Semantic Tools: Using Observation Ontologies for Data Discovery and Integration

M. Jones[1], S. Bowers[2], B. Leinfelder[1], M. O' Brien[3], M. Schildhauer[1], C. Jones[1]

? Why Use Ontologies?

Metadata languages in the environmental sciences have traditionally focused on high-level and human-readable descriptions of data sets and their syntactic properties, but often lack semantic information needed to fully identify and align heterogeneous data. Rich semantic information describing the content, structure, and relationships of environmental data can be used to build effective tools for discovery and integration of disparate data.

Software Tools

We have developed software tools within the Semtools project that use **formal annotations** on environmental data to link domain-specific ontology terms from the **Extensible Observation Ontology (OBOE)** to particular facets of environmental data sets. This annotation approach allows us to maintain data in traditional scientific formats (e.g., CSV text, NetCDF) while still establishing formal semantics for the data. The annotations can then be used to **materialize a full knowledge model** of the data (e.g., an RDF graph compatible with Linked Open Data conventions), which in turn **drives discovery and integration** tools.

We enable this through extensions to the **Metacat** data repository system that can be used for **semantic search** and **semantic data subsetting** for heterogeneous environmental data. The Metacat software uses the Ecological Metadata Language (EML) Data Manager Library for loading EML-described data, and incorporates the Semtools Semantic Mediation System (SMS) Java Library to enable query term expansion, semantic data subsetting, and data integration.

[1] National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara
[2] Deptartment of Computer Science, Gonzaga University
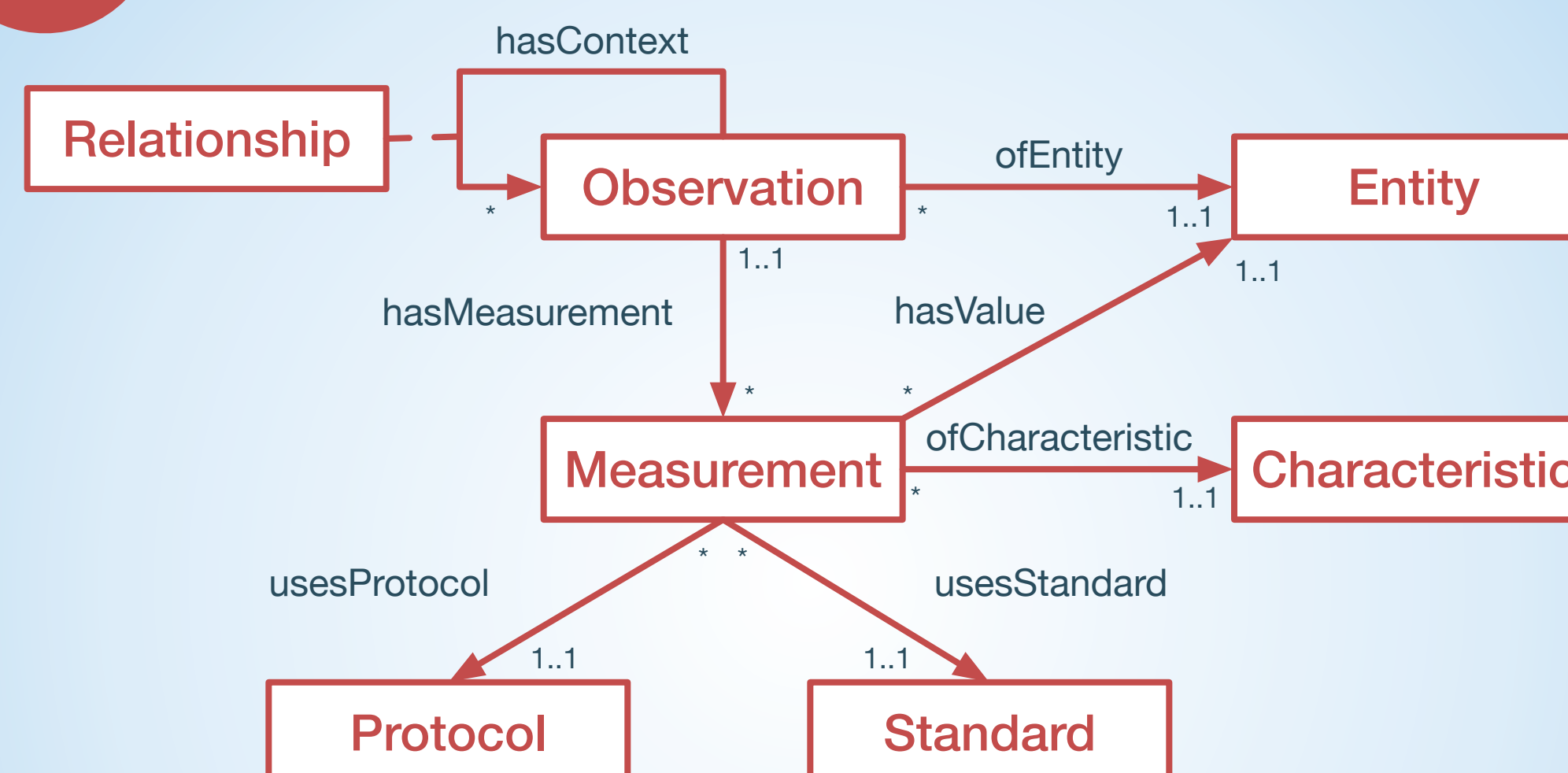[3] Marine Science Institute, Univ. of California, Santa Barbara

