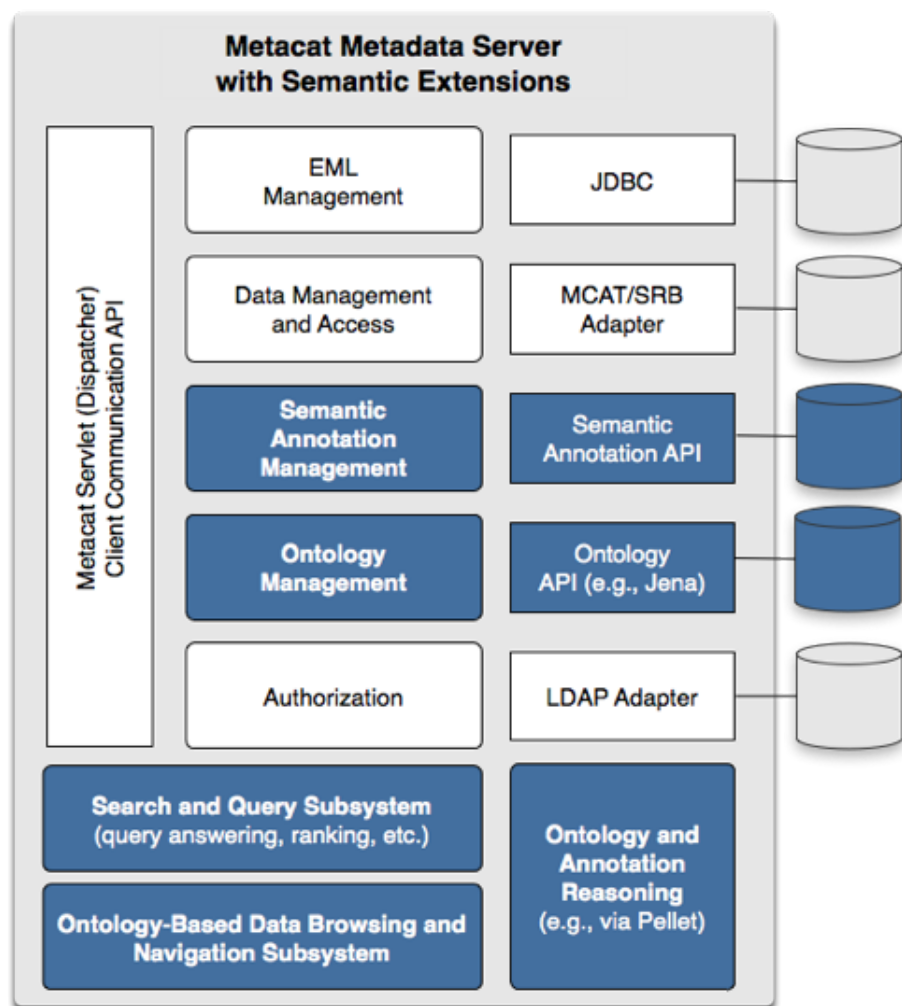


Figure 1

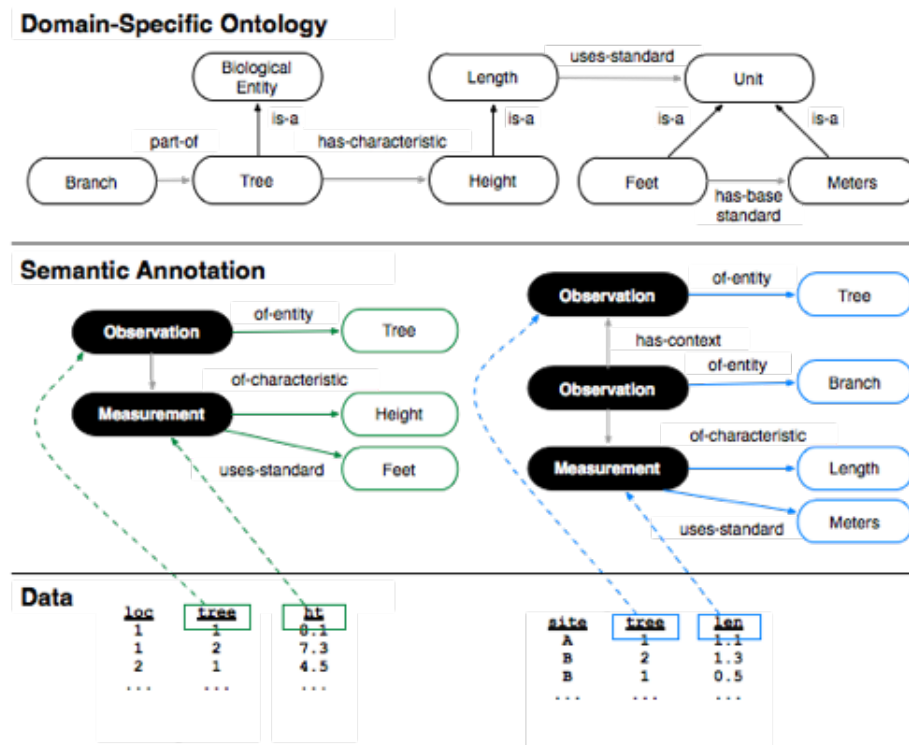


Figure 3

Figure 4. OBOE Annotation.

Partial OBOE semantic annotation for Kelp sampling data. Shaded nodes represent ontological concepts; rectangular nodes are data table attributes mapped to OBOE measurement characteristics.

