Ontology for Big Data

**Story**

**Slides**

The Ontology Summit Operation in Perspective
- Slide 1 Ontology Summit 2014 Symposium
- Slide 2 Motivation ...
- Slide 3 The beginning ...
- Slide 4 Continuing to innovate ...
- Slide 5 What can I say, when I'm old enough to retire?
- Slide 6 References ...

Overview of Semantic Technologies and Ontologies
- Slide 1 The 6 W's of Semantic Technologies and Ontologies
- Slide 2 Topics for Discussion
- Slide 3 Background, Concepts and Standards
- Slide 4 Terms - Semantic Web
- Slide 5 Terms - Ontology
- Slide 6 Basic Concepts
- Slide 7 Semantic Web Standards "Layer Cake"
- Slide 8 Web Ontology Language (OWL)
- Slide 9 Semantic Web Rule Language (SWRL)
- Slide 10 SWRL (continued)
- Slide 11 SPARQL Query Language
- Slide 12 How/Why Are Ontologies Used?
- Slide 13 Semantic Search
- Slide 14 Mapping and Merging
- Slide 15 Knowledge Management
- Slide 16 Current Work and Future Directions
- Slide 17 BioPortal BioOntologies
- Slide 18 IBM Watson I
- Slide 19 IBM Watson II
- Slide 20 BBC, July 2010 Blog
Intelligent Information Management Tools in a Service-Oriented Software Environment

Papers About Ontology

The Value of Ontology-Based, Service-Oriented, Distributed Systems in a High Bandwidth Environment (with Steven J. Gollery), Collaborative Agent Design Research Center White Paper - GOLL-HBW (2002)


The Value of Ontology-Based, Service-Oriented, Distributed Systems in a High Bandwidth Environment (with Steven J. Gollery), Collaborative Agent Design Research Center White Paper - GOLL-HBW (2002)


Abstract

Need for Adaptive Planning Tools

Information-Centric vs. Data-Centric

Service-Oriented Architecture (SOA)

Figure 1: Principal components of a conceptual SOA implementation

Figure 2: Primary ESB components

Typical Service Requester and Service Provider Scenario

Figure 3: Conceptual Cloud operations

Business Process Management (BPM)

Figure 4: BPM design requirements

Figure 5: BPM design components

In Conclusion: Cloud Computing

Footnotes
E-MAPS on Ontology and Big Data

Ontology

Basic Value of Ontology

Warfighters, Ontology, and Stovepiped Data

The Operational Problem
This Document’s Contribution
Reality is not Segmented
Three Orders of Reality
Role of Words and Other Symbols
Creation and Use of Words and other Symbols
Avoid Conflation
Base Definition on Essential Properties
Non-Essential or Accidental Properties
Form Definitions Properly
Reality versus Convention
Relationships
Invest in Ontological Foundation
Purpose and Scope of Document Review
Conclusion
Story

Ontology for Big Data

The theme of this year's Ontology Summit was Big Data and Semantic Web Meet Applied Ontology and the highlights for me were:

- Peter Yim announced his retirement and the significant accomplishments which I asked him to elaborate on - see Research Notes below
- The only keynote presenter who has done ontology work is Dr. Phil Bourne who I have done a story on recently entitled Data Culture at NIH

I asked Andrea Westerinen and Gary-Berg Cross, active members of the Ontolog Forum, to present at our Federal Big Data Working Group Meetup as follows:

- Andrea Westerinen, Nine Points Solutions, Overview of Semantic Technologies and Ontologies: The 6 W's of Semantic Technologies and Ontologies Also see SlideShare
- INTEROP--Spatial Ontology Community of Practice: an Interdisciplinary Network to Support Geospatial Data Sharing, Integration, and Interoperability

Ontology development for Big Data needs automation with knowledge modeling tools like:

- Be Informed - See our Healthcare.gov example
- Two SIRA-based products; Research Assistant™ and Research Librarian™, Chuck Rehberg, Semantic Insights (limited beta test in process)
- Something Very Big Is Coming: Our Most Important Technology Project Yet—Stephen Wolfram Blog

Ontology development for Big Data needs work with actual big data as suggested by:

- Barry Smith - See our State Health Databases
- Eric Little - See our Modus Operandi
- Olivier Bodenreider and Tom Rindflesch - See our Semantic Medline and January Kickoff Meetup

So I need to find some more ontology and/or knowledge modeling work that uses big data from Peter Yim's responses and do some data science on both the ontology and the big data. A recent quote that I used in my story NSF Big Data Publications was: "With enough data you don't need semantic search. You can just use statistics." so let's see if that is true for some specific examples.

MORE TO FOLLOW
On a personal note, three thoughts come to mind:

- Of the three people who I worked with on Doug Engelbart's type of collaboration (Peter Yim, Susan Turnbull, and myself), I am now "the last leaf on the tree" so to speak;

- In answer to Peter Yim's question: What can I say, when I'm old enough to retire?, I said to myself: "I cannot do that because I have discovered a whole new exciting career as a data scientist/data journalist that builds on semantic technologies and ontologies, and am having too much fun at it to stop now."

- I forgot the third thought (maybe my age is showing), but I will think of it eventually:

So Peter, I wish you well in retirement, but I cannot say "Thank you and goodbye" yet.

However, I do have to say "Thank you and goodbye" to my good friend and esteemed colleague, George Thomas, who I gave an award for excellence to from the Federal CIO Council years ago, and hope that his many friends and admirers will do the same by signing George's Guest Book and viewing a video from April 13, 2012: "U.S. Chief Technology Officer Todd Park, David Forrest, Lead Project Manager, and George Thomas, the Chief Architect, discuss the future of Healthdata.gov."

---

Slides

The Ontology Summit Operation in Perspective

Slide 1 Ontology Summit 2014 Symposium

Ontology Summit 2014 Symposium
The Ontology Summit Operation in Perspective
Peter Yim
28 April 2014
Slide 2 Motivation ...

Motivation ...

- I asked the Chairs for this chance to speak
- want to let everyone know I am retiring
- will make an attempt to share a few stories I've learnt in the past 9 Ontology Summit seasons, my almost 15 years of working with the Ontology community, and over 4 decades of working with information systems
- and, of course, want to the take the opportunity to say “Thank You and Good Bye!”

Slide 3 The beginning ...

http://ontolog.cim3.net/cgi-bin/wiki...OntologySummit

The beginning ...

- 2006 Upper Ontology Summit
  See: http://ontolog.cim3.net/cgi-bin/wiki?UpperOntologySummit
- The motivation: “we are part of the problem”

The players
- Co-convener: 4  (Cassidy, Obrist, Ray, Yim)
- Co-organizers: 2  (ONTOSLOG and NIST)
- the stewards of all open upper ontologies: BFO, CYC, DOLCE, DOLCE-D3S, GUM, ISO-15926, PSL, SUMO.

... but no, the Heidelberg meeting of 1998 was really the “first” Ontology Summit

Slide 4 Continuing to innovate ...

http://ontolog.cim3.net/cgi-bin/wiki...OntologySummit
Continuing to innovate ...

- Fast forward: almost 9 season under our belts now
  see: http://ontolog.cim3.net/cgi-bin/wiki?OntologySummit
- clever choice of themes
- more co-organizers, more organizing team members, more contributors, more participants, reaching out to more
- organization and process innovations
  - from a 2-day conference, to a 4 month series of both virtual and on-site events
  - from just the Ontology community, to reaching out and engaging potential collaborators
  - establishing a model in what a broader networked community of practice might accomplish

Slide 5 What can I say, when I'm old enough to retire?

What can I say, when I'm old enough to retire?

