Ontology Summit 2014 Session 06 Track D: Tackling the Variety Problem in Big Data – Summary

Ken Baclawski

Anne Thessen

Track D Co-Champions

February 20, 2014

Introduction to Big Data

- Big Data can be “big” in terms of volume, velocity and variety
- Can address important societal and commercial needs
- Current techniques are inadequate, ontologies can help
  - Metadata
  - Integration
  - Privacy
  - Provenance

Track D-1 First Talk

- Eric Chan (Oracle)
- Industrial use of ontologies for Big Data
- Developed ontology for OODA loop
- Excellent example of ontology reuse in vastly different domains
  - Fighter pilots (original)
  - Operating system performance monitoring
  - Scientific hypothesis generation

Track D-1 Second Talk

- Nathan Wilson (Marine Biological Laboratory)
- Using ontologies to manage biodiversity data
- Filling in gaps themselves
- Ontologies changing over time is an issue
  - Terms change
  - Knowledge changes
  - Ontology is being adapted/reused as the domain evolves
- Crowdsourcing curation

Track D-1 Third Talk

- Ruth Duerr (National Snow and Ice Data Center)
- Using ontologies to manage data about ice
- Building ontologies from small modules to maximize reuse
- Working with existing data types to build the ontological structure (Egg Notation)
- Integrating natural science with social science
- Employing multiple languages

Track D Community Input

- Data/metadata annotation
- Semantic tagging
- Ontology mapping can help with data set merging
- Aid machine learning algorithms

Summary and Identified Challenges

- Reuse in many forms is a recurring theme
- Evolution of ontologies and terms
- Adapting to existing workflows of domain experts
- Tool incompatibility problems
- Gaps in existing ontologies
- Ontology mapping within domains