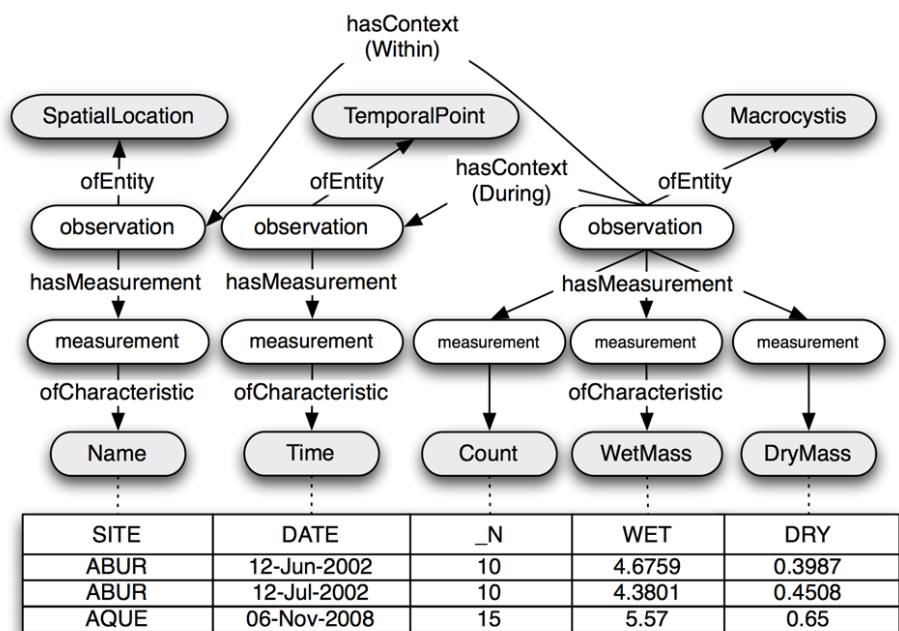


Figure 4

Figure 5. Semantics in Morpho.

Morpho metadata editor with Semantic plugin. The fill-in-the-blank interface uses natural language descriptions for intuitive editing. A searchable, hierarchical browser is used to select concepts from domain-specific ontologies.

