

Finding or Making Ontology Fit for your Purpose

*Points worth special attention in Ontology
Quality Assurance*

Amanda Vizedom
Wind River Consulting
Ontology Summit 2012

X-Track A1 Panel Discussion: "Implementing Ontology
Quality Measures in Big Systems Engineering"

Ontology Quality Factors

Ontology Quality Factors = Ontology features and characteristics which frequently factor into the ability of an ontology to meet the requirements of a particular usage.

- Particular Ontology Quality Factors may be
 - more or less universally relevant
 - more or less significant for any particular usage.
 - all-or-nothing or satisfiable in degrees
 - quantitative or qualitative
 - more or less readily expressible as requirements using well-understood metrics
 - more or less measurable using well-understood methods
 - evaluated by manual, automated, or mixed techniques
 - more or less costly to measure and/or satisfy

A Partial Inventory of Ontology Quality Factors (1)

- **Formal soundness:** logical consistency, KR language use accords with formal semantics
 - Absolutely necessary
 - Almost never sufficient (for ontology selection or development)
- **Accuracy (verisimilitude):** model captures reality of whatever is being modeled.
 - Only a usable requirement when accompanied by an identification of what is to be modeled. May be "concrete" things, maybe abstractions, may be conceptualizations, may be elements of information.
- **Coverage:**
 - Only a usable requirements when accompanied by moderately detailed specification of coverage or defining/scoping corpus (e.g., data streams or messages whose content must be expressible in the ontology, documents in which mention of domain concepts must be indexible) Includes representation of entity types and relationships, and either includes representation of particulars or includes holders and connectors to represent changing particular

A Partial Inventory of Ontology Quality Factors (2)

- **Logical characteristics** & reasoning supported
 - None? Subsumption only? Non-monotonic? Uncertainty? Backward? Forward?
- **Linguistic characteristics** & support for required NL-related functions
- **Reusability**
- **Extensibility**
- **Usability**
 - By whom?
 - for what?
- **Performance characteristics**
- **Compliance** with mandated or applicable standards
- **Dynamic Management**
 - TMS and regression as ontology evolves
 - effects of use of externally-controlled ontologies
 - collaborative, distributed, and/or authoritative development
- **Traceability & Trust**
 - Provenance, bookkeeping, meta-reasoning

Sample Ontology QA Measures (1)

- Formal, logical validation:
 - **Solid Foundations:**
 - Principles are well-established and specified. Ontology languages have formal semantics.
 - **Tools and Techniques exist; improvement and extension needed.**
 - Available for most common ontology languages. Needed for more. Sufficiently usable by those who understand foundations.
 - Absolutely necessary.
 - *Soapbox:* Ontology dev without minimum reasoner testing is like programming without a DE or even a compiler. Don't do it! Humans cannot manually, reliably, detect inconsistencies, mistakes and unintended inferences over a large ontology.
 - **Solid core of trained and qualified personnel, *But***
 - Much smaller than crowd currently developing ontologies
 - Not easily distinguished, by others, from the many who don't understand the formal semantics, may not have the background, and don't know how to use a reasoner.

Sample Ontology QA Measures (2)

- Accuracy/Verisimilitude, Coverage:
 - **Partial Foundations, care needed:**
 - Some established principles regarding model accuracy
 - Confusion about object of model often confounds
 - **Tools & Techniques exist; need more, better, & guidance**
 - Partial but predictive if applied appropriately
 - Usually involves Domain / Subject Matter Experts (SMEs)
 - Effective only insofar as review material is intelligible to SMEs while they are also accessing their expertise.
 - Often more effective using not just ontology artifact, but output of some application of ontology in SME-familiar context
 - Visualization, lexical content, NL mappings matter.
 - Many variants on NLP + application to corpus (e.g., indexing/markup, statistical analysis and comparison).
 - For broad domain & usage, effective lexical/conceptual distinction and handling of polysemy is necessary
 - Human, cognitive factors make/break effective SME Validation
 - **Personnel issue:**
 - Ontologists are not often trained in experimental & cognitive factors that affect validation accuracy.
 - If these factors are significant part of fitness for your purpose, spending resources here will save you much more later.

Approaching Ontology Quality Assurance

- **Ontology Requirements Specification**

- Analyze (SE) relationship of ontology to rest of system, identifying what ontology is expected to do and provide
 - NB: consider whether each expectation involves an ontology artifact; an ontological analysis process; an ontology-driven subsystem (reasoning, text analysis, data mediation...); all of the above. Surprisingly common source of error: skipped components
- Turning usage-dependent ontology quality factors into usable requirements

- **Ontology Quality Management**

- From requirements, identify metrics and measures that together will provide a good assessment of the fitness of an ontology for the usage
- Developing and implementing plans to evaluate the relevant quality factors at appropriate times.
- Planning ahead for resources (ontologists, SMEs, test corpora, other)
- Enlisting researcher help where pieces are missing (see next slide)

- **Don't skip!**

- *<Insert your favorite horror stories here>*

Good to know

- Beyond formal validation, ontology quality factors have been too much omitted in research
- This is changing
- See X-Track A1 subcollection in group library for a start on helpful resources <https://www.zotero.org/groups/ontologysummit2012/items/collecti onKey/TV2QAN4H> (and help grow it by adding resources you've found helpful)
- Check program of any major related conference; likely to see usage-inspired events, including workshops on ontology quality, evaluation, dynamics, modularity....
- Some missing metrics and methods are missing because no one has prioritized them yet, not because they are impossible.
 - Communication between systems engineers and ontology researchers has been insufficient from perspective of many researchers, too
 - Publicize your problematic use cases
 - Let a crowd of researchers know what you need and why; you may soon find that you have it.