Ontology Summit 2014
Big Data and Semantic Web Meet Applied Ontology **Track A-Semantic Content Reuse:** Synthesis

MikeBennett, GaryBergCross, AndreaWesterinen

April 3, 2014

Introduction, Mission and Scope of Track A

- Semantic technologies such as ontologies and related reasoning play a major role in the Semantic Web and are increasingly being applied to help process and understand information expressed in digital formats.
- The mission of Track A is to leverage common semantic content to reduce the burden of new, quality ontology creation while avoiding silos of different ontologies.
- The range of semantic content reuse being used on the Web and Big Data is broad.
  - Content includes whole or partial ontologies, ontology modules, ontological patterns and archetypes, and common, conceptual theories related to ontologies.
  - The role and relation of methods, bottlenecks and tools was discussed.

Approach

- Enlisted 6 speakers and the community to discuss reuse issues and problems, and present their efforts and experiences to address these
- Referenced and integrated aspects of past Ontology Summits (for example, the Ontology Repositories discussions)
- Promoted discussion of track session topics on the Ontolog/Summit forum both before and after sessions
- Participated in Hackathons related to reuse and tools
- Distilled the virtual meeting topics to a useful summary and set of speakers for the face-to-face Symposium

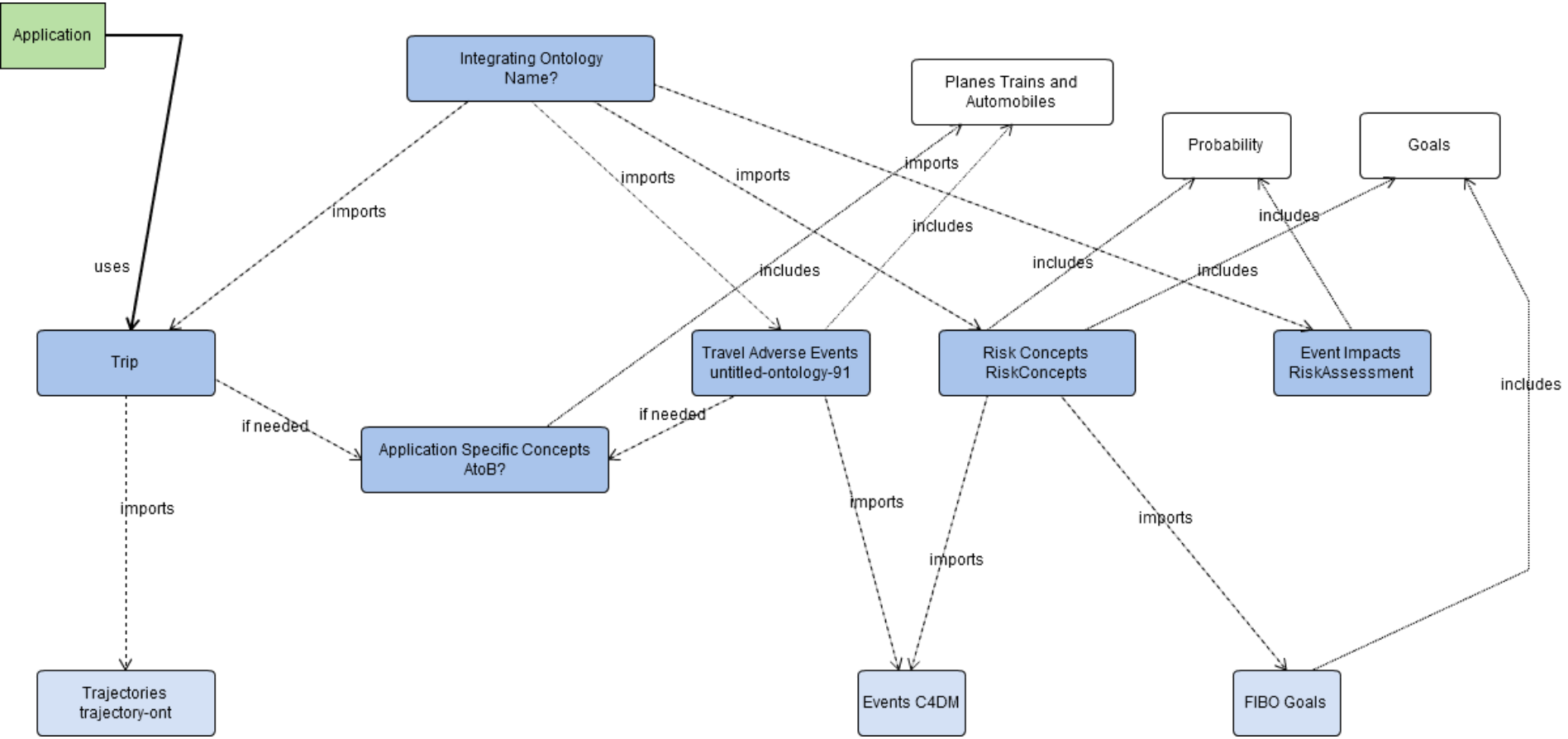
Our 8 Speakers & Their Talks

1. Mike Bennett- Overview of the track
2. Gary Berg-Cross – Overview of Reuse and Issues
3. John Sowa- "Historical Perspectives: On Problems of Knowledge Sharing"
4. Pascal Hitzler (Wright State U) - "Towards ontology patterns for ocean science repository integration"
5. Andrea Westerinen - "Reuse of Content from ISO 15926 and FIBO"
6. Ms. Megan Katsumi & Michael Gruninger "Reasoning about Events on the Semantic Web"
7. Michel Dumontier - "Tactical Formalization of Linked Open Data"
8. Kingsley Idehen - "Ontology Driven Data Integration & Big Linked Open Data"

ODPs and Integration Hackathon

- Aim: Explore re-use of diverse ontologies
- Context: Selected risk as a use case which requires concepts across a range of subjects
- Hackathon: Chose travel risk as a specific area to develop
- Ontologies
  - 1 created from analysis of available data sources
  - 2 created by extension or extraction from existing ontologies
  - 1 ontology with additional risk assessment concepts
- Agreed on common ontology pattern for risk and applied this to the ontologies
- Used Visual Ontology Modeler tool to visualise and compare ontologies

ODPs and Integration Hackathon Architecture



VOCREF Hackathon

- VOCREF: Vocabulary and Ontology Characteristics Related to Evaluation of Fitness.
- Highlights:
  - Use of GitHub (https://github.com/vocref/vocref) to store and allow collaboration on the ontologies, and hold current issues
    - Work will continue to address the issues and add content
  - Use of OWL Functional Syntax for serialization (to ease merging issues)
  - Defined modularity with a top-level framework ontology, and smaller ontologies that subclass from it to capture:
    - Ontology characteristics and other metadata
    - Mappings to existing ontologies

Reuse Issues

- Reuse issues are not unique to ontologies/schemas.
  - There are parallels and differences with software reuse.
- For successful reuse of semantic content … it is important to understand how content is being used, what methods to coordinate reuse are available, and what tools are helpful.
- Tooling for modularity, documentation, etc. is critical.
  - Broader use by mainstream efforts including Big Data is bottlenecked by the paucity of semantic tools integrated into mainstream tools, along with the inherent learning curve of understanding semantics.
- In practice, reuse is dependent on both the availability of well-documented content AND tooling that supports finding and incorporating this range of content.

Conditions for Reuse

- Two of the most critical aspects are that the content is "understood" (documented) and in a form conducive to reuse (or convertible to such a form).
  1. Documentation must include the basic details of the semantics, but also the range of conditions, contexts and intended purposes for which the content was developed.
  2. It was recommended that standard metadata for reuse be defined and complete exemplars provided.
- Specific items for consideration (capture and retrieval)
  - Content is accessible and can be found
  - The re-user is motivated to find the content
  - The content is in a form conducive to re-use or can be converted/transformed to a usable form
  - The re-user knows how to do the conversion/transformation
  - The content is logically consistent with the micro-theories of the re-user and this can be established
  - The re-user trusts the content and its quality, and believes that this quality will be maintained
  - When it is best to reuse content within the lifecycle – it can be confusing as a starting point
  - Original ontology creators may have had different range of applications in mind for a given class or property (especially if developed for an application rather than standard)

Tooling (I)

- Ontology repositories with good search capabilities and governance help with the task of finding relevant content
  - Including both topical ontologies and linked data schemes
  - One possible repository is the Open Ontology Repository
- Another resource from the LOD world is the Linked Open Vocabulary (LOV)
  - In an increasingly linked data world, vocabularies rely more and more on each other through reusing, refining or extending, stating equivalences, and declaring metadata
  - LOV provides a service to find relevant vocabularies

Tooling (II)

- What is needed to support the development of modular ontologies and schemas?
  - Is this different than the tooling to query/find/reuse the modules?
  - It may be, but the tools should "play well together" to support a complete life cycle.
  - Tools for modular design / architecture of ontologies a big gap!
- Controlled natural language tools (to generate candidate ontologies) may ease the KA bottleneck.
  - Again, these should be integrated with development and query tools.
  - These should also be integrated with standard IT tools.

Best Practices (I)

- Small, more modular ontologies and schemes
  - More possibilities for reuse due to greater focus and cohesiveness, and likely less dependency on the original context
  - Modularity viewed from the perspective of the user, not the creator
- Collect and document approaches to modularization, best practices and specific patterns
- Dimensions of variability should be understood and addressed to improve modularity
  - Variability across the contexts (for example, a certain concept or property may be present or absent in different contexts and uses)
  - Variability over time

Best Practices (II)

- "Integrating" modules defined for an application or domain
  - Employing owl:equivalentClass and OWL axioms to map between the concepts, properties, etc. of the complete set of modular ontologies that address an application/domain
- Each module and its concepts, properties, axioms, ... well-documented via well-established labels and predicates
  - SKOS, etc.
- Patterns of concepts separated from patterns of usage, analysis, traversal and diagnosis
- Multiple domains represented such that the ontologies and schemes represent "common needs"
  - No single domain focus
- Plans for variability and change documented with the modules
- Constraints or axioms distinguished as:
  - Definitive ("defining" the concepts that are necessary in the core module)
  - Pragmatic (related to the business uses or a particular domain)

Best Practices (III)

- Separate reuse of classes/concepts, from properties, from individuals and from axioms
  - Easier to target what is possible to reuse and reduces the amount of transformation and cleaning that is necessary
- Define and discuss concept naming
  - Names can be surrogate or human-readable identifiers, both approaches have their advocates, and pros and cons
  - Labels as documentation (such as from SKOS) are valuable regardless of the identifier scheme that is chosen

Ontology Management for Reuse

- Ontologies must include consistent, supporting metadata for query
  - Possible metadata includes context, use cases, labels, governance information, etc.
  - Building on the Ontology Metadata Vocabulary and concepts (or ontologies) from the Hackathons
- Reuse enhanced by feedback and user input
  - Possibly include both a recommendation system and feedback mechanisms in the repository
- Governance needs a process and its enforcement
  - Process should include open consideration, comment, revision and acceptance