Contact: jones@nceas.ucsb.edu

http://semtools.ecoinformatics.org

1 OBOE: Extensible Observations Ontology

An **Observation** is made of an **Entity** (e.g., a biological organism or an environmental feature). Other Observations may provide **context** for a particular Observation.

A **Measurement** assigns a **value** to a **Characteristic** of the observed entity (length of a fish, discharge rate of a stream)

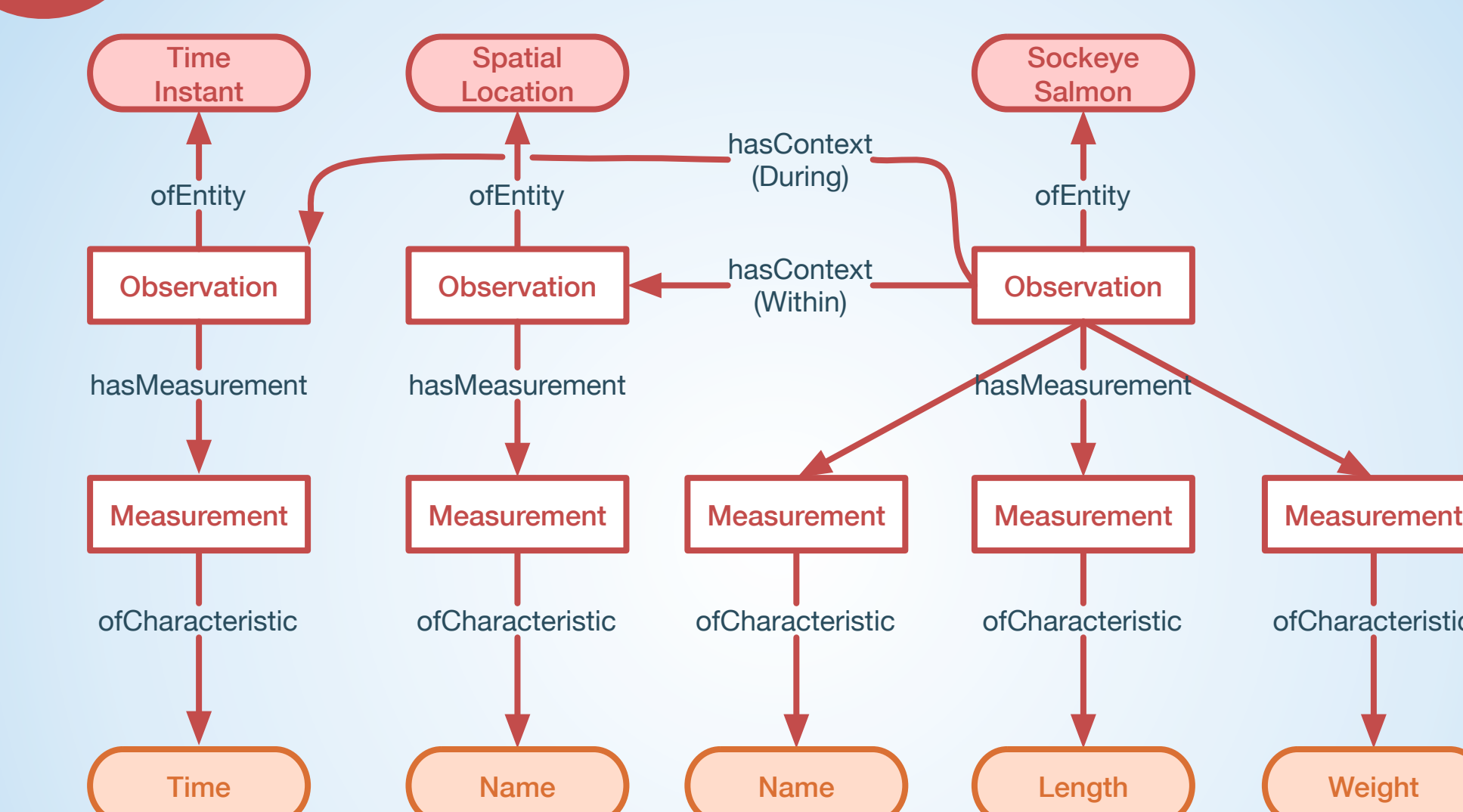A Measurement may use a **Standard** (e.g., meter) or **Protocol**

