2014 Ontology Summit & Symposium

Big Data and Semantic Web Meet Applied Ontology

Summary

Presented by

Ram D. Sriram
Chief, Software and Systems Division
Information Technology Laboratory
National Institute of Standards and Technology, USA
sriram@nist.gov

On behalf of Ontology Summit 2014 Organizers and Participants

Overview of Ontology Summits

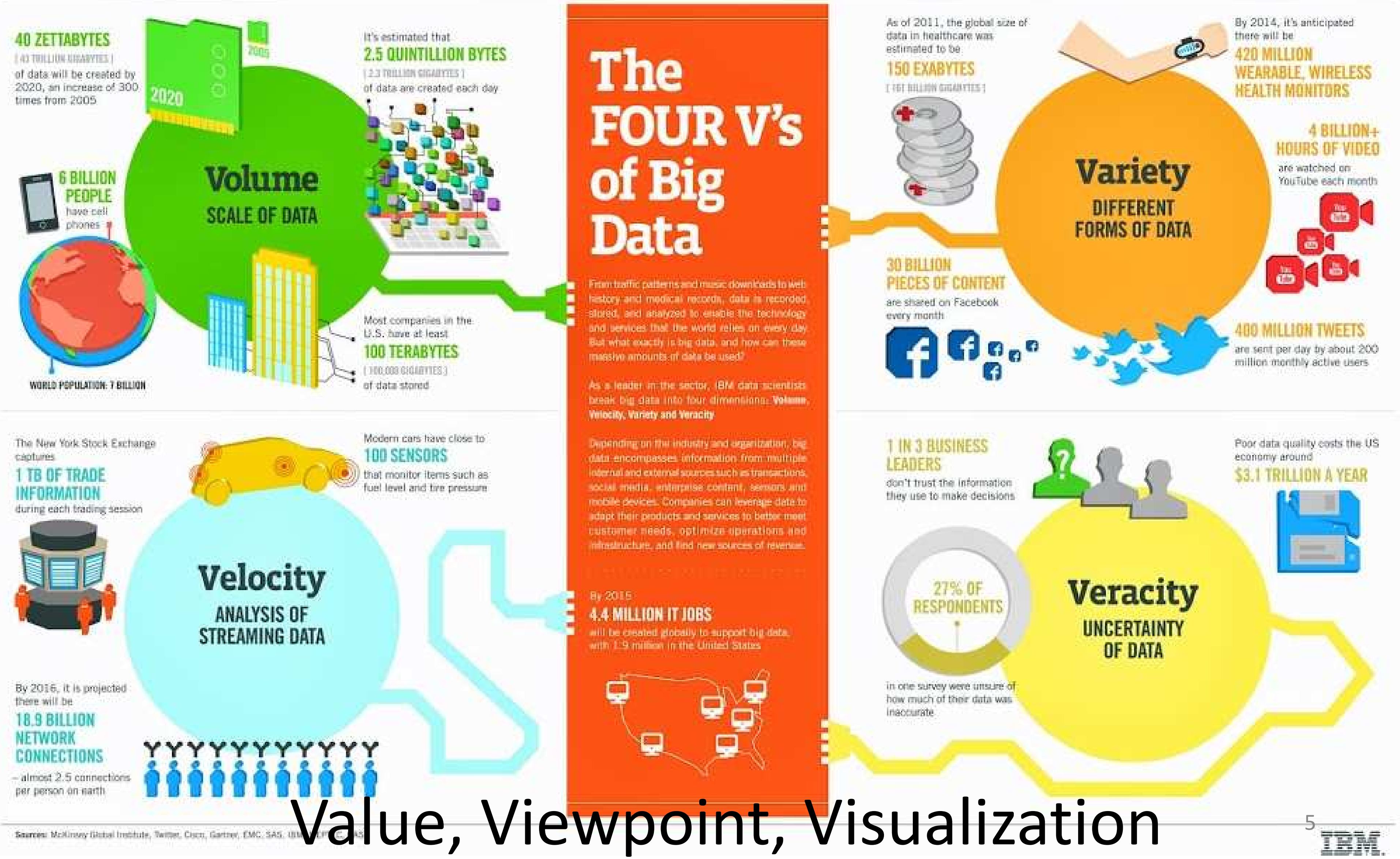
- The Ontology Summit is an annual series of events that started in 2006 with the joint sponsorship of Ontolog and NIST
- The summit is largely a self-organizing, bottom-up, volunteer driven effort, that solicits contributions from participants around the world in both industry and academia
- Each year's Summit (different theme every year) consists of a series events and continued discourse spanning three months, culminating in a free, two-day face-to-face workshop and symposium
- URL: http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit

Summit History

- 2006: Upper Ontology
- 2007: Ontology, Taxonomy, Folksonomy: Understanding the Distinctions
- 2008: Toward an Open Ontology Repository
- 2009: Toward Ontology-based Standards
- 2010: Creating the Ontologists of the Future
- 2011: Making the Case for Ontology
- 2012: Ontology for Big Systems
- 2013: Ontology Evaluation across the Ontology Lifecycle.
- 2014: Big Data and Semantic Web Meet Applied Ontology

BIG DATA

Issues in Big Data



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States

Volume
SCALE OF DATA

40 ZETTABYTES [43 TRILLION GIGABYTES] of data will be created by 2020, an increase of 300 times from 2005

It's estimated that **2.5 QUINTILLION BYTES** [2.3 TRILLION GIGABYTES] of data are created each day

6 BILLION PEOPLE have cell phones

WORLD POPULATION: 7 BILLION

Most companies in the U.S. have at least **100 TERABYTES** [100,000 GIGABYTES] of data stored

Variety
DIFFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be **150 EXABYTES** [161 BILLION GIGABYTES]

By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

4 BILLION+ HOURS OF VIDEO are watched on YouTube each month

30 BILLION PIECES OF CONTENT are shared on Facebook every month

400 MILLION TWEETS are sent per day by about 200 million monthly active users

Velocity
ANALYSIS OF STREAMING DATA

The New York Stock Exchange captures **1 TB OF TRADE INFORMATION** during each trading session

Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure

By 2016, it is projected there will be **18.9 BILLION NETWORK CONNECTIONS** – almost 2.5 connections per person on earth