- I've been blessed ... and very lucky
- I've contributed: some time, work, money ... and lots of passion
- Want money, jobs, ... change the world?
  They're just 'round the corner, but "Are We ready?"
- Note these significant accomplishments:
  W3C-RDF, ISO-CommonLogic, BioPortal, Apple-Siri, IBM-Watson,
  GoodRelations Ontology, Google Knowledge Graph, schema.org, ... 
- I've seen it before ... magic does happen!
- ... keep dreaming; don't short change your vision
- Thank you and good bye!
  ... It's been great knowing and working with you all.
- I'm most proud to have helped keep it "free and open" for the community

Slide 6 References ...

http://ontolog.cim3.net/cgi-bin/wiki...OntologySummit

http://ontolog.cim3.net/file/work/On...s_20140116.pdf

http://ontolog.cim3.net/cgi-bin/wiki...ents_Synthesis

http://ontolog.cim3.net/file/work/SI...m_20100325.pdf


http://ontolog.cim3.net/file/work/pr..._20140116a.pdf
Overview of Semantic Technologies and Ontologies

Slides

Slide 1 The 6 W's of Semantic Technologies and Ontologies

The 6 W’s of Semantic Technologies and Ontologies

Who, What, Where, When, Why and how

Andrea Westerinen
andrew@ninepts.com

© Nine Points Solutions, LLC 2012-2013
Slide 2 Topics for Discussion

Topics for Discussion

- Background, concepts and standards
  - What and when (from Ancient Greece to today)
- How are ontologies used?
  - How and for what
- Current work and future directions
  - Providing some specific who’s, what’s, where’s, when’s and why’s

Slide 3 Background, Concepts and Standards

Background, Concepts and Standards

- Semantic Technologies
  - The “web” part of the Semantic Web
  - Using triples
- Semantic Technology Standards
  - RDF and RDF Schema : Linked Data
  - OWL : Ontologies
  - SWRL : Inference Rules
  - SPARQL : Query
Slide 4 Terms - Semantic Web

Terms - Semantic Web

- Term introduced in Scientific American article in 2001, authored by Tim Berners-Lee, James Hendler and Ora Lassila

"Most of the Web's content today is designed for humans to read, not for computer programs to manipulate meaningfully. ... here a header, there a link to another page—but in general, computers have no reliable way to process the semantics ... The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

© Nine Points Solutions, LLC 2012-2013

Slide 5 Terms - Ontology

Terms - Ontology

- Based in philosophy and the desire to understand and categorize reality
  - Ancient Greece and Parmenides (circa 540-515 BC)
- Evolved to use in computer science
  - To describe models of the concepts, relationships and axioms of a domain
- Tom Gruber's 1993 papers defined and popularized current usage
  - A Translation Approach to Portable Ontology Specifications
  - Toward Principles for the Design of Ontologies Used for Knowledge Sharing

© Nine Points Solutions, LLC 2012-2013
Slide 6 Basic Concepts

Basic Concepts

- Linked Data
  - Relationships between resources
- Standardized by W3C in RDF
  - Resource Description Framework
  - Built on URIs (standard naming – for ex, http://www.google.com)
  - Statements are triples
    - Subject, Relationship (Predicate), Object
    - Multiple encodings are possible
      - XML, annotation of HTML, etc...

© Nine Points Solutions, LLC 2012-2013

Slide 7 Semantic Web Standards "Layer Cake"

Semantic Web Standards “Layer Cake”

© Nine Points Solutions, LLC 2012-2013
Slide 8 Web Ontology Language (OWL)

Web Ontology Language (OWL)

- Start with RDF and RDF-Schema (RDF-S)
  - Adds subclassing, domains/ranges of properties, ... to RDF
- Move onto OWL
  - Adds first-order-logic
    - Union, intersection, complement, equivalent, disjoint, symmetric, transitive, ...

```
owl:intersectionOf(ex:Action, ex:pollution)
```

© Nine Points Solutions, LLC 2013-2015

Slide 9 Semantic Web Rule Language (SWRL)

Semantic Web Rule Language (SWRL)

- Combines parts of OWL and Datalog (a subset of Prolog)
- Rules are of the form:
  - If body (antecedent) then assert head (consequent)

We learn that Joe is Mary’s uncle ...

Via a SWRL rule and a semantic reasoner inference

```
hasParent(?x1,?x2), hasBrother(?x2,?x3) -> hasUncle(?x1,?x3)
```

Where "?" indicates a variable

So, if variable x1=Mary, x2=John and x3=Joe

Then x1 (Mary) has an uncle x3 (Joe)

© Nine Points Solutions, LLC 2013-2015
Slide 10 SWRL (continued)

SWRL (Continued)

- Built-in functions for comparisons, mathematical manipulation, and datatype processing
- Can be exchanged via RIF (Rule Interchange Format)

We learn that John and Joe are adults ...

Person(?p), hasAge(?p, ?age),
swrlb:greaterThan(?age, 17) -> Adult(?p)

Slide 11 SPARQL Query Language

SPARQL Query Language

- Can FILTER, return DISTINCT values, etc. as with SQL
- Different results can be returned
  - SELECT returns variables and their values
  - ASK tests whether a query pattern has a solution, returns T/F
  - CONSTRUCT creates a set of triples/an RDF graph
  - DESCRIBE returns a set of triples/an RDF graph about resources

© Nine Points Solutions, LLC 2012-2013
Slide 12 How/Why Are Ontologies Used?

How/Why Are Ontologies Used?

- Semantic search and query expansion
  - When you get too many or too few results
  - When search terms are too vague or too specific
- Mapping and merging data
  - Lots of implicit and explicit vocabularies
  - Data has to flow across groups and is used in business processes
  - Needs to be exchanged and interoperable
- Knowledge Management

Slide 13 Semantic Search

http://company.hakia.com/techoverview.html  
http://semanticweb.com/semantic-sear...-basics_b36678  
http://www.gopubmed.org/web/gopubmed/

Semantic Search

- Some examples:
  - Hakia,  
    http://company.hakia.com/techoverview.html
  - Google search,  
    http://semanticweb.com/semantic-search-at-google-the-seo-basics_b36678
  - PubMed,  
    http://www.gopubmed.org/web/gopubmed/

Slide 14 Mapping and Merging

http://semanticweb.com/webcast-fight...-data-2_b17051
Mapping and Merging

- Fighting Fires with the Semantic Web
  - [http://semanticweb.com/valtime/digging-fires-with-linked-data-0_31703]
- Problem – Wide variety of data from numerous sources
  - Police emergency DBs, construction road reports, geospatial DBs, fire department DBs/spreadsheets, ...
- Solution - An infrastructure based on accessing raw (RDF) data
  - Data loaded in a generic form (RDF triples)
    - To retrieve and utilize any relevant information
  - Data accessed and assembled on the fly
    - URI and its meta-data defined how data was displayed
    - No need for multiple applications with their own UIs, systems, etc. - One display for all data
      - “We already had 8 screens in the fire truck... We didn’t need 9”

© Nine Points Solutions, LLC 2012-2015

Slide 15 Knowledge Management

Knowledge Management

- Ontologies to organize and search, and to infer new knowledge

© Nine Points Solutions, LLC 2012-2015

Slide 16 Current Work and Future Directions

Current Work and Future Directions

- BioPortal BioOntologies
- IBM Watson
- BBC and the World Cup
- Google
- Reuters
- Bechtel Corporation
- Wells Fargo
IBM Watson (II)

- Open domain Question Answering (QA)
- DeepQA project
  - Integration of Natural Language Processing, Information Retrieval, Machine Learning, Knowledge Representation and Reasoning, and massively parallel computation
- Semantic Web Technology
  - Enhanced by open linked data
    - For example, Dbpedia and Geonames
    - Support in triple stores
    - Generate candidate answers and score them under “multiple points of view” such as type coercion and geographic proximity
  - Web provided training data via the connection between linked data and natural language text in Wikipedia
    - For relation detection and entity recognition systems

© Nine Points Solutions, LLC 2012-2013

---

BBC, July 2010 Blog

- Query Expansion
- World Cup 2010 website
  - “… deeper and richer use of content than … traditional CMS-driven publishing solutions”
- Metadata
  - “… framework … publishes data about the content - metadata. … By querying this published metadata we … create dynamic page aggregations for teams, groups and players.”
- Ontology
  - Concepts, individual entities, groups and relationships for the World Cup, plus additional journalist-authored data (stories, blog entries, statistics, …)
Slide 21 Google, March 2009 Blog

Google, Mar 2009 Blog

- Semantic search
  - "...we’re deploying a new technology that can better understand associations and concepts related to your search, and one of its first applications lets us offer you even more useful related searches. For example, if you search for [principles of physics], our algorithms understand that “angular momentum,”” “special relativity,” “big bang” and “quantum mechanic” are related terms that could help you find what you need.”