2 Domain-specific ontology extensions
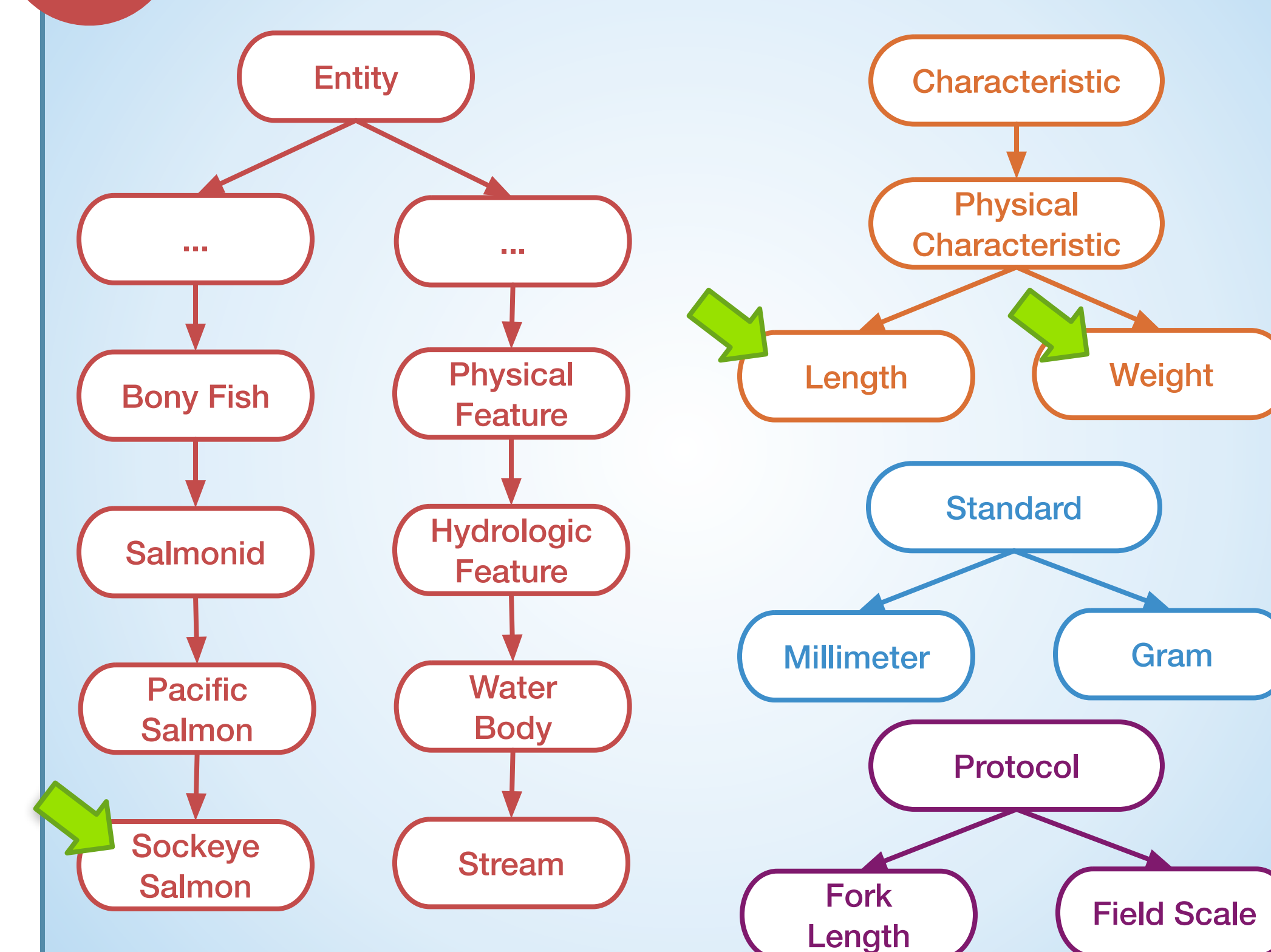
3 Annotations: Mapping Data to Concepts

date	site	spp	length	wt
01-Apr-2011	Grays	SCK	200	92.5
07-May-2011	Grays	SCK	187	90.1
04-Jun-2011	Grays	SCK	194	88.9
03-Jul-2011	Grays	SCK	211	78.2

4 Improved Data Discovery and Integration

Query recall is enhanced by using term-expansion enabled by traversing the ontology class subsumption hierarchy

Result filtering is strengthened by limiting datasets that match combinations of the components of the OBOE model

Data integration is supported by determining equivalence mapping based on dataset annotations to the OBOE model

Combined Data			
Site	Package	ForkLength	Weight
Grays	A	200	92.5
Grays	A	187	90.1
Grays	A	194	88.9
Grays	B	211	78.2
Snake	B	222	89.3
Snake	B	203	97.0
Snake	B	192	90.0
Snake	B	200	89.9