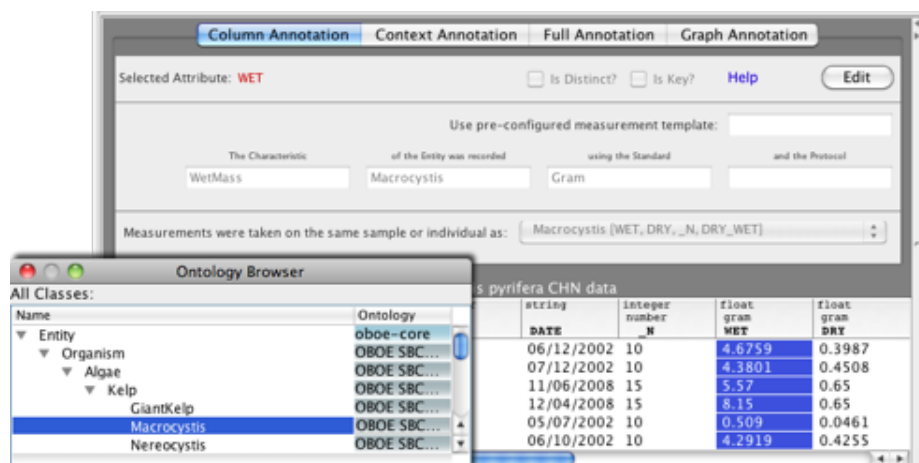


Figure 5

Figure 6. Semantic Search Web UI.

The first prototype of the semantic data query web interface. Data packages containing observations of Kelp Wet Mass less than or equal to 5 [grams] are returned. Additional search options and compound query criteria can be specified within the other tabs. Matches can be saved in the data cart for later exploration.

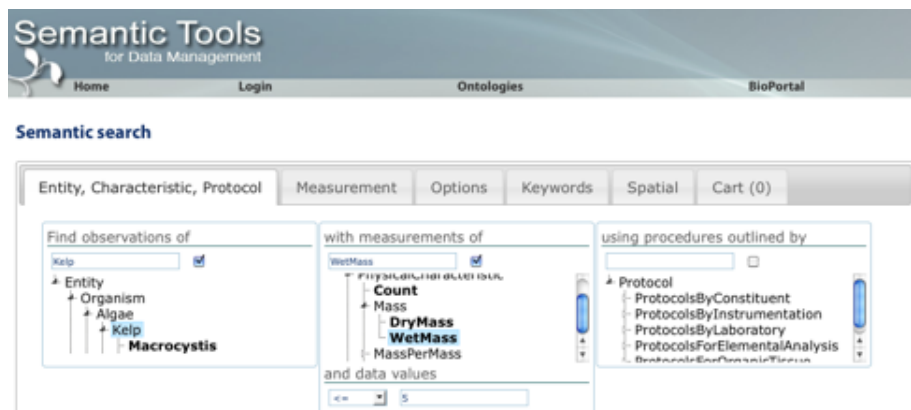


Figure 6

Figure 7. Semantic Data Integration.

Integration query across multiple data packages (A, B). Annotations (C, D) determine semantically equivalent data attributes contained in the data objects (E, F). Attributes plot and site are considered equivalent measurements of the characteristic Location; attributes weight and wt both map to the same characteristic Mass. The Semantic Mediation API utilizes the Data Manager Library to load and query the source data informed by semantic similarities between the structurally disparate data attributes.