Veracity
UNCERTAINTY OF DATA

1 IN 3 BUSINESS LEADERS don't trust the information they use to make decisions

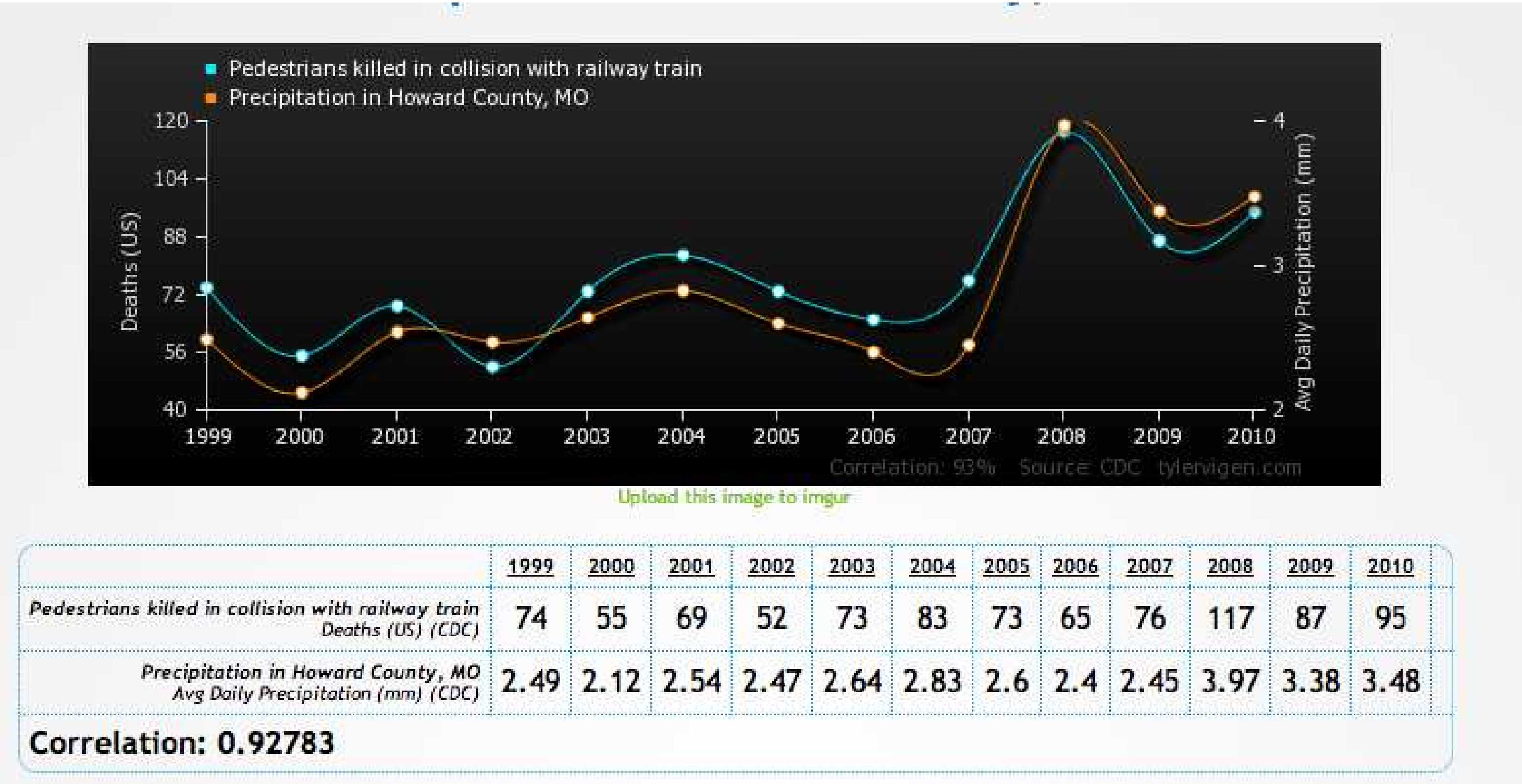
Poor data quality costs the US economy around **$3.1 TRILLION A YEAR**

27% OF RESPONDENTS in one survey were unsure of how much of their data was inaccurate

