



Comparing semantic technologies - strengths and weaknesses

David Norheim

Norwegian Semantic Days 2008



Outline



- Characteristics of semantic technologies
- Extending current technologies with semantics
- Where are the differences relevant?
- Comparing technologies
- An organization's need for semantic technologies



Based on...



The SIM Report

A Comparative Study of Semantic Technologies

The SIM Report

Page 1

...and ongoing work in Norstella Interop





“Semantics and all that stuff...”

“Does standardization help?”

“Why is XML successful and RDF not?”

“Wouldn't it be better for both parties to USE the reference model?”



...when something is about to become successful it's name tends to become overloaded...

...or was it just hype?





How can we say that a technology is “semantic”?

- Technologies that enable
 - explicit, unambiguous and shared definition of domain terms and relations (for humans and machines to interpret)
- And preferably
 - a global system for identification
 - a global system for reuse

Where are the differences relevant?

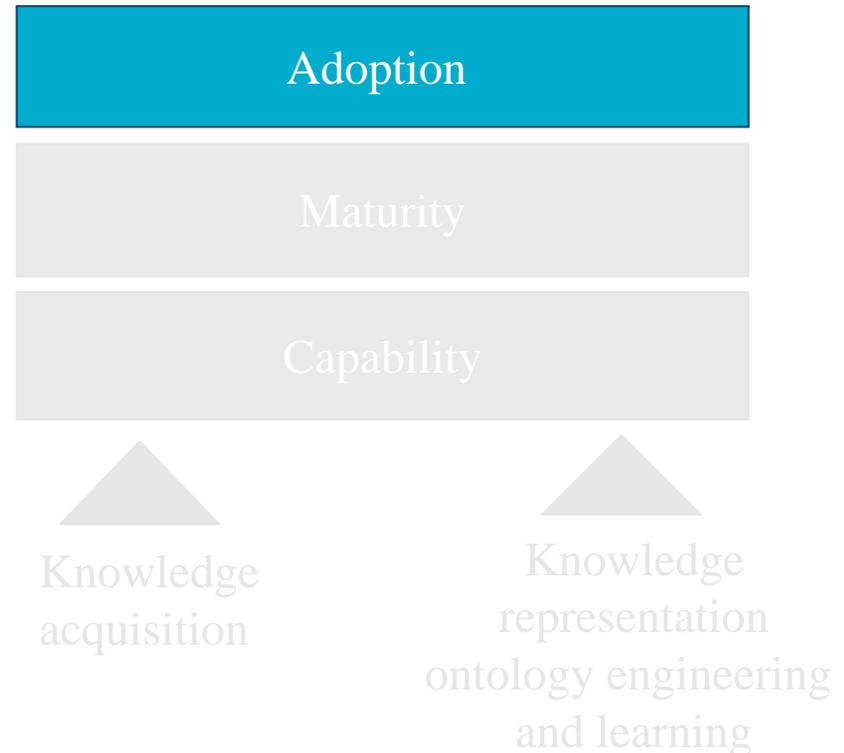


Use cases:

- information retrieval (search, navigation)
- information integration and fusion
- harmonization and documentation for interoperability (b2b)

Sector use of the technology

- reuse potential



Use case - Information integration and fusion



- Integration of data from various silos by providing a uniform access layer to heterogeneous data sources. Discovering new relations.
- Techniques involved: wrapping existing data sources, domain ontologies, reasoning



Use case - harmonization and documentation for semantic interoperability



- To ensure that the meaning is preserved in an information exchange across organizational boundaries by systemization of concepts internally and externally using explicit vocabularies in messages
- Techniques include: ontologies, ontology alignment, identifiers



Where is technology differences relevant?