© Nine Points Solutions, LLC 2012-2013

Slide 22 Reuters

Reuters

- Eikon for the Trading Industry
  - "Maps/merges data “that in the past was on disparate products, each with a different look and feel... We can logically bring together information in seconds that might have taken a user hours before.”

- Calais
  - Web service and API
  - Discovers entities, events and relations in submitted content and attaches semantic metadata
  - Uses natural language processing, machine learning and other methods, to categorizes and links data with entities (people, places, organizations, etc.), facts, and events
  - Returns the results in RDF which are also centrally stored and identified by a GUID
  - Using the GUID, a consumer can retrieve the metadata via a call to Calais

© Nine Points Solutions, LLC 2012-2013

Slide 23 Bechtel - iRing Mapping and Merging

http://iringtoday.com
Slide 24 Wells Fargo and FBO

Wells Fargo and the Financial Industry Business Ontology (FIBO)
- Collaborative effort across the financial industry, regulatory community, academic institutions and semantic technology vendors
- Open standard ontologies for the financial system
- From the SemTech 2013 conference session:
  "Following the 2008 global financial crisis it became clear that the absence of certain data standards led to a lack of awareness of the magnitude of risk that the industry was exposed to. The goal of FIBO is to introduce semantic capabilities that will better enable common terminology for business entities, financial contracts and instruments, improve data transparency, improve data integration and linkage, and improve financial reporting and analytics, so that greater “data” quality and health can be introduced into the financial system."

Spotfire Dashboard

Research Notes

Peter, Again congratulations on your Ontology Summit 2014 and Ontolog Forum contributions.

In preparation for our June Federal Big Data Working Group Meetups, for which I have invited several members of Ontolog to provide tutorials on Semantic Technologies and Ontology Development, I would like to invite you to present (or at least elaborate) on the list of accomplishments that were in your slides.

What I have in mind is a link to a presentation for each, the role of ontology in each, and the role of the Ontolog Form in the development of that ontology.
I also have been asked to cover the following questions:

What will happen with all the wiki content (you mentioned meeting with the people you provide services to to discuss that)?

Will the Forum continue its work in some other organization (I heard mention of the Applied Ontology…. that Leo is involved in so maybe I should ask him)?

What are you plans for retirement?

Is there anything else you would like to present that addresses the role of ontology in semantic web technologies and especially big data?

Thank you in advance for your response and your cooperation in providing content of interest to our Meetup Members.

Best regards, Brand

Brand, Thank you for this kind message, and for joining us at the Summit Symposium (sorry, we didn't have a chance to even say "hello!")

> [BN] I would like to invite you to present ...

[ppy] since I am going into my retirement, I am now more focused in pulling myself out, rather than developing new speaking gigs.

Therefore, I am not an appropriate candidate speaker for you. LeoObrst co-convened Ontolog, OntologySummit, OOR, etc. with me, and has been a key driver of the content (I was only providing support on the side.) he was one of the General Co-chairs of this OntologySummit on "Big Data and Semantic Web Meet Applied Ontology;" so, maybe he should be your target speaker.

> [BN] (... at least elaborate) on the list of accomplishments that were in your slides.

[ppy] I assume you mean the bullet on slide#5 of my Symposium Remarks (http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2014_Symposium#nid4CUG)

) ...

//

* Note these significant accomplishments:

W3C-RDF, ISO-CommonLogic, BioPortal, Apple-Siri, IBM-Watson,

GoodRelations Ontology, Google Knowledge Graph, schema.org, ...

//

First of all, in the context of the presentation, the bullet should read "Note these significant accomplishments by the Ontology community, collectively, towards "advancing the field of ontology, ontological engineering and semantic

More details ref. significant involvement by various individuals (from our community) and references from the Ontolog archives, include:

* W3C-RDF

** significant contribution from BobSchloss, R_V_Guha, PatHayes, et al. on RDF (1.0)

*** ref. http://www.w3.org/TR/WD-rdf-syntax-971002/

** significant contribution from PatHayes and Peter Patel-Schneider, et al. on RDF 1.1 Semantics

*** ref. http://www.w3.org/TR/2014/REC-rdf11-mt-20140225/

** significant contribution from R_V_Guha, DanBrickley, et al. on RDF Schema 1.1

*** ref. http://www.w3.org/TR/WD-rdf-syntax-971002/

* ISO-CommonLogic

** significant contribution by ChrisMenzel, PatHayes, JohnSowa, with Editor: to the current standard: ISO/IEC IS 24707:2007


** significant contribution by by MichaelGruninger (editor), FabianNeuhaus, TaraAthan, JohnSowa, et al. to the ongoing work to update the CL standard (dubbed "CLv2")


* BioPortal

** significant contribution by MarkMusen, RayFergerson, PaulAlexander


(esp. MarkMusen's talk) and the 4 sessions under http://ontolog.cim3.net/cgi-bin/wiki.pl?OOR/ConferenceCall_2013_12_10#nid42SL
* Apple-Siri

** significant contribution by AdamCheyer and TomGruber


* IBM-Watson

** significant contribution by DaveFerrucci, ChrisWelty et al.

*** ref http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2006_05_11

& http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2014_01_30#nid455F

* the GoodRelations Ontology

** significant contribution by MartinHepp

*** ref. http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2008_10_16#nid1NDM


* Google Knowledge Graph

** ref. discussion thread:

http://ontolog.cim3.net/forum/ontolog-forum/2012-05/threads.html#00028

* schema.org

** significant contribution by R_V_Guha, DanBrickley, et al.


& http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2014_03_27#nid4AOT

... you can generally pick up on the role of ontology in each from the cited references. Ontolog Form (being just a "water cooler conversation") does not have a role in these accomplishments or the development of their ontologies ... its the cited individuals, their institutions and the team of people involved that accomplished them.

> [BN] What will happen with all the wiki content ... (you mentioned meeting with the people you provide services to to discuss that)? ... Will the Forum continue its work in some other organization ... What are you plans for retirement?

[ppy] I am just starting to explore these ... I'll find out more in the next few weeks.

> [BN] anything else ... that addresses the role of ontology in semantic web technologies and especially big data?
Ken Baclawski (and Anne Thessen) whom you know well, championed our Track on "Tackling the Variety Problem in Big Data" for Ontology Summit 2014. That track has pooled together significant work and insight on the subject. You might consider asking Ken to present.

* Ref.

** http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2014_02_13

** http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2014_03_27

** http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2014_Tackling_Variety_In_BigData_Synthesis

Best wishes to your event.

Regards. =ppy

---

**Intelligent Information Management Tools in a Service-Oriented Software Environment**

Source: [http://works.bepress.com/jpohl/90](http://works.bepress.com/jpohl/90)

[http://digitalcommons.calpoly.edu/cg...&context=cadrc](http://digitalcommons.calpoly.edu/cg...&context=cadrc) (PDF)

Plenary Session Keynote Paper: InterSymp-2009, Baden-Baden, Germany, 3-7 August, 2009 [RESU98]

Jens Pohl, Ph.D. Executive Director, Collaborative Agent Design Research Center (CADRC)

California Polytechnic State University (Cal Poly)

San Luis Obispo, California, USA

---

**Papers About Ontology**

My Note: See [http://works.bepress.com/jpohl/subje...html#Articles](http://works.bepress.com/jpohl/subje...html#Articles) and Google Chrome Find: Ontology. See 8 hits below:

**The Value of Ontology-Based, Service-Oriented, Distributed Systems in a High Bandwidth Environment** (with Steven J. Gollery), *Collaborative Agent Design Research Center White Paper - GOLL-HBW (2002)*

The TRANSWAY® software application is an adaptive, ontology-based toolset with collaborative agents, designed to assist...

This paper proposes the construction of a Conceptual Model as a logical step prior to...

The Value of Ontology-Based, Service-Oriented, Distributed Systems in a High Bandwidth Environment (with Steven J. Gollery), *Collaborative Agent Design Research Center White Paper - GOLL-HBW* (2002)


The purpose of this paper is to provide an introduction to the OWL Web ontology...


In San Luis Obispo we have seventy-seven SEAWAY systems which are being prepared for fielding,...


This paper provides a basic description of the concept of an ontology. It then describes...


Traditional approaches to building intelligent information systems employ an ontology to define a representational structure...


The support of context-level interoperability demands increasing attention in today’s arena of semantics-oriented decision-support systems....

