Ontology Summit 2012

Track 3
Challenge: Ontology and Big Data

Co-Chairs

Mary Brady

Ernie Lucier

Thursday, February 9, 2012

Mission Statement for Track 3 - Challenge: Ontology and Big Data

The mission of Track 3 is to identify appropriate objectives for an Ontology and Big Data challenge, prepare problem statements, identify the organizations and people to be advocates, and identify the resources necessary to complete a challenge. The goal will be to select a challenge showing benefits of ontology to big data.

Action Plan

- **Prepare problem statements**
- **Identify objectives**
- **Identify advocates**
  - **Organizer**
  - **Organizations**
  - **People**
- **Identify resources**
- **Perform Ontology and Big Data Challenge**

Meeting Big Data Challenges through Ontology

- The mission of this track is to identify appropriate objectives for an "Ontology and Big Data" challenge, prepare problem statements, identify the organizations and people to be advocates, and identify the resources necessary to complete a challenge. The goal will be to select a challenge showing benefits of ontology to big data.
- One of the NCO's goals is to enhance collaboration and accelerate agencies' adoption of advanced IT capabilities. NITRD seeks to accelerate deployment of promising research technologies; share protocol information, standards, and best practices; and coordinate and disseminate technology assessment and testbed results. NITRD coordinates federally supported IT research under the leadership of OSTP. Ontologies and the semantic web support Open Government Directive.
- The goal of "**Meeting Big Data Challenges through Ontology**" Track 3 is to identify issues that can be addressed using an ontology challenge. Challenges can take many forms and target many issues.

Potential issues to be addressed by challenges (1 of 3)

- Enhance collaboration and accelerate agencies' adoption
- Accelerate the adoption of ontological methods, maximize public awareness, and impact of research.
- Increase the number of agencies using ontologies, i.e., earlier adoption
- Where should our focus be to accelerate agencies' adoption of ontology capabilities?
- How many scientists, physicists, engineers, programmers, big data administrators, etc. have experience with ontologies?
- Is the growth of ontological implementations and technologies with Big Data constrained by the shortage of qualified personnel?
- Inform, educate, and include the public in scientific research and discovery. Public involvement could be a critical component of our success

Potential issues to be addressed by challenges (2 of 3)

- A mismatch between those with data and those with the skills to analyze the data
- Are programmers able to optimize the use of unstructured or semi-structured data sets for scientists and engineers?
- What are the talent and skill set issues impacting the use of ontologies?
- The skills important to the growth of ontological technologies with Big Data include a combined understanding of a scientific or engineering discipline and knowledge of ontology-based technologies.
- Programmers are not able to optimize the use of unstructured data for scientists and engineers
- Scientists and engineers without ontology training may use brute force programming – this can be inefficient and the scientists and engineers without training may not be aware of options and capabilities using ontology-based technologies
- Strategic significance to the economy, e.g. enabling competitive products.

Potential issues to be addressed by challenges (3 of 3)

- How long does it take to become productive in the ontology environment?
- Can universities expand coursework in ontologies and integrate ontological methods into the requirements for science degrees? At the undergraduate level? At the graduate level?
- Identify individuals who have both domain experience and an understanding of what it means to apply ontology technologies.
- Increase the number of individuals capable of applying ontology technology
- Ontology-based technology evolution for big data may be slow or non-existent
- Advances in the use of ontology technology can be difficult or unattainable without an adequate number of properly trained personnel, including scientists, engineers, programmers, system administrators, technologists, and all others that make up the big data systems.
- Expanding the markets for ontologies could make the field a more attractive career path. What is the growth rate of the ontology market? Further expansion could spark investment and make ontologies an even more vibrant, attractive market for young people to enter.
- A Challenge may seed and transform the current status quo
- Software dilemma analogy with ontology
  - Ontologies and software perceived to be a commodity resulting in little or no investment in research. Projects use ontologies as one of their tasks

Potential Challenge Goals

- Increase the awareness of ontology technology among programmers/database managers
- Increase the number of qualified personnel to facilitate the growth of the ontology technologies
- Accelerate agencies' adoption of semantic and ontology capabilities, i.e., help spread the use of ontologies
- How to implement semantic and ontology technologies
- People in the domains of science, engineering, software, computer science, etc. can benefit from a combined knowledge of their domain and application of ontology-based technologies. A combined understanding of these domains and ontology-based technologies may encourage the growth of technology.
- Ameliorate any mismatch between those with data and those with the skills to analyze it
- Enable scientists and engineers to make maximum use of big data
- Enable scientists and engineers to understand the potential of ontology-based systems integration
- Enable ontologists to understand scientists and engineers needs
- Improve critical areas of current practice

Process

- Process can benefit from Challenge organizations
- Identify organizations, interests, and establish agreements
- Select Data Sets
- Set rules for their use during the challenge
- Define products
- Set competition period, i.e., one day, three weeks, etc.
- Evaluate and Post results

Panelists

- Professor Barry Smith (University at Buffalo, SUNY, Ontology)
    - Big Data that might benefit from ontology technology, but why this usually fails
- Mr. Chris Musialek (for Dr. Jeanne Holm, Evangelist, Data.gov)
    - Data.gov and how its very large collection of datasets (>400,000) could benefit more from ontological approaches and what we have done so far in that area.
- Mr. Bryan Thompson and Mr. Mike Personick (SYSTAP, LLC., U.S.)
    - Managing Scale in Ontological Systems
    - We will take a look at large scale ontological systems from several angles, including limitations placed by the CAP Theorem, tradeoffs in expressivity versus scale, tradeoffs in timeliness versus completeness, and implications for information architecture and system architecture.
- Mr. James Kirby (Naval Research Laboratory)
    - Ontology for Software Production