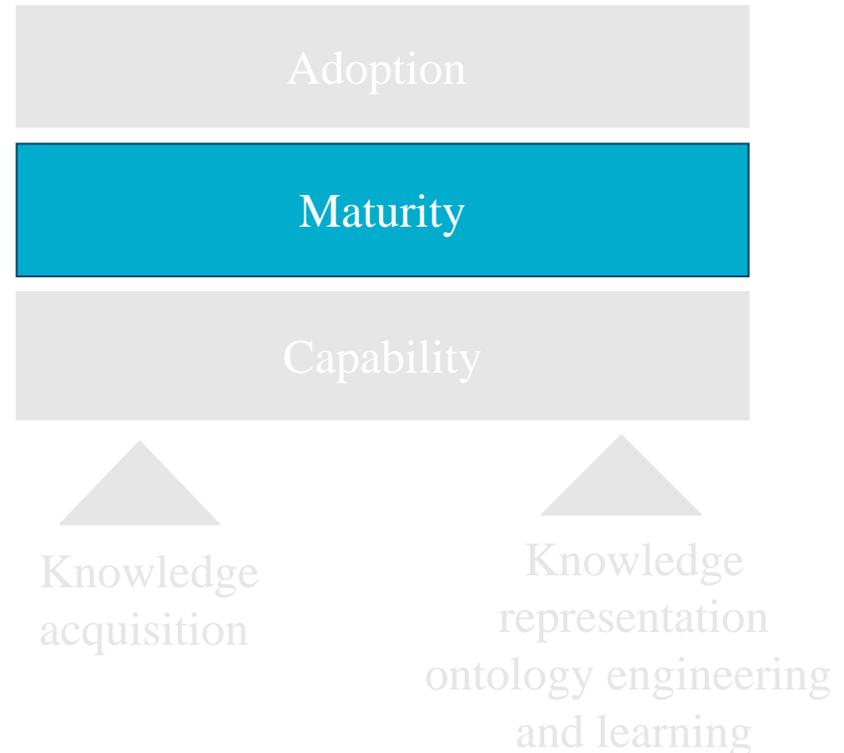
Maturity of technology

- standardization

Tool support and available competency

- developer tools
- scalability
- vendor adoption

Methods and experiences



Where is technology differences relevant?

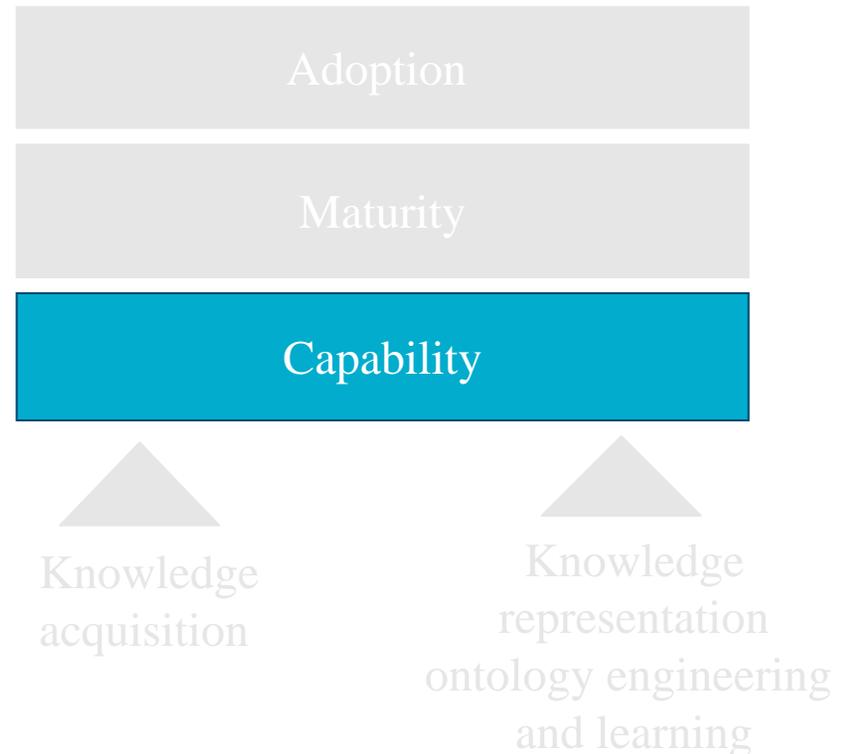


Semantic expressiveness

- logical foundation
- reasoning support

Sharing and extensibility:

- extension by explicit reuse
- extension by restriction

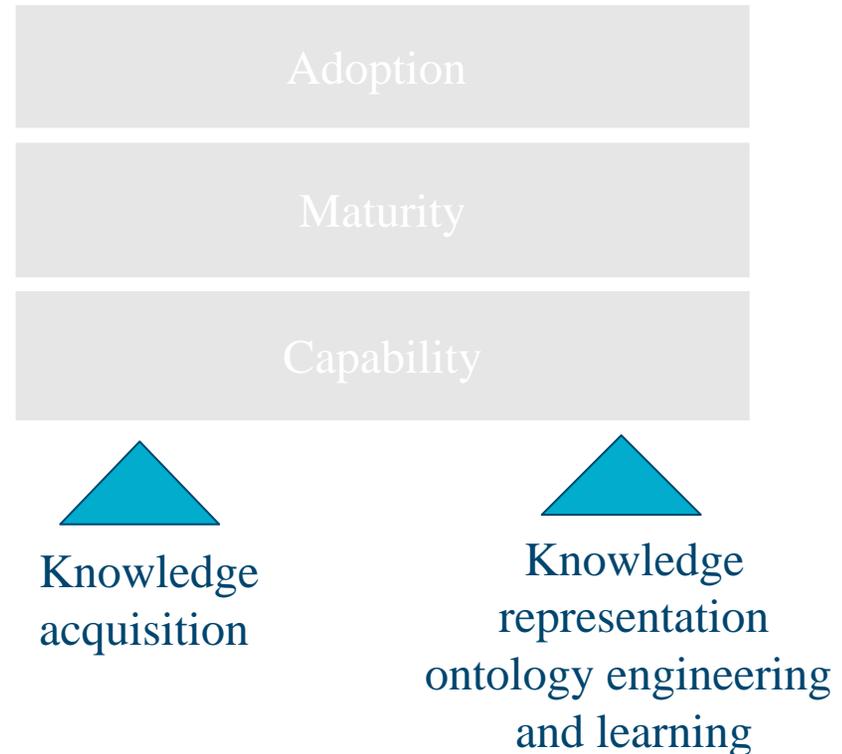


Where is technology differences relevant?



Knowledge acquisition and representation

- Techniques are relative independent of technology choice



Knowledge acquisition

Knowledge representation
ontology engineering
and learning

Technology “stacks” compared



- Selected technologies
 - W3C Semantic Web (RDF, OWL...)
 - ISO Topic Maps (TM)
 - ISO 15926
 - UN Core Components (CC)
 - OMG UML
- Are they comparable?
- How do we compare?
- Terminology: concepts and relations...



W3C Semantic Web (RDF and OWL)



- Effort from W3C to extend the Web creating a global database
- Global focus, a large set of common metadata vocabularies and open linked data. Focus on all use cases.
- Includes standards for global data representation (RDF), ontology language (OWL) and query language (SPARQL).
- Associated standards and techniques for wrapping legacy systems.
- A large set of tools in most areas, but with varying maturity
- Focus on data representation and formal ontologies, less focus on methodology