**Abstract**

This paper draws attention to the increasing need for agile and adaptive software environments that are capable of supporting rapid re-planning during the execution of time-critical operations involving commercial end-to-end supply chain transaction sequences, as well as disaster response and military missions. It is argued that such environments are currently best served by information-centric software tools executing within a service-oriented paradigm. Service-oriented architecture (SOA) design concepts and principles are described, with a focus on the functions of the services management framework (SMF) and enterprise service bus (ESB) components. Differentiating between data-centric and
information-centric services, it is suggested that only intelligent software services, particularly those that incorporate an internal representation of context in the form of an ontology and agents with reasoning capabilities, are able to effectively address the need for agile and adaptive planning, re-planning and decision-support tools.

The paper concludes with a description of the design components of a business process management (BPM) system operating within a SOA-based infrastructure, followed by a brief discussion of Cloud computing promises and potential user concerns.

Keywords: adaptive, agile, APEX, cloud computing, BPEL, business process execution language, BPM, business process management, choreographer, data-centric, enterprise service bus, ESB, information-centric, mediator, registry, services management framework, SMF, service-oriented architecture, SOA

Need for Adaptive Planning Tools

There is an increasing need in industry and government for planners and decision-makers to be able to rapidly re-plan during execution. Experience has shown that the best-laid plans will likely have to be changed during implementation. Operational environments are often impacted by events or combinations of factors that were either not foreseen during the planning stage or were thought to be unlikely to occur. In commerce, where just in time inventories have become an acknowledged cost-saving measure, suppliers and shippers are particularly vulnerable to the disruption of end-to-end supply chain sequences, such as inclement weather conditions, traffic congestion, accidents, equipment malfunction, and human error.

Military commanders, who often deal with extremely time-critical and human life endangering operations have learned from bitter experience that agile planning tools are essential for their ability to rapidly adapt to changing mission conditions. It can be argued that an information management environment, with an agile planning capability of the type implied by the stated objectives of the Adaptive Planning and Execution (APEX) process recently adopted by the U.S. military forces, requires both the ability to automatically interpret data in context and the flexibility to provide access to decision-support tools regardless of whether these are part of the same software application or another application.

This argument is based on the definition of agility as the ability to rapidly adapt to changing conditions, and has two implications. First, in a real world environment the operational data that enter a particular application may not adhere exactly to the specifications on which the design of the software was originally based. An agile software application will therefore need to have the ability to automatically interpret the incoming data within the appropriate context and make the necessary processing adjustments. Second, under such dynamic conditions it is likely that the user will have a need for tools that were not foreseen during the design of the application and are therefore not available. An agile software environment will therefore have to provide access to a wide range of tools, at least some of which may not be an integral component of the particular application that the operator is currently using. This suggests a system environment in which software tools can be seamlessly accessed across normal application domain boundaries. This is the objective of an information management environment that is based on the service-oriented concepts and principles described in this paper.
Information-Centric vs. Data-Centric

There are several reasons why computer software must increasingly incorporate more and more intelligent capabilities (Pohl 2005). Perhaps the most compelling of these reasons relates to the current data-processing bottleneck. Advancements in computer technology over the past several decades have made it possible to store vast amounts of data in electronic form. Based on past manual information handling practices and implicit acceptance of the principle that the interpretation of data into information and knowledge is the responsibility of the human operators of the computer-based data storage devices, emphasis was placed on storage efficiency rather than processing effectiveness. Typically, data file and database management methodologies focused on the storage, retrieval and manipulation of data transactions, rather than the context within which the collected data would later become useful in planning, monitoring, assessment, and decision-making tasks.

The term information-centric refers to the representation of information, as it is available to software modules, not to the way it is actually stored in a digital machine. This distinction between representation and storage is important, and relevant far beyond the realm of computers. When we write a note with a pencil on a sheet of paper, the content (i.e., meaning) of the note is unrelated to the storage device. A sheet of paper is designed to be a very efficient storage medium that can be easily stacked in sets of hundreds, filed in folders, folded, bound into volumes, and so on. As such, representation can exist at varying levels of abstraction. The lowest level of representation considered is wrapped data. Wrapped data consists of low-level data, for example a textual e-mail message that is placed inside some sort of an e-mail message object. While it could be argued that the e-mail message is thereby objectified it is clear that the only objectification resides in the shell that contains the data and not the e-mail content. The message is still in a data-centric form offering a limited opportunity for interpretation by software components.

A higher level of representation endeavors to describe aspects of a domain as collections of interrelated, constrained objects. This level of representation is commonly referred to as an information-centric ontology. At this level of representation context can begin to be captured and represented in a manner supportive of software-based reasoning. This level of representation context can begin to be captured and represented in a manner supportive of software-based reasoning. This level of representation (i.e., context) is an empowering design principle that allows software to undertake the interpretation of operational data changes within the context provided by the internal information model (i.e., ontology).

Even before the advent of the Internet and the widespread promulgation of SOA concepts it was considered good software design and engineering practice to build distributed software systems of loosely coupled modules that are able to collaborate by subscription to a shared information model. The principles and corresponding capabilities that enable these software modules to function as decoupled services include (Pohl 2007):

- An internal information model that provides a usable representation of the application domain in which the service is being offered. In other words, the context provided by the internal information model must be adequate for the software application (i.e., service) to perform as a useful adaptive set of tools in its area of expertise.
- The ability to reason about events within the context provided by the internal information model. These reasoning capabilities may extend beyond the ability to render application domain related services to the performance of self-monitoring maintenance and related operational efficiency tasks.
- Facilities that allow the service to subscribe to other internal services and understand the nature and capabilities of these resources based on its internal information model.
- The ability of a service to understand the notion of intent (i.e., goals and objectives) and undertake self-activated tasks to satisfy its intent. Within the current state-of-the-art this capability is largely limited by the degree of context that is provided by the internal information model.
Additional capabilities that are not yet able to be realized in production systems due to technical limitations, but have been demonstrated in the laboratory environment, include: the ability of a service to learn through the acquisition and merging of information fragments obtained from external sources with its own internal information model (i.e., dynamically extensible information models); extension of the internal information model to include the internal operational domain of the software application itself and the role of the service within the external environment; and, the ability of a service to increase its capabilities by either generating new tools (e.g., creating new agents or cloning existing agents) or automatically searching for external assistance.

Service-Oriented Architecture (SOA)

The notion of service-oriented is ubiquitous. Everywhere we see countless examples of tasks being performed by a combination of services, which are able to interoperate in a manner that results in the achievement of a desired objective. Typically, each of these services is not only reusable but also sufficiently decoupled from the final objective to be useful for the performance of several somewhat similar tasks that may lead to quite different results. For example, a common knife can be used in the kitchen for preparing vegetables, or for peeling an orange, or for physical combat, or as a makeshift screwdriver. In each case the service provided by the knife is only one of the services that are required to complete the task. Clearly, the ability to design and implement a complex process through the application of many specialized services in a particular sequence has been responsible for most of mankind's achievements in the physical world. The key to the success of this approach is the interface, which allows each service to be utilized in a manner that ensures that the end-product of one service becomes the starting point of another service.

Figure 1: Principal components of a conceptual SOA implementation

In the software domain these same concepts have gradually led to the adoption of Service- Oriented Architecture (SOA) principles. While SOA is by no means a new concept in the software industry it was not until Web services became available that these concepts could be readily implemented (Erl 2005). In the broadest sense SOA is a software framework for computational resources to provide services to customers, such as other services or users. The Organization for the Advancement of Structured Information (OASIS) defines SOA as a “…paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains” and “…provides a uniform means to offer, discover, interact with and use capabilities to produce desired effects with measurable preconditions and expectations”. This definition underscores the fundamental intent that is embodied in the SOA.
paradigm, namely flexibility. To be as flexible as possible a SOA environment is highly modular, platform independent, compliant with standards, and incorporates mechanisms for identifying, categorizing, provisioning, delivering, and monitoring services.

The principal components of a conceptual SOA implementation scheme (Figure 1) include a Services Management Framework (SMF), various kinds of foundational services that allow the SMF to perform its management functions, one or more portals to external clients, and the enterprise services that facilitate the ability of the user community to perform its operational tasks.

**Services Management Framework (SMF):** A Services Management Framework (SMF) is essentially a SOA-based software infrastructure that utilizes tools to manage the exchange of messages among enterprise services. The messages may contain requests for services, data, the results of services performed, or any combination of these. The tools are often referred to as foundational services because they are vital to the ability of the SMF to perform its management functions, even though they are largely hidden from the user community. The SMF must be capable of:

- Undertaking any transformation, orchestration, coordination, and security actions necessary for the effective exchange of the message
- Maintaining a loosely coupled environment in which neither the service requesters nor the service providers need to communicate directly with each other; or even have knowledge of each other.

A SMF may accomplish some of its functions through an Enterprise Service Bus (ESB), or it may be implemented entirely as an ESB.

**Enterprise Service Bus (ESB):** The concept of an Enterprise Service Bus (ESB) greatly facilitates a SOA implementation by providing specifications for the coherent management of services. The ESB provides the communication bridge that manages the exchange of messages among services, although the services do not necessarily know anything about each other. According to Erl (2005) ESB specifications typically define the following kinds of message management capabilities:

- **Routing:** The ability to channel a service request to a particular service provider based on some routing criteria (e.g., static or deterministic, content-based, policy-based, rule-based).
- **Protocol Transformation:** The ability to seamlessly transform the sender’s message protocol to the receiver’s message protocol.
- **Message Transformation:** The ability to convert the structure and format of a message to match the requirements of the receiver.
- **Message Enhancement:** The ability to modify or add to a sender’s message to match the content expectations of the receiver.
- **Service Mapping:** The ability to translate a logical business service request into the corresponding physical implementation by providing the location and binding information of the service provider.
- **Message Processing:** The ability to accept a service request and ensure delivery of either the message of a service provider or an error message back to the sender. Requires a queuing capability to prevent the loss of messages.
- **Process Choreography and Orchestration:** The ability to manage multiple services to coordinate a single business service request (i.e., choreograph), including the implementation (i.e., orchestrate). An ESB may utilize a Business Process Execution Language (BPEL) to facilitate the choreographing.
- **Transaction Management:** The ability to manage a service request that involves multiple service providers, so that each service provider can process its portion of the request without regard to the other parts of the request.
• **Access Control and Security**: The ability to provide some level of access control to protect enterprise services from unauthorized messages. There are quite a number of commercial off-the-shelf (COTS) ESB implementations that satisfy these specifications to varying degrees. A full ESB implementation would include four distinct components (Figure 2): Mediator; Service Registry; Choreographer; and, Rules Engine. The Mediator serves as the entry point for all messages and has by far the largest number of message management responsibilities. It is responsible for routing, communication, message transformation, message enhancement, protocol transformation, message processing, error handling, service orchestration, transaction management, and access control (security).

The Service Registry provides the service mapping information (i.e., the location and binding of each service) to the Mediator. The Choreographer is responsible for the coordination of complex business processes that require the participation of multiple service providers. In some ESB implementations the Choreographer may also serve as an entry point to the ESB. In that case it assumes the additional responsibilities of message processing, transaction management, and access control (security). The Rules Engine provides the logic that is required for the routing, transformation and enhancement of messages. Clearly, the presence of such an engine in combination with an inferencing capability provides a great deal of scope for adding higher levels of intelligence to an ESB implementation.

**Figure 2: Primary ESB components**

![Figure 2: Primary ESB components](image)

**Typical Service Requester and Service Provider Scenario**

The following sequence of conceptual steps that must be taken by the SMF to support a SOA system environment is not inclusive of every variance that might occur. It is intended to provide a brief description of the principal interactions involved (Figure 3).

While the Service Requester knows that the Mediator is the entry point of the ESB component of the SMF and what bindings (i.e., protocols) are supported by the Mediator, it does not know which Service Provider will satisfy the request because it knows nothing about any of the other enterprise services that are accessible through the Mediator. Therefore, the conceptual SOA-based infrastructure shown in Figure 1 is often referred to as a Cloud.

The Mediator is clearly in control and calls upon the other primary components of the ESB if and when it requires their services. It requests the handle (i.e., location and mappings) of the potential Service Providers from the Service Registry. If there are multiple Service Provider candidates then it will have to select one of these in Step (6) to provide the requested service. The Mediator will invoke any of the foundational services in the SMF to validate (i.e., access
control), translate, transform, enhance, and route the message to the selected Service Provider. The latter is able to accept the message because it is now in a data exchange format that the Service Provider supports.

Similar transformation and mapping actions are taken by the Mediator after it receives the reply message from the Service Provider, so that it complies with the data exchange format supported by the Service Requester. On receiving the response message the Service Requester does not know which service responded to the request, nor did it have to deal with any of the data exchange requirements of the Service Provider.

**Figure 3: Conceptual Cloud operations**

---

**Business Process Management (BPM)**

From a general point of view, Business Process management (BPM) is the orchestration of activities between people and systems. More specifically, BPM is a method for actively defining, executing, monitoring, analyzing, and subsequently refining manual or automated business processes. In other words, a business process is essentially a sequence of related, structured activities (i.e., a workflow) that is intended to achieve an objective. Such workflows can include interactions between human users, software applications or services, or a combination of both.

In a SOA-based information management environment this orchestration is most commonly performed by the Choreographer component of the ESB (Figure 2). Based on SOA principles, a sound BPM design will decompose a complex business process into smaller, more manageable elements that comply with common standards and reuse existing solutions.

The BPM design solution should be based on an analysis of the problem within both its local and global contexts (Figure 4). It must describe and support the local business process requirements as its primary objective and yet seamlessly integrate this micro perspective into a global view. Successful integration of these two perspectives will require an understanding of external interactions and the compliance parameters that apply to interprocess protocols. The principal components of a BPM design solution include a Business Process Execution Language (BPEL) engine, a graphical...
modeling tool, business user and system administration interfaces, internal and external system interactions, and persistence (Figure 5).

**Figure 4: BPM design requirements**

- **Problem Analysis:** Step-by-step flow chart of the business sequence.
- **Code Generation:** Model-driven architecture capable of at least some automated code generation.
- **Run-Time Tools:** For administrative monitoring and management during execution.
- **Interfaces:** To facilitate both human interaction and data exchange with external systems.

**Figure 4: BPM design requirements**

**Figure 5: BPM design components**

**BPEL Engine**: BPEL, which is the preferred process language, is normally XML-based and event driven. The BPEL Engine is responsible for detecting events, executing the appropriate next step in the business process sequence, and managing outbound message calls.
**Graphical Editor**: Effective communication during design is greatly facilitated by a standard system of notation that is known to all parties involved in the design process, and a graphical tool that allows design solutions to be represented in the form of diagrams. Both the Business Process Modeling Notation (BPMN) and the Unified Modeling Language (UML) Activity Diagram provide the necessary capabilities. However, BPMN is normally preferred because it incorporates BPEL mapping capabilities and is considered to be the more expressive notation. Whichever graphical modeling tool is chosen it should be capable of representing the different views of the process that are desired by the business user and the technical user. The business user is interested in the overall flow of the process, while the technical user is interested in the more detailed behavioral characteristics of each step.

**User-interfaces**: Typically, separate user-interfaces are required for the business user who has a functional role in the business process and may from time to time be required to interact with the BPEL Engine, and the system administrator who may be monitoring the task flow for reactive or proactive system maintenance reasons. The business users essentially require a worklist interface that allows them to contribute manual tasks to the automated BPM process. This should be a user-friendly, role-based interface with process status reports and error correction capabilities. The system administrators require a user-interface that allows them to perform a host of management tasks including: defining a process (i.e., find, activate, deactivate, remove, or add); controlling the execution of processes known to the BPEL Engine and worklist tasks or activities (i.e., find, suspend, resume, or terminate); managing user roles (i.e., add, modify, or remove users and roles from applications); and, configuring application connections. Both the business and system administration user-interfaces must incorporate security measures to prevent unauthorized access and ensure that only authorized role-based actions can be executed.

**System interactions**: A business process is likely to involve both internal and external system interactions. In general terms these interactions may be characterized as four distinct modes: process receives a message from another system; process receives a message and sends a response; process sends a message to another system; and, process sends a message and waits for a response. External interactions are typically choreographed as web services, with a wide variety of system interfaces being supported through a generic adapter facility. This means that the BPEL Engine must include a web services listener capable of accepting an inbound message (e.g., in SOAP format), insert it into the runtime engine, obtain a response (if any), and send out the response as a SOAP message. Internal interactions are typically either client-server interfaces to other systems executing on the enterprise network or inline code snippets.

**Persistence**: To survive the inevitable need to restart the BPEL Engine the current process state must be stored in a database. Tables in the database typically include: process definition; process execution state; message content and identification code; process variables; activity execution state; and, worklist task execution state.

While BPM and SOA concepts are closely connected, they are certainly not synonymous. Described more precisely, a SOA-based system environment provides the enabling infrastructure for BPM by separating the functional execution of the business process from its technical implementation.

---

**In Conclusion: Cloud Computing**

The concept of Cloud computing as a massively scalable, user-transparent computing resource that can be readily accessed by multiple users across a global network is indeed a compellingly attractive proposition. Combined with the...
SOA design and implementation principles described above, the Cloud not only takes care of all of the intricate technical interoperability and data exchange incompatibility issues that have plagued computer users in the past, but also provides essentially ubiquitous access to powerful and seamlessly integrated computer-based capabilities as services. Naturally, multiple Clouds can be linked in a manner that is quite similar to the way services are registered within a particular Cloud. In such an environment neither the service requester nor the service provider needs to know, or even care, where the request originated and where it was processed, even if the request for services had to traverse several Clouds before the necessary service provider could be found.

It is of interest to note that this view of computing as a service is not new. During the 1960s and 1970s time-share computer systems, which linked multiple remote user terminals through modems to a central computing facility, provided a similar computing service. However, there were some major differences. First, access and data exchange was strictly confined to a single computer center and in most cases to the particular application that the user was authorized to use. Second, very little of the underlying computing environment was transparent to the user. Third, the users were almost as rigidly tied to their access terminal location as the service provider was tied to the location of its computer center. The time-share concept became obsolete as soon as the advent of microcomputers brought the computing power to the user.

We might ask: Was it a desire by the computer users to have complete control over their computing resources or convenience that led to the preference of ownership over service? While Cloud computing promises to overcome the inconvenience, immobility, and lack of interoperability constraints of the time-share service environment, it does pose other problems that will need to be overcome. Chief among these is the issue of data security. Will organizations be willing to entrust their proprietary data to a remote Cloud environment over which, in reality, they have little control? They must trust the Cloud service provider to not only maintain adequate internal security, but to resist even the most sophisticated and continuously changing external intrusion attempts. Also, as Robert Lucky (2009) recently wrote “... once all your petabytes of data are out there in the Cloud, can you ever get them back?

Finally, there is the question of user autonomy and control. Are current and will future privacy laws be sufficient to protect the user from a plethora of potential consumer abuses, for example, the automated collection of data about a user’s activities in the Cloud without the need to actually trespass the data repositories themselves. Such data are already being collected by Internet service providers and utilized to determine collective and individual preferences for advertising and directed marketing purposes. Perhaps users will not be greatly concerned about the potential privacy infringements of such activities, and in the end the convenience and inexpensiveness of Cloud computing may become the deciding factors.

Footnotes

1
Adaptive Planning and Execution Roadmap II, AO Review (Draft), Joint Chiefs of Staff, 8 February 2007.

2
Most large organizations, including the Military, are currently forced to dedicate a significant portion of their operating budget, staff, project budgets, and time, on the piecemeal resolution of ad hoc problems and obstacles that are
symptoms of an overloaded data-centric environment. Examples include: data bottlenecks and transmission delays resulting in aged data; temporary breakdown of data exchange interfaces; inability to quickly find critical data within a large distributed network of data-processing nodes; inability to interpret and analyze data within time constraints; and, determining the accuracy of the data that are readily available. This places the organization in a reactive mode, and forces it to expend many of its resources on solving the symptoms rather than the core problem. In contrast, an information-centric environment is capable of supporting: (1) the automatic filtering of data by placing data into an information context; (2) the automated reasoning of software agents as they monitor events and assist human planners and problem solvers in an intelligent collaborative decision-making environment; and, (3) autonomic computing capabilities.

3

This must be considered a minimum system capability. The full implementation of a web services environment should include facilities that allow a service to discover other external services and understand the nature and capabilities of these external services.

4

OASIS is an international organization that produces standards. It was formed in 1993 under the name of SGML Open and changed its name to OASIS in 1998 in response to the changing focus from SGML (Standard Generalized Markup Language) to XML (Extensible Markup Language) related standards.

5

The Extensible Markup Language (XML) is a general purpose specification that allows the content of a document to be defined separately from the formatting of the document.

6

BPMN provides a graphical representation for describing a business process in the form of a workflow diagram. It was developed by the Business Process Management Initiative (BPMI) and is now maintained by the Object Management Group following the merging of these two organizations in 2005.

7


8

A BPM worklist allows a manual task to be assigned to a user and track the progress of that task. In this way the human user can be the source of events that trigger the BPEL Engine.
The Simple Object Access Protocol (SOAP) is a protocol specification for the exchange of data among web services. It utilizes XML as its message format and depends on other protocols, such as Remote Procedure Call (RPC) and Hypertext Transfer Protocol (HTTP) for transmitting the message.

Reference

Ref 1

Ref 2

Ref 3

Ref 4
Havey M. (2005); ‘Essential Business Process Modeling’; O'Reilly, Sebastopol, California.

Ref 5

Ref 6
Lucky R. (2009); ‘Cloud Computing’; (under Reflections) IEEE Spectrum, Institute of Electrical and Electronics Engineers, 46(5), May (pp. 27).

Ref 7

Ref 8
Ref 9


E-MAPS on Ontology and Big Data

Ontology
Source: [http://www.e-mapsys.com/Problem_Space2.html#Ontology](http://www.e-mapsys.com/Problem_Space2.html#Ontology)

Ontology is the science of representing reality consistently across domains, organizations, and IT systems. Everyone performs ontology as they make sense of their circumstances and plan their future actions. The concepts and methods of ontology are to the development and use of IT as weather is to development and use of aircraft and ships. Yet, while there are first-rate, readily available references for pilots and seamen on weather, easy-to-access and easy-to-use references on ontology are not available to IT developers and users. E-MAPS and its partners provide such references and instruction.

- Basic Value of Ontology (Ontology for the Above Average Manager)
- Warfighters, Ontology, and Stovepiped Data, (Information, and Information Technology)

Basic Value of Ontology
Source: [http://www.e-mapsys.com/Basic_Value___y_%28v1%29.pdf](http://www.e-mapsys.com/Basic_Value___y_%28v1%29.pdf) (PDF)

Ontology has three common definitions: (1) the science of representing reality, (2) products such as Web Ontology Language (OWL) files that contain representations of reality, and (3) software tools such as TopQuadrant used to create OWL files and other ontological products.

Managers need a basic understanding of ontology because their effectiveness and efficiency is determined largely by their ability to understand, access, and share information in various situations. Ontology, in the sense of the science of representing reality (i.e., the first definition above), provides concepts and methods that facilitate developing and managing the information managers need from data and information drawn from multiple sources (hopefully from all available sources).

Ontology’s importance is growing because computer networks have the potential to connect all data and information. Realizing this potential, however, requires the right tools – and the right concepts and methods. The concepts and methods are important because they provide context for the users of TopQuadrant and similar tools. By understanding context, an ontology user can take the actions necessary to produce OWL files with contents that managers and other users need and can use.
Ontology addresses reality and how we represent it. We represent different slices of reality (i.e., different perspectives) with different constructs. But the essential characteristics of elements of reality that are common to two or more “slices” should be represented the same way in both slices. A bridge is a bridge whether we are talking about toll-bridge finance or trucking operations. What is different about a bridge in the context of highway-bridge finance and the context of trucking operations are the bridge’s roles – what ontologists call “accidental qualities.” Finance specialists are concerned about the bridge as a generator of toll payments for bond holders. A trucking company owner is concerned about the bridge as a means for his truck to cross a river so his company can deliver cargo. We have one bridge with one “essential” quality (i.e., a structure that provides a route to cross a river or other area) and multiple “accidental” qualities (e.g., a source of revenue and the means to cross a river).

