

SemantEco Annotator

Patrice Seyed^{1,2} (seyeda2@rpi.edu), Katherine Chastain² (chastk@rpi.edu), Brendan Ashby² (ashbyb@rpi.edu),

Timothy Lebo² (lebot@rpi.edu), Evan Patton² (pattoe@rpi.edu), Deborah L. McGuinness² (dlm@cs.rpi.edu)

(¹DataONE, University of New Mexico, 1 University Boulevard, Albuquerque, NM 87131) (² Rensselaer Polytechnic Institute 110 8th St., Troy, NY, 12180 United States)

The Idea:

- Utilize community standard vocabularies to describe data and metadata.
- Provide a broadly useable tool that makes it easy to annotate data for reuse, understandability, and integration.
- Make linked data conversion easy and accessible, even to non-experts.

Vocabularies:

- Employ existing ontologies, including OBO-E (Extensible Observation Ontology) and W3C PROV-O (Provenance Ontology).
- The annotator provides a large, searchable index of ontologies to use and re-use.

Bundling:

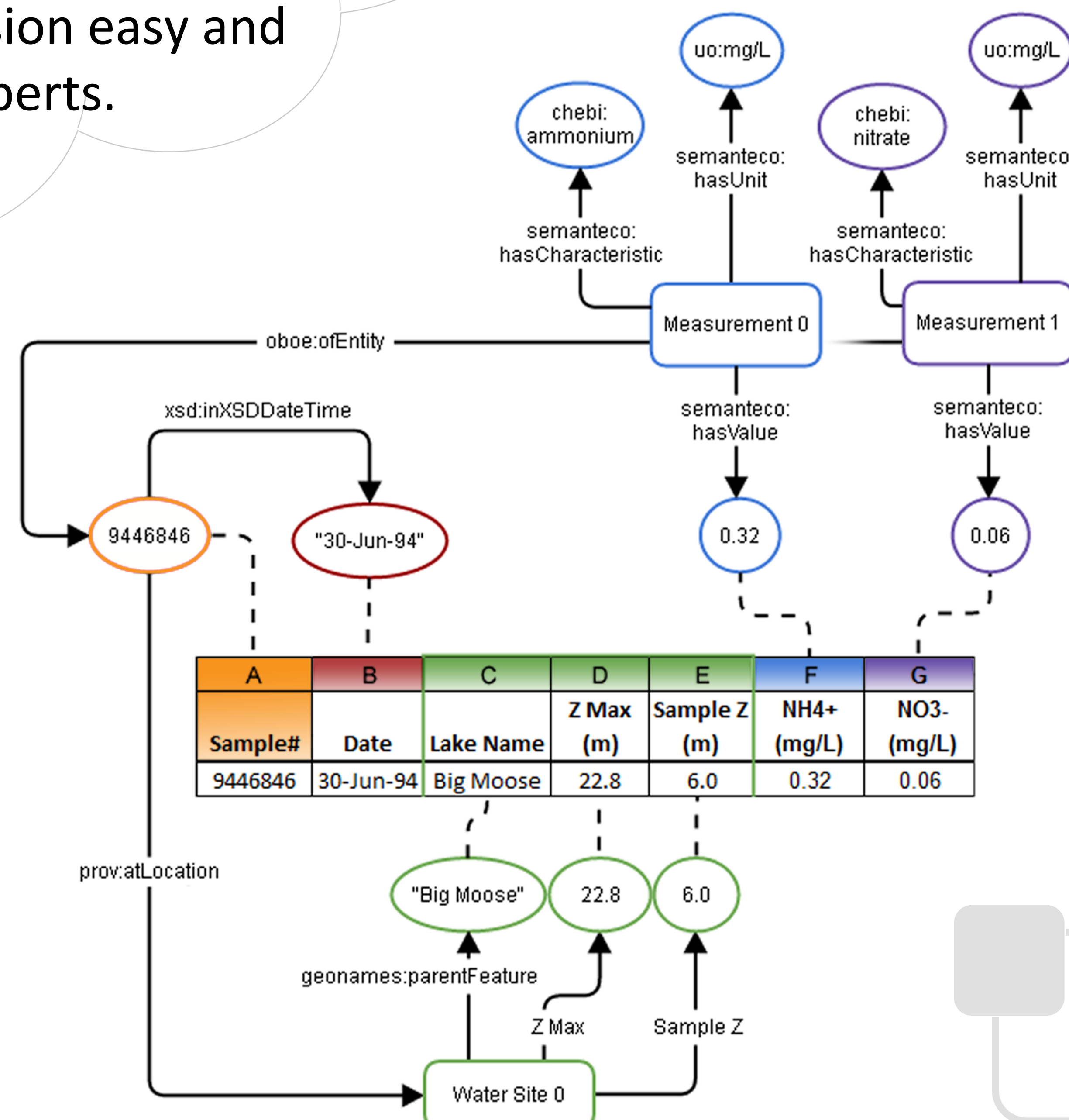
- Several columns together represent aspects of one entity.
- Z Max and Sample Z describe maximum lake depth and sample depth, respectively. Coupled with the Lake Name, these three columns describe a “Water Site” for each Sample.
- The researcher knows this entity exists in the data. The annotator allows him or her to explicate it in a machine-readable format.

Future Work:

- Allow incorporation of the annotator in a variety of semantic applications.
- Provide templates of expected relations for additional conversion guidance.
- Refine ontology search and expand available index.

Standard Conversion:

- One column typically represents the subject of triples in that row.
- Other columns relate to that row subject, either directly or indirectly.



Cell-Based Conversion:

- When the subject of a triple is not the subject of the row.
- For example, to save space, Units and Characteristics of the Measurements appear in the column headers.
- We can represent information that is contained in places other than the data table itself.

Current Annotator Features:

- Graphical, browser-based drag-and-drop interface – work directly on the data table.
- Methods for representing many types of relationships between columns.
- Visual feedback and guidance to assist newcomers to linked data.
- The key is use, reuse, and sharing of data, so the SemantEco Annotator will be open-source.



Scan to visit our project page:

<http://tw.rpi.edu/web/project/SemantEcoAnnotator>

Acknowledgments:

Our thanks to DataONE for post-doctorate, intern, and working group support. Also thanks to the Tetherless World for research support, and semantic tools and resources. Also thanks to the Darrin Freshwater Institute for sharing the water quality data from which we derived our example conversion.

Glossary:

RPI – Rensselaer Polytechnic Institute
TWC – Tetherless World Constellation at Rensselaer Polytechnic Institute
DataONE – Data Observation Network for Earth

Sponsors:

