Semantic Interoperability for Web Services: The Needs and Challenges for the Demanding Consumer

Ken Laskey

W3C Workshop on Frameworks for Semantics in Web Services

10 June 2005

Perspectives from Typical Customers

- Technology is fun, but we are interested only to the extent to which it solves problems
- Focus of this presentation
  - What problems do we need to solve?
  - What do we need in the way of tools to support the most effective solutions?

Typical Problems, Typical Solutions

- US Department of Defense (DoD) Directives have moved from single vocabulary approach
  - 8320.1 (1991)
    - To determine what data elements should be standardized, how grouped
    - Results stored in DoD Information Resource Dictionary System (DoD IRDS)
    - Directive cancelled in 2004
  - 8320.2 (2004)
    - Net-Centric approach
    - Data shall be made visible, accessible, and understandable to any potential user
      - Accessible by being published in shared spaces
      - Understandable by publishing associated metadata
    - Semantic and structural agreements for data sharing shall be promoted through Communities of Interest (COIs)

New Solutions, New Problems

- In SOA, distributed resources (data and processing) developed by independent entities

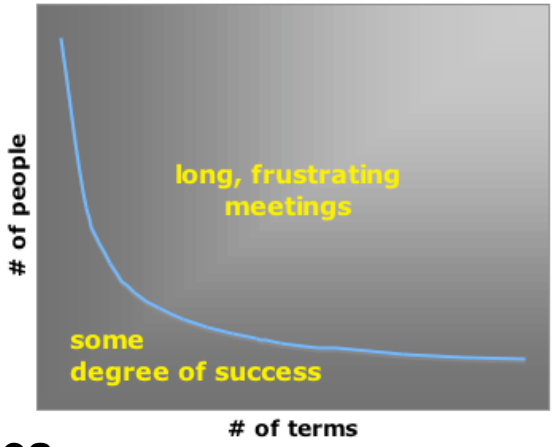
  ⇒ likely different vocabularies, independent semantics

- US Department of Defense (DoD) Net-Centric Data Strategy assumes Communities of Interest (COIs) will develop community vocabularies, provide and find services
  - Institutional (long-standing) and expedient COIs
  - But how will these (especially expedient COIs) interoperate?
- Current emphasis on and mandates for information sharing
  - Vocabularies often rooted in historical developments of domains and their functional role within an enterprise
  - Need for semantic translation/negotiation

A single standard vocabulary across diverse domains is not feasible but how do you share across vocabularies?

Underlying Challenges (1)

- Avoiding n^2 -- the compromise vocabulary
  - Some success when need limited number of people to agree on limited number of terms
  - Helps if pressing need, can avoid tough issues
  - Success tends not to scale as numbers increase, flexibility of what can avoid decreases
  - Still some drawbacks
    - Minimum vocabularies squeeze out subtleties to collapse differences ⇒ lose information
    - Verbose vocabularies preserve subtleties but miss higher level similarities ⇒ everything different
    - Compromises that work for one application may not be suitable for another



The world isn't n^2 but it isn't n either - deal with it!

Underlying Challenges (2)

- Description through use of multiple ontology subsets
  - Ex: different manufacturers using same materials may use ontology subsets emphasizing different properties
  - Ex: different vendors of same product may emphasize different uses, require different ontology from each perspective
- Timely evolution of vocabulary to support new products, new features
  - Must describe in order to convey what new terms and relationships to standardize
  - Must support variations while learning what are useful descriptions

Underlying Challenges (3)

- Imprecise description wanted when it provides advantage
  - Vendors want to be associated with desirable categories, not necessarily most accurate
  - Vendors want to differentiate their products, show uniqueness
  - Vendors want ambiguity to their benefit
    - ex: neighborhood where house for sale is located
  - Degree of precision wanted depends on use of information
    - Medical diagnosis looking to identify disease, correct intervention
    - Merchandise sales looking for customer, varying degrees of satisfaction
- Incomplete or inconsistent information
  - Provided when required or gives advantage but may not be available
  - Must be fused from multiple sources to support valid decision making

Specific Needs (1) – The Many Vocabularies of XML Tags and Metadata

- Metadata as the magic bullet for discovery but requester must understand
  - organizing schema by which resources of interest have defined metadata ⇒ the XML tags
  - values that correspond to schema and define the metadata instances ⇒ the values between the tags
- Many solutions point to searching registry
  - Typically define property tags but not precise semantics of properties or property values
  - How to provide well-defined properties and interoperable target values for search?
- Implies need for
  - associating tags (such as WSDL) and property values to ontologies describing semantics
  - tools that support disambiguating vocabularies through descriptive connections ranging from namespaces to ontologies

Specific Needs (2) – Versioning as a Variation of Multiple Vocabularies

- Schema versioning has become topic of immense interest
  - Can you modify schemas without breaking validation?
  - Backward and forward compatibility
- Necessary but not sufficient to have version numbers
  - What are semantics of version numbers?
- Not enough to have tags indicating expected compatibility
- Need to capture specific differences and effectively move between versions
  - Important for tools, information, and information structure
- Web services makes it easy to have multiple versions BUT we have to make multiple versions usable

Specific Needs (3) – The Many Uses of Constraints

- Describes array of assumptions, restrictions, conditions
  - Preconditions and post-conditions of every task and sub-task
  - Results of constraint evaluation
- Covers both requester and requested resource
  - Should requester be permitted access?
  - Does resource satisfy requester needs and requirements?
  - *Ex:* requester looking for service with specific signal processing capability, service looking for requester with paid license
- Reuse of general constraints
  - Written in vocabulary of those creating (e.g. lawyers)
  - Later associated with different domains (e.g. software engineers)
- Annotations to indicate how and when information used
  - Labeled deduction to indicate logic in which expressed
  - Pedigree as identity of rules and tools used to evaluate compliance

Overriding Need for Reuse Across Domains

- Perspectives on the challenge
  - Use of overlapping/complementary information generated from separate efforts and described using separate vocabularies
  - Preservation of investment in existing schemas
  - Need to relate structures created for a given problem to broader knowledge base
    - Increase information value
    - Enhance semantic interoperability
- Emphasis for standards needs to be on framework on which to design solutions, not on proliferation of specific solutions
  - Today's new integration paradigm becoming tomorrow's legacy to integrate
  - Do not need another data model with dozens of pages of detailed APIs
  - Need standard means to capture and accommodate for change rather than building yet another standard with each change

Possible Approaches of Interest

- Annotations that connect structured documents with ontologies describing related domains
- Upper ontologies to encode higher level concepts and provide consistent relationships for more detailed ontologies to leverage
  - Basic assumptions regarding nature or state of the world and encoding past thought on issues such as part/whole, identity in the face of change over time
  - Associate more specific descriptions using higher level concepts
  - Lattice of theories – macroscopic relation between individual ontologies
- Ontology of mapping concepts to provide basis for generating and capturing specifics of mapping instances
- Transitive reasoning to leverage specific relationships in making connections to previously unmapped entities
- Polymorphic reasoning to leverage entities related to one ontology when working with another ontology with which there are formal relationships

Thoughts on Moving Forward

- The promise of the power and flexibility of Web services will not be realized until one can find, access, and understand semantic context as easily as one can find a service interface
- In mapping between vocabulary structures
  - What information is needed to sufficiently capture a mapping?
  - What kind of structures are needed to capture the mapping?
  - What tool support is needed to extract information from catalogued mappings and infer new relationships?
- What is the difference between a compromise integration ontology and an upper ontology?
  - Understanding this may be the difference between struggling with standard integration and finding a new way to attack problem
  - If you can map two ontologies to an ontology of higher concepts, can there be yet higher order ontologies and higher order mappings?
- If we can find ontological approaches to relating semantics of semantics, is this solution for others to describe their semantics?