Adoption

Maturity

Capability

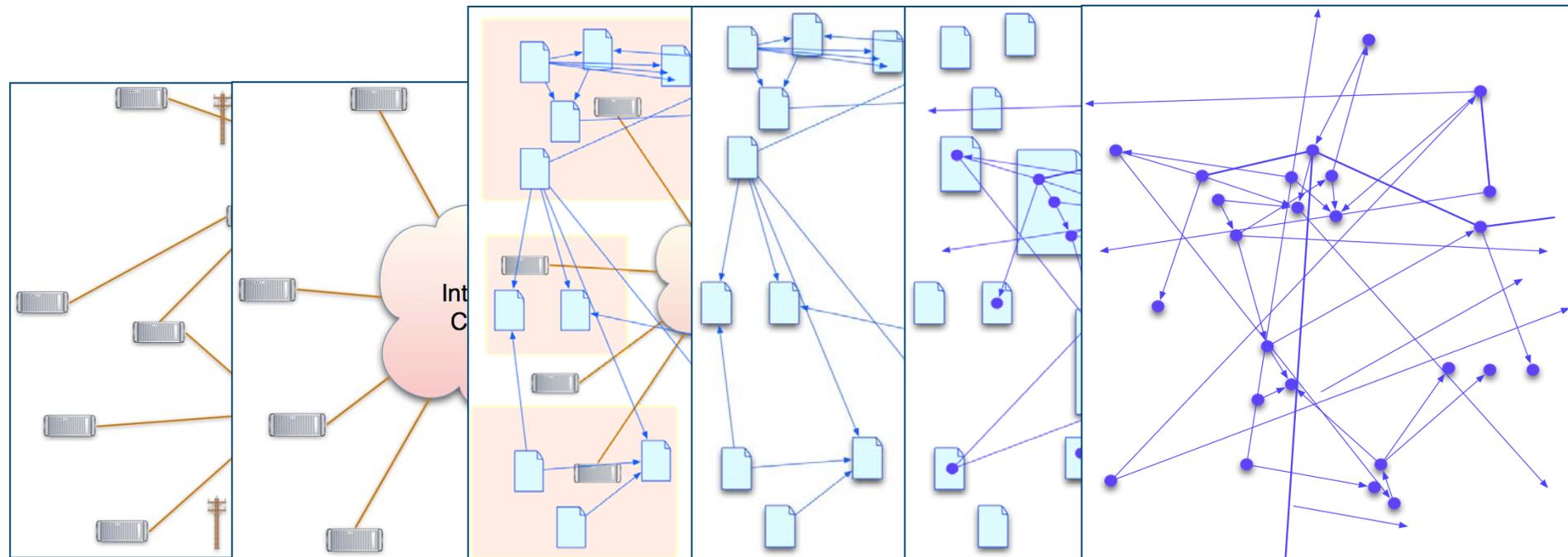


Internet -> WWW -> GGG



“Progress in communications technology has been characterized by a movement from lower to higher levels of abstraction.”

Tim Berners-Lee



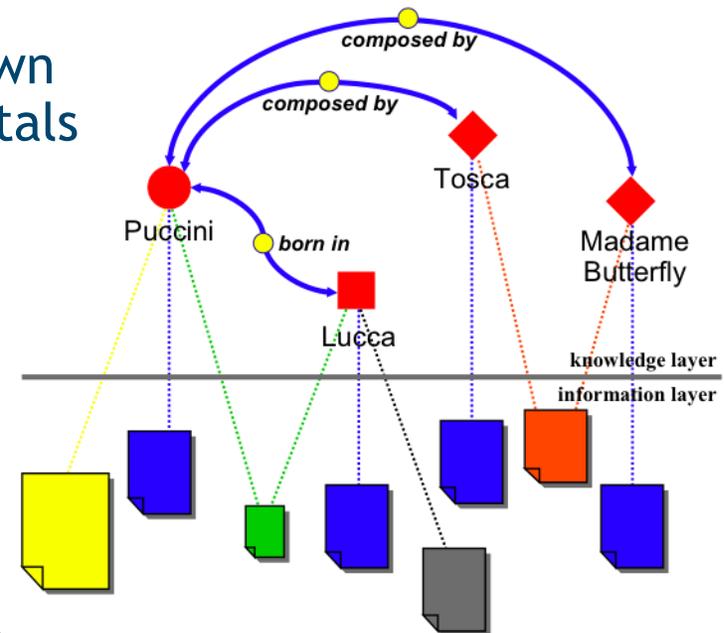
<http://www.w3.org/DesignIssues/Abstractions.html>



ISO Topic Maps



- Standard managed by ISO, most known for enabling relevance linking in portals
- Strong in the information retrieval use case, but also examples in information integration
- Strong base in Norway especially in public sector portals.
- Includes standards for data representation, ontology-language, and query language and tool support
- Associated techniques for integration with e.g. RDF
- Currently less focus on the constraint language (formal ontologies) than classification
- *Semantic technology stack and ontology language mainly for less formal use cases*



Adoption
Maturity
Capability

UN/CEFACT Core Components



- Managed by UN/CEFACT most known as subset of ebXML.
- Relevant for machine-to-machine transactions in the e-business domain (interoperability)
- Include standards for knowledge representation, one core ontology
- Limited tool support
- Extension by restriction (top-down)
- *an ontology and methodology for the e-business domain*

Adoption

Maturity

Capability



ISO 15926



- Managed by ISO, focus on life-cycle management of engineering objects.
 - Strongest adoption in the oil- and gas industry, but focus also on defense.
 - Specifies a generic data model and upper ontology (part 2), and an real-life objects (part 4 and RDL).
 - Standardized and managed by ISO and Posc Ceasar
 - Limited tool support, but OWL tools increasingly in use
 - Includes time and space in modeling methodology.
 - Increasingly align with with W3C Semantic Web standards
- *an ontology and methodology for lifecycle management*

Adoption

Maturity

Capability

OMG UML



- Managed by OMG is a standardized specification language for object modeling
- Strong in the field of software system modeling.
- Standardized, good tool support and maturity
- Model exchange in XML (models≈ontology)
- UML modeling notations used in semantic tools
- Semantics has not been the core of UML.

- *Language for modeling, but not a rich ontology language*

Adoption

Maturity

Capability

Adoption



- “Sector” adoption of technologies
 - Public sector (UML, CC, TM, SW)
 - Oil and Gas (ISO15926, SW)
 - Life science and health (SW)
 - Media and digital libraries (SW)
 - Web communities (SW)
 - Software development (UML)
- Where is my data?
 - Large amount through Open linked data (SW)
 - ...
- Availability of reusable vocabularies
 - Very Large in SW (Documents, People, Licenses, Thesauri, ...)
 - Moderate in TM (Documents/DCTM)
 - ISO 15926 and UN/CEFACT have a clear scope and reusability in their fields

Maturity of technology



- Is the technology stack complete and stable?
 - Most stacks are standardized and stable, but new versions will arrive
- Tool support during design time
 - Semantic Web, good especially in OWL DL
 - Topic Maps, good especially in topic hierarchies
 - UN/CEFACT, limited, but UML plug-ins
 - ISO15926, small but aligning with OWL tools
 - UML, very good
- Tool support at run time
 - Semantic Web, very good, several libraries, databases etc.
 - Topic Maps, a handful (e.g. ontopia, network planet, ovitas)
 - UN/CEFACT, limited
 - ISO15926, small but aligning with OWL
 - UML, N/A (?)

Maturity of technology



- Available competencies
 - All have strong advocates, but, few multi-disciplinary houses.
- Governance
 - In W3C Semantic Web, ISO Topic Maps and UML concrete ontologies/models are not developed and maintained in the organizations..
 - ISO 15926 and UN/CEFACT do define and maintain the concrete ontologies.
- Scalability
 - Semantic Web now in the area of tens of billions triples..., ontologies the size of >10k

Capability



- Currently OWL is the most expressive language with good tool support
- Also Semantic Web has a strong build-in reuse capability



Figure 3.4 Expressiveness shown in an ontology spectrum. Based on (Davis 2006)

How do I start?



- Understand your needs
 - Information retrieval?
 - Information integration and fusion?
 - Interoperability?
 - Other?
 - Ask competency questions
- Understand technologies
 - adoption
 - maturity
 - capabilities

How do I start?



- Try to reuse!
- Expect challenges with modeling methods
- Expect challenges with tools
 - Expect new more user friendly tools, e.g. semantic wikis
- Share your experiences!!
 - Many forums; e.g., this conference, Dataforeningen and Norstella.
- Is there a way to define an organization's maturity to take up semantic technologies?



Finally, a word of advise!



- Semantic technologies provide a menu!
- Be open to new tastes, but don't take on too much at once!
- But try!





Thank you!
david.norheim@computas.com

