Protégé Extensions for Scientist-Oriented Modeling of Observation and Measurement Semantics (Application Note)

Shawn Bowers\textsuperscript{1}, Wesley Saunders\textsuperscript{1}, Margaret O’Brien\textsuperscript{2}

\textsuperscript{1} Dept. of Computer Science, Gonzaga University
\textsuperscript{2} Marine Science Institute, UC Santa Barbara
Observational data

Ecological data is largely focused on “observations”

• Data sets stored in tables (e.g., spreadsheets)
• Represent collections of “field” measurements
• Highly heterogeneous (format, content, semantics)

<table>
<thead>
<tr>
<th>site</th>
<th>plot</th>
<th>spp</th>
<th>ht</th>
<th>dbh</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCE1</td>
<td>A</td>
<td>piru</td>
<td>21.6</td>
<td>36.0</td>
<td>4.5</td>
</tr>
<tr>
<td>GCE1</td>
<td>B</td>
<td>piru</td>
<td>27.0</td>
<td>45</td>
<td>4.8</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>GCE9</td>
<td>A</td>
<td>abba</td>
<td>23.4</td>
<td>39.1</td>
<td>3.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>yr</th>
<th>trns</th>
<th>m2</th>
<th>acidity</th>
<th>piru</th>
<th>abba</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>1</td>
<td>4</td>
<td>5.1</td>
<td>20.8</td>
<td>14.1</td>
</tr>
<tr>
<td>2009</td>
<td>1</td>
<td>4</td>
<td>5.4</td>
<td>21.1</td>
<td>15.3</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>4</td>
<td>5.8</td>
<td>22.1</td>
<td>16.9</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>
Broad Scale Research Studies

Analyses require a **wide range** of data

- Study phenomena at broad **geo, temporal, biological** scales

  *Examples*: influence of nitrogen fertilization on grasslands, effects of climate change on tree allometry (growth)

Researchers struggle to …

- *Discover relevant* datasets for a study
- *Combine them* into an integrated product to analyze

- **Semtools** and **SONet** projects (NSF funded)
  - Develop tools to exploit ontologies for improved data discovery and integration
Observation Models

Emerging models for observations

- Examples: O&M, ODM, OBOE, VSTO, ...

- Higher-level representations of observations
  - A standard set of “core concepts”
  - Entities, properties measured, units, protocols, etc.

- Better interoperability and uniform access
  - Standard representation of underlying datasets
  - Abstract away underlying format issues
Observational Data Model

- An OWL-DL ontology
  - Basic concepts for describing observations
  - “Extension points” for domain-specific terms
Observational Data Model

Observations are of entities (e.g., Tree, Plot, …)
  – Represents an “observation event”

subClassOf(Observation ObjectSomeValuesFrom(ofEntity Entity))
FunctionalObjectProperty(ofEntity Entity)
Observational Data Model

A *measurement* consists of

– The characteristic measured (e.g., Height)
– The standard used (e.g., unit, coded value)
– The protocol used
– A measured value
Observational Data Model

Observations can have **context**

- Geo, temporal, or biotic/abiotic environment in which measurements were taken
- Context is **transitive**
Observational Data Model

• Terms drawn from domain-specific **ontologies**
  – Specific entities, characteristics, standards, protocols
Semantic Data Annotation

Each dataset has a specific model “configuration”

Data attributes linked to measurements

Facilitates search & integration
**Challenge**: Domain-Specific Ontologies

Ontology elicitation is difficult in Ecology
- Breadth of terms & technical expertise needed

**Our Goal**: Tools to help ecologists define ontology terms
- Focus on specialized “measurement types” (annotation)
- Initially developed a spreadsheet-based approach
  - Natural for ecologists, but too open-ended
  - No immediate feedback
- Developed a “form-based” approach within Protégé …
Protégé Forms for Observational Ontologies

- Specialized forms in Protégé for defining
  - Basic types: Entities, Characteristics, Protocols
  - Units (simple, composite, derived) and Unit conversions
  - Measurement types (combines other types)

- Each form a simple “fill in the blank”
- Constrains choices to appropriate classes
- Generates underlying OWL-DL axioms
Protégé Forms for Observational Ontologies

Example Measurement Type: “Nitrate Concentration in Fresh Water”
Protégé Forms for Observational Ontologies

Example Measurement Type: “\textit{Nitrate Concentration in Fresh Water}”

Corresponding OWL-DL Axioms
Ongoing Work

• Protégé forms currently used in three domains
  – Santa Barbara Coastal LTER (standard attribute types)
  – Plant trait measurements
  – Salmon data collections

• Extending to support richer measurement types
  – E.g., for required context measurements

• Generalize forms-based approach
  – Construct forms for other “core” ontologies
  – Layout form “widgets” (one form per class)
  – For each widget, specify corresponding class “paths”