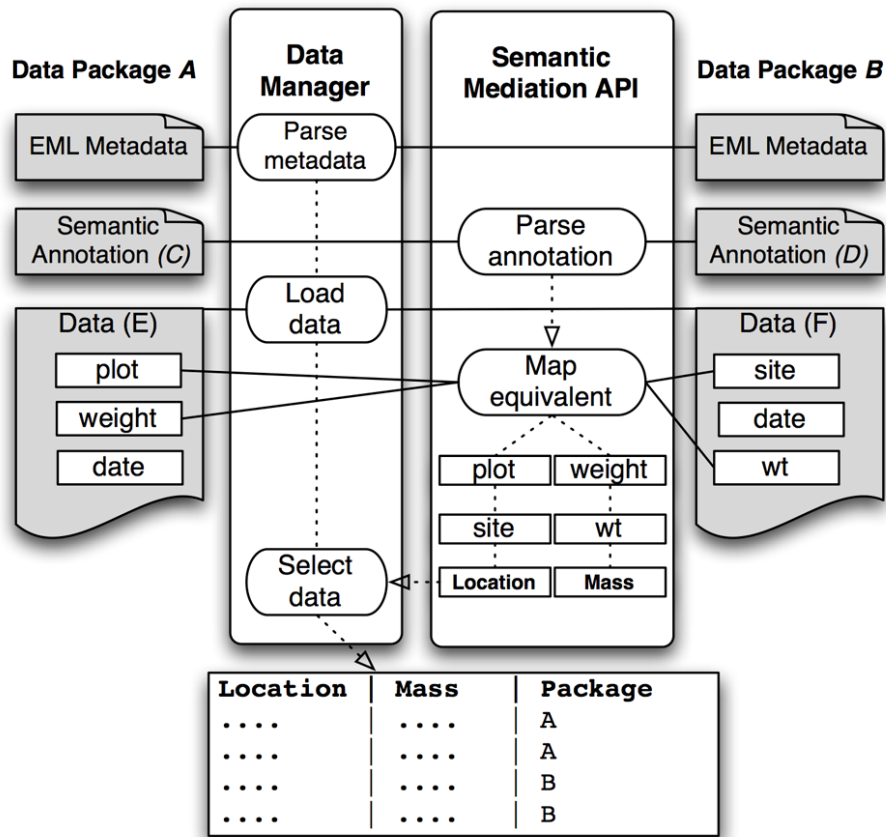


Figure 7

Figure 8. Annotation Components.

The final system produced by Semtools contains 7 major components, including 1) an Ontology repository for storing OWL ontologies, 2) an Annotation generator for creating annotations that link ontology concepts to data, 3) an Annotation Store for storing those annotations, 4) an Indexer that processes annotations using reasoning from an 5) Triple Store that is used to process the OWL annotations. The final indexed semantic content is stored in a 6) SOLR Index that provides rapid searching across the semantic concepts and is used to drive the 7) Web User Interface which provides the ability for research scientists to explore and utilize the semantic content.



Figure 8

Figure 9. Search User Interface.

The search system provides an easy-to-understand dropdown of Measurement Types, allowing scientists to precisely select the kinds of measurements they are looking for and integrate those with other search criteria.

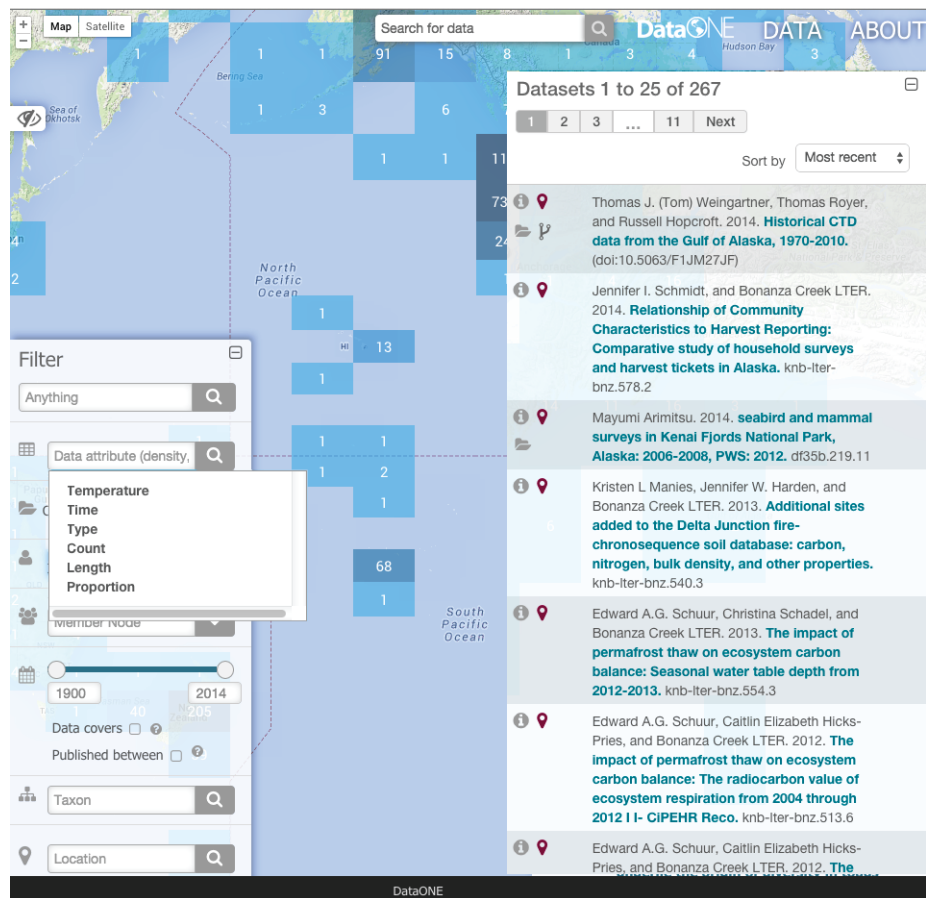


Figure 9

Figure 10. Annotation User Interface.

The Annotation user interface is integrated into the metadata display for a data set, allowing a researcher to view the annotations that are associated with a particular column in a data set, to view comments that people left with the annotation, and to see the definition of the annotated terms. This user interface allows the science community to have a structured conversation about the semantics of measurements and come to consensus on how to model data semantics.

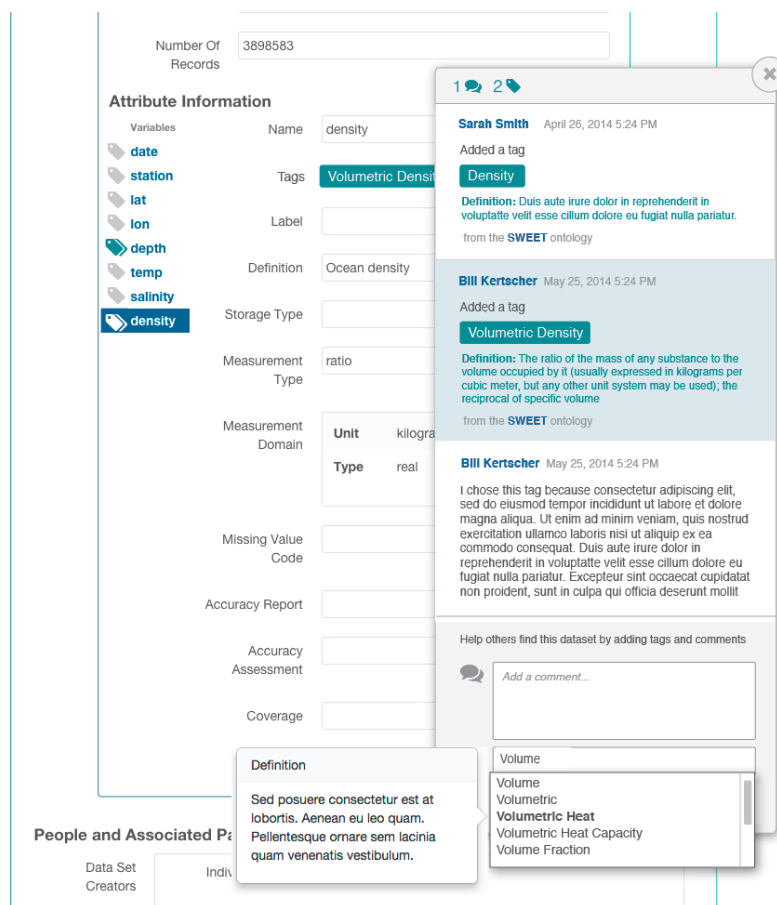


Figure 10

Figure 11. Open Annotation Model.

The final annotation model for Semtools builds on the Open Annotation and PROV specifications from the W3C. This model utilizes Open Annotation to link annotations that are expressed in OBOE or other models to data and metadata using selectors that target specific structures within the data and metadata. This model is flexible enough to accommodate multiple heterogeneous data and metadata formats while still providing enough structure to predictably interpret the annotations.