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

Value, Viewpoint, Visualization

Spurious Relationships



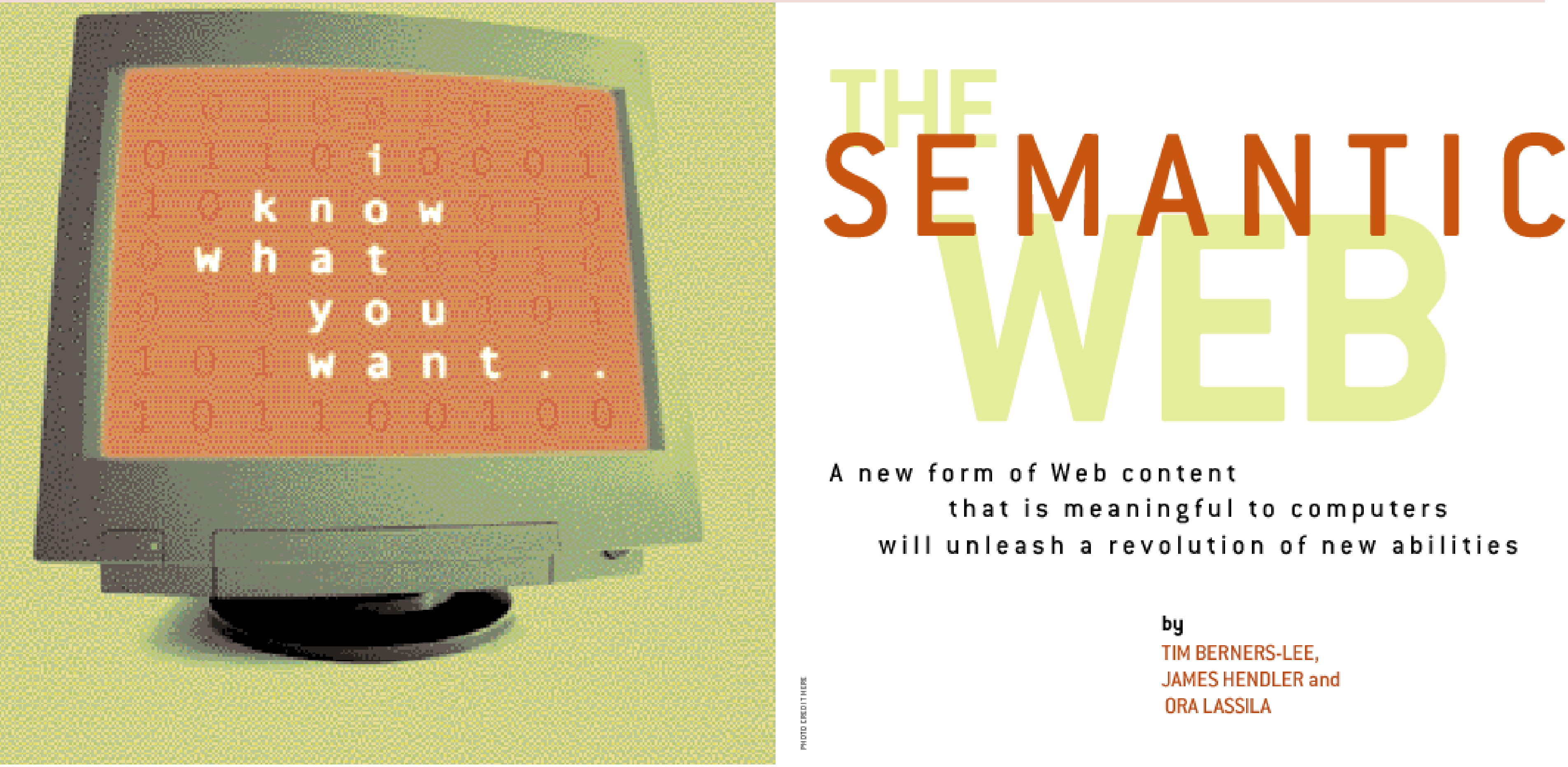
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Pedestrians killed in collision with railway train *Deaths (US) (CDC)*	74	55	69	52	73	83	73	65	76	117	87	95
Precipitation in Howard County, MO *Avg Daily Precipitation (mm) (CDC)*	2.49	2.12	2.54	2.47	2.64	2.83	2.6	2.4	2.45	3.97	3.38	3.48

Correlation: 0.92783

Courtesy: http://www.tylervigen.com

THE SEMANTIC WEB

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.



THE SEMANTIC WEB

A new form of Web content
that is meaningful to computers
will unleash a revolution of new abilities

by
TIM BERNERS-LEE,
JAMES HENDLER and
ORA LASSILA

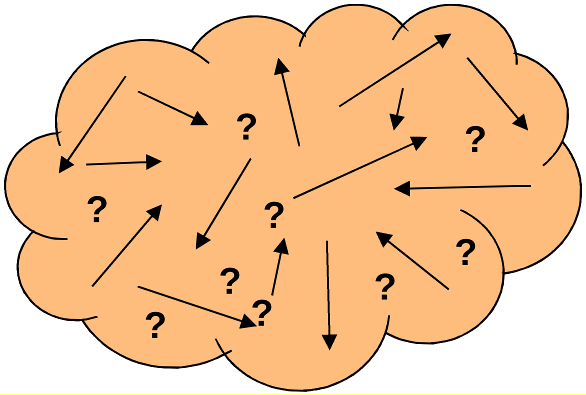
PHOTO CREDIT HERE

From Berners-lee, Hendler, J., and Lassila, The Semantic Web, Scientific American, May 2001.

The Semantic Web

