Ontology Summit 2014 Session 12
Synthesis II: Technical Tracks and Hackathon - Track D

Ken Baclawski
Anne Thessen
Track D Co-Champions
Tackling the Variety Problem in Big Data with Ontologies

• Big Data can be “big” in terms of volume, velocity and variety

• Sources of variety
  - Domain-specific knowledge
  - Governance
  - Data structure and completeness
  - Provenance
  - Workflows
  - Privacy
  - Identity/Terminology
How Can Ontologies Help?

- Represent background knowledge
- Annotate data and metadata
- Describe data structure
- Record provenance of data
- Record workflows
- Describe data governance
- Facilitate data integration
- Facilitate hypothesis generation
- Record privacy needs
Use Cases

- Harvest from data partners (Wilson, Uceda-Sosa)
- Modular development (Duerr)
- Reuse (Every speaker)
- Formalize existing informal models (Chan, Duerr)
- Develop ontology with extension points (Chan)
- Involve communities (Duerr)
- Governance framework (Chisholm)
- Vocabulary pipeline (Brickley)
- Pattern matching (Brickley)
- Information ecosystem (Uceda-Sosa)