Complex Data Modeling for Simpler Data Access

TDWG 2014, Jönköping, Sweden

Ramona Walls
Robert Guralnick
A Canonical Example of “Opportunistic Collecting” typical in biocollections
Figure 2. Schematic representation of transect lines within the forest fragments and in the pasture matrix. Solid lines represent transects Tomahawk, Sherman, and snap traps and dashed lines represent transect lines with pitfall traps.
transect

*sample collection point

depth

water sample at depth X

aliquot

metagenome
Bag of terms

DwC
http://vegbank.org/vegbank/general/faq.html#datamodel
OBO-E:

Madin et al. 2007 Ecol. Informatics doi: 10.1016/j.ecoinf.2007.05.004

O&M:
Most biology requires work at the intersection of disciplines.
Material entities, information entities, and processes in the Basic Formal Ontology

![Ontology Diagram]

- BFO:entity
  - BFO:continuant
    - BFO:dependent continuant
    - IAO:information content entity
  - BFO:independent continuant
  - BFO:material entity
  - BFO:occurrent
  - BFO:process

(subClassOf (is a))
observations versus specimens

BFO:entity

- BFO:occurent
- BFO:material entity
- OBI:planned process
  - OBI:specimen collection process
  - BCO:observing process

- BFO:independent continuant
  - OBI:specimen

- BFO:continuant
  - IAO:information content entity
    - BFO:dependent continuant
      - BCO:observation

subClassOf

has output
Specimen data from a Darwin Core Archive: VertNet
**sampling process logical definition:**

- assay and (achieves_planned_objective some ‘biological feature identification objective’)

- has_specified_input some ‘sampling feature’
- has_specified_output some ‘sample data item’

**material sampling process logical definition:**

- 'planned process' and (achieves_planned_objective some ‘material sampling objective’)

- has_specified_input some ‘material sampling feature’
- has_specified_output some ‘material sample’

**specimen collection process logical definition:**

- 'planned process' and (achieves_planned_objective some ‘specimen collection objective’)

- has_specified_input some ‘material entity’
- has_specified_output some ‘specimen’
BCO Taxonomic Inventory Process Class and Sub-classes of different kinds of processes
BCO:taxonomic inventory \rightarrow IAO:information content entity \rightarrow IAO:protocol \rightarrow OBI:planned process \rightarrow BCO:taxonomic inventory process

IAO:protocol \rightarrow \textit{has output}
BCO:taxonomic inventory

IAO:information content entity

IAO:protocol

BCO:geographically structured search protocol

OBI:planned process

BCO:taxonomic inventory process

BCO:restricted search

Park Survey 1

Park Survey 1 protocol

follows protocol

has output

Park Survey 1 process

follows protocol

has output

Type

has output
BCO:taxonomic inventory

IAO:information content entity

IAO:protocol

BCO:geographically structured search protocol

Park Survey 1 protocol

Park Survey 1

OBI:planned process

BCO:taxonomic inventory process

BCO:restricted search

Park Survey 1 process

follows protocol

has output

type

follows protocol

has output

has output

follows protocol

has output

follows protocol

has output
Conclusions

• BCO splits the middle ground between the high level OBO-E world view and the flat way of representing a process that has a single output to allow us to represent all kinds of different content.

• BCO can serve as a sandbox to test out new models and terms for describing sampling processes and data, to inform standards like DwC.
Acknowledgments

• Dozens of participants at BCO workshops and hackathons over the past two years

• NSF-EAGER: An Interoperable Information Infrastructure for Biodiversity Research (I3BR)

• NSF: Research Coordination Network for GSC (RCN4GSC)

• VertNet and University of Kansas Biodiversity Institute