Ontology is to data and information as meteorology is to air temperature, air pressure, clouds, and wind. We can make good use of data and information without the concepts and methods of ontology, but we will make better use of data and information if we understand and use the concepts and methods of ontology. A person with little knowledge of the concepts and methods of meteorology can sometimes look at the sky or feel the wind and know it is going to rain or snow. However, when we want the best possible weather forecast, we turn to meteorologists because they use their understanding of the concepts and methods of meteorology to produce accurate forecasts.

To learn how E-MAPS can help you to understand and exploit the concepts and methods of ontology to improve your use of data and information, call us at 703-385-9320 or send an email to ontology@e-mapsys.com.

---

**Warfighters, Ontology, and Stovepiped Data**


**The Operational Problem**

Warfighters and others in DoD need to share warfighting and business data and information across and beyond DoD easily. Today’s impediments to such sharing need to be remedied because they hinder DoD realizing the efficiency and effectiveness required to remain affordable and effective. Data and information that cannot be easily shared machine-to-machine between domains, specialties, organizations, and information technology (IT) systems is characterized as being stovepiped (i.e., a system that does not interoperate with other systems).
This Document’s Contribution

This document explains concepts and methods essential to:

(1) Creating data, information, and IT systems that are not stovepiped, and

(2) Integrating data, information, and IT systems that are stovepiped.

Reality is not Segmented

The first and most important concept is that reality is an integrated whole, not a collection of mutually exclusive domains. Each specialty (e.g., medicine, logistics, intelligence, etc.) is a different perspective of our common reality. In medicine, a person can be a caregiver and patient. In logistics, a person can be a passenger, a customer, or a worker. In intelligence, a person can be an intelligence specialist or a target. Indeed, a person can fill all these roles simultaneously.

Three Orders of Reality

Situational awareness (i.e., understanding reality) is generally understood to be essential for success. It is also generally understood that people often misunderstand reality. The theory of ontology, the modeling of reality, addresses the disparities between reality and what people think is reality by defining three orders of reality.

(1) 1st Order. Reality as it is. In the action in the upper image to the right, reality is what is, not what we think is happening as we peer through the fog of war.

(2) 2nd Order. What we believe is happening as we peer through the fog of war. Examples: what a participant in the action shown in the upper image or a member of an operations center in the lower image believes is occurring in the engagement.

(3) 3rd Order. Reality as we record it. In the lower image, the computer displays are 3rd order reality.

We create 2nd and 3rd order realities with symbols (e.g., words and map icons). Gaps between orders of reality introduce risk. These gaps are not the only form of risk, but reducing these gaps contributes to reducing risk.

Role of Words and Other Symbols

Closing these orders-of-reality gaps starts with aligning the words and other symbols we use to create our 2nd and 3rd order realities with 1st order realities. People have difficulty comprehending something for which they do not have a word or other symbol. It is nearly impossible to communicate or record an aspect of reality for which we lack a word or other symbol.

Creation and Use of Words and other Symbols

People usually create words and symbols in response to specific problems in specific domains. The natural result of this reality is sets of words or vocabulary focused on specific domains (i.e., stovepiped data, information, and IT systems). However, organizational effectiveness and efficiency require that data and information in one domain or IT system be
shared and used by members of other domains and the IT systems these individuals use. The challenge, therefore, is how to facilitate creating terminology and other symbols based on a particular type of problem or specialty (e.g., improvised explosive devices [IED]) that are easy to integrate and use with terminology and other symbols for other types of problems and specialties.

The solutions include:

1. Using the same words and other symbols across specialties (e.g., for command and control, operations, logistics, and intelligence);
2. Grouping words and other symbols that are common among different specialties according to common categories (e.g., the category or class “sensor” can be used by logistics for devices that sense engine problems on trucks and for the instruments used by intelligence agencies to generate images); and
3. Developing a vocabulary for a specialty or domain that represents its basic elements (e.g., a model of an infantry company intended to facilitate personnel management that extends beyond the concept of a company to include individual soldiers, not just platoons, squads, and fire teams).

Avoid Conflation

Conflation is when two or more independent concepts are combined into a single concept with a single term and definition. Example:

*Attack Geography* – a description of the geography surrounding the...incident, such as road segment, buildings, foliage, etc. Understanding the geography indicates enemy use of landscape to channel tactical response, slow friendly movement, and prevent pursuit of enemy forces.

In this example, the definition for attack geography conflates (i.e., aggregates) two related but separate categories. The first category is the aspect of reality that is of interest (i.e., the area where an attack occurred). The second category is a description (e.g., verbal statement, written report, or map overlay) of the area. Going back to the orders of reality, an area where an attack occurs is part of first-order reality. A description of that area is an element of second- or third-order reality depending upon whether the description is in someone's memory or is in a document. Usually, there are multiple descriptions produced by various observers of any area where an attack occurred. None of the descriptions is completely accurate and no report contains all the details. In practice, members of an operations center distinguish between reports they receive about an incident and the incident itself. Experienced members of operations centers know that they need to collect and study multiple reports to understand an incident because each report is likely to (1) contain some inaccurate information and (2) have only some of the needed details.

The concept of attack geography as defined above should have been divided into two concepts (i.e., two pairs of terms and definitions) – attack geography and description of attack geography. Because the definition above is conflated, it obscures the crucial distinction between what is described and the description. This conflated definition also invites mistaken beliefs such as there can be one authoritative and complete and accurate description (e.g., report) of the area in which an attack occurred when, in fact, there will almost always be (1) multiple reports and (2) some information that is believed to be true but is not true.
Conflation by its nature introduces inaccuracies.

**Base Definition on Essential Properties**

Avoiding conflation is aided by basing a definition on the essential property of what is being defined. Returning to the definition above, the essence of attack geography is that it is an area or geography where an attack is planned, is occurring, or has occurred.

**Non-Essential or Accidental Properties**

For attack geography, accidental properties include the types of attacks (e.g., ambush, frontal assault, and attack by fire). Basing a definition on an accidental property causes conflation and produces an inaccurate definition. Returning to the definition of attack geography, if we define attack geography as an area where an IED incident occurred, then we have a definition which states that the concept of attack geography cannot be applied to ambushes, assaults, and attacks by fires.

**Form Definitions Properly**

The need to focus on essential properties leads to the following two steps when creating a definition.

1. **Step 1:** refer to a parent class (e.g., Infantry Battalion: A Military Organization or Military Engagement: A Military Event).
2. **Step 2:** add differentia (i.e., those properties that distinguish the thing being defined from all other things in its parent class).

These steps force several necessary considerations.

- First, what is the parent class? This will prompt someone developing a vocabulary for a particular problem or specialty (e.g., artillery fires) to ask what is the parent class for that domain (e.g., fire support).

- Second, this facilitates inquiries into related problems and specialties (e.g., naval fire support and close air support) with the aim of identifying words and other symbols defined for the related problem or specialty that can be reused in the new vocabulary. It also should prompt an inquiry into the classes used by related specialties.

**Example Definitions from Joint Publications**

- **Fires** — the use of weapon systems to create specific lethal or nonlethal effects on a target. (JP 3-09) (Note that the definition refers to a parent class [the use of weapons systems] and lists a certain differentia [to create specific lethal or nonlethal efforts on a target]. This points to fires as part of a larger realm - the use of weapons systems for any purpose)

- **Final protective fire** — an immediately available prearranged barrier of fire designed to impede enemy movement across defensive lines or areas. (Note that the definition starts with reference the class “fires” and then states the differentia – an immediately available prearranged barrier intended to impede enemy movement across defensive lines or areas)
Counterfire — fire intended to destroy or neutralize enemy weapons. Includes counter-battery and countermortar fire. (JP 3-09) (Note again that the definition refers to the parent class and distinguishing element [i.e., destroy or neutralize enemy weapons])

Suppressive fire — fires on or about a weapons system to degrade its performance below the level needed to fulfill its mission objectives during the conduct of the fire mission. (Note that again the definition starts with a reference to the parent class)

**Reality versus Convention**

The orders of reality imply a difference between reality and the symbols we develop to represent reality. The most important difference between reality and our symbols for representing reality is that we cannot change reality but we can represent a feature in reality with any symbol we want to use. However, DoD efficiency and effectiveness are promoted by establishing and following conventions or standards that map words and other symbols to features in reality. The English word “water” and the French word “l’eau” are equally good symbols for H2O. Which is appropriate depends on convention or the language of those seeking to communicate.

**Relationships**

If one, as is suggested above, models reality with terms that represent the smallest relevant elements (thus avoiding conflation), these elements must be connected with relationships. In the image below, various smallest-possible elements are connected. The reality of the person is represented by the photograph. This cannot be changed by someone performing ontology, the modeling of reality. Use of the terms and concept of person to represent the real individual, however, is a matter of convention (i.e., agreement or authoritative direction). Any number of terms and definitions might be used. What is required is uniform understanding and application of a term in modeling reality. The connection between a real person and his representation is the relationship **“Instance_of”**. This and other relationships used in modeling reality need to be codified by convention (e.g., recorded in an authoritative data source).

Note that the essential property of the individual (i.e., that he is a person or homo sapien) is separated from his various names. To avoid conflating a specific name with the person, an intermediate concept or class of “Personal Name” has been used and connected to person by the relationship **“denotes”** and to specific names used by that individual with the relationship **“is_a”**. Note also that the designer of this model sought and found the class to which “Personal Name” belongs – “Name”. Understanding the broader class may facilitate representing information about this person.

![Relationship Diagram](https://semanticommunity.info/Data_Science/Ontology_for_Big_Data)

**Invest in Ontological Foundation**

The model on the previous page is irrelevant if one is simply interested in knowing a name of the individual shown in a picture. Operational success in Iraq and Afghanistan, however, required a capability to associate one individual with
multiple names because a single Afghan or a single Iraqi sometimes used multiple names. The mental model that each Afghan and Iraqi uses only one name is not aligned with reality and has facilitated adversaries slipping through our fingers by using different names at different times and places. A database that treats a person’s name as his essential property is not only ontologically incorrect but also hinders or precludes recording multiple names for the same individual.

Purpose and Scope of Document Review

This document is intended to alert those seeking to rapidly develop sound situational awareness about the need to understand and apply the concepts and methods of ontology. This document also seeks to illuminate the need to understand and use these concepts and methods when attempting to implement information sharing and reuse across an organization. Everyone uses ontology (i.e., models reality) as they puzzle through what is occurring around them and determine how to accomplish assigned tasks or missions. Those who understand the concepts and methods of ontology have an advantage over those who (1) create stovepipe data, information, and IT systems through inadvertent conflation, poor definitions, and domain models that are poorly aligned with reality and (2) do not support simple but critical capabilities such as relating a single person with multiple names.

Conclusion

Because there are many more ontological concepts and methods than those explained in this short document, those who would exploit ontology need to learn more about ontology or gain access to advisors schooled and experienced in ontology. It is important to understand that there is a difference between someone who uses software such as TopQuadrant that is intended to support ontology and someone educated in the concepts and methods of ontology. This is similar to the distinction between someone who can use Microsoft Word software and someone who can write a good book.

For further information

Contact E-MAPS, Inc.: Email: ontology@e-mapsys.com Phone: 703-385-9320 Visit the E-MAPS website: www.e-mapsys.com

Big Data

Source: http://www.e-mapsys.com/Problem_Space2.html#Big_Data

The term Big Data represents the reality that advances in (1) sensors and other computer-based data-generating tools and (2) computer networks force far more data on users of IT than they can comprehend or even sort out without using IT tools. Big Data conversations and tools tend to focus on manipulating semi-structured or unstructured data. Manipulation is important, but it is more important to start by identifying the information problem to be solved and the ontology (i.e., representation of reality) that describes a problem’s domain. E-MAPS and its partners provide support in representing domains.

• Enabling Big Data Solutions
Enabling Big Data Solutions


The term **Big Data** has different meanings for different people. For some people, it is large quantities of disparate data. For others, it is the activity of making effective use of such data. Large quantities of disparate data are the result of advances in (a) data storage facilities, (b) data communications and (c) sensors and other data-generating tools.

The effective use of Big Data requires hardware and software for collecting, communicating, and processing large volumes of data. Demystifying Big Data: A Practical Guide to Transforming the Business of Government notes:

"Although there clearly is an intense focus on Big Data, there remains a great deal of confusion regarding what the term really means, and more importantly, the value it will provide...This confusion may be due in part to the conversation being driven largely by the information technology community versus line of business community.

Successful Big Data initiatives seem to start not with a discussion about technology, but rather with burning business or mission requirements that...leaders are unable to address with traditional approaches."

The white papers listed below address the use of ontology and other resources to facilitate discussions about "burning business or mission requirements" that are not met with existing information technology concepts, methods, and tools.

**Big Data and Why It Cannot Be Ignored**

**Big Data and Ontology**

**Big Data and Why It Cannot Be Ignored**


This paper discusses the importance of Big Data and that it cannot be ignored because people and organizations that understand Big Data are using it to gain competitive advantages.

The term Big Data has come into use because just purchasing an information technology (IT) tool does not lead to better capabilities, efficiency, and effectiveness, even though intuition may suggest that more data and better tools should lead to better operations. At this point in the evolution of computers, Big Data is shorthand for the question, "**What do we have to do to realize the potential of ever-better computer-based data generators, ever-larger data stores, and increasingly complex computer networks that connect evermore data stores?**"

The last 30 years’ investments in (1) computer-based tools for generating data and (2) information storage facilities have left organizations with very large volumes of data and information. A few people are proving adept at exploiting this data and information. However, most people and organizations only sense unrealized potential.

Most investments in computer-based tools for generating and storing data and information have been focused on supporting a particular user community (e.g., sales, operations, or finance). The organizations that have performed the
best have integrated data and information from all their divisions. One result of well-integrated information has been that corporations focus their advertising in areas where they have outlets. This may sound obvious, but some large corporations have spent millions of dollars advertising in areas where they did not have outlets. IT expenditures in such corporations often reinforce impediments to information sharing, degrade corporate synergy, and lead to financial problems.

Eliminating such waste and starting to answer the Big Data question above requires IT managers, developers, and users to shift from (1) concerning themselves only with their computer-based tools and their corner of their enterprise to (2) a perspective that (a) understands the details of their computer-based tools and their corner of the enterprise and (b) relates those concerns to accomplishing the enterprise’s goals.

The heart of Big Data is all elements of an organization taking actions that generate, share and use data and information in ways that improve the performance of (1) each element of the organization and (2) the organization as a whole.

To ignore Big Data is to ignore the concepts, methods, and tools being developed by many diverse efforts across governments, academic institutions, and business with the intent of realizing the largest possible advantage from investments in IT.

**Big Data and Ontology**


This paper introduces and discusses the role in Big Data of ontology, the science of representing reality across disciplines and IT systems.

Big Data needs to be understood and exploited because of the ever-increasing volumes of data being (1) generated by sensors and other computer-based data generators and (2) made accessible through computer networks. The question now is how and not whether to exploit Big Data.

Exploiting Big Data starts with understanding that while the computer-based tools that give us today’s Big Data challenges and opportunities are relatively new, Big Data challenges have arisen and been mastered repeatedly throughout history. When the ancient Greeks created data and information on a wide range of subjects from mathematics to architecture and medical science, Aristotle saw the need for and wrote out theory and methods for categorizing elements of reality consistently across individual subjects. These concepts are the core of today’s concepts and methods of ontology and are used in IT systems that produce a few elements of important information from large volumes of data that have little value until processed.

Ontology’s primary contributions and value are concepts and methods for categorizing elements of reality consistently across disciplines and IT systems. Google Maps demonstrates the importance of ontology. A Google Maps user enters a street number and name, city, and state and gets a quick response regardless of the city or state. This capability is possible because cities inventory addresses within their boundaries for 911 systems, tax rolls, and predicting school enrollment. Google Maps can exploit these inventories because cities addresses are composed of standard categories of data (i.e., street number, street name, city, state, and zip code).
Users of Google Maps are impressed by the software - the interface, the detail of the maps, and the clever icons for stores and houses. Google’s success at applying IT to create and maintain this capability to exploit data produced by thousands of organizations around the world depends as much on the theory and methods of ontology as on databases, networks, and other elements of IT. Google maps would not exist without the IT, but selecting and tailoring the IT rests on understanding and exploiting the concepts and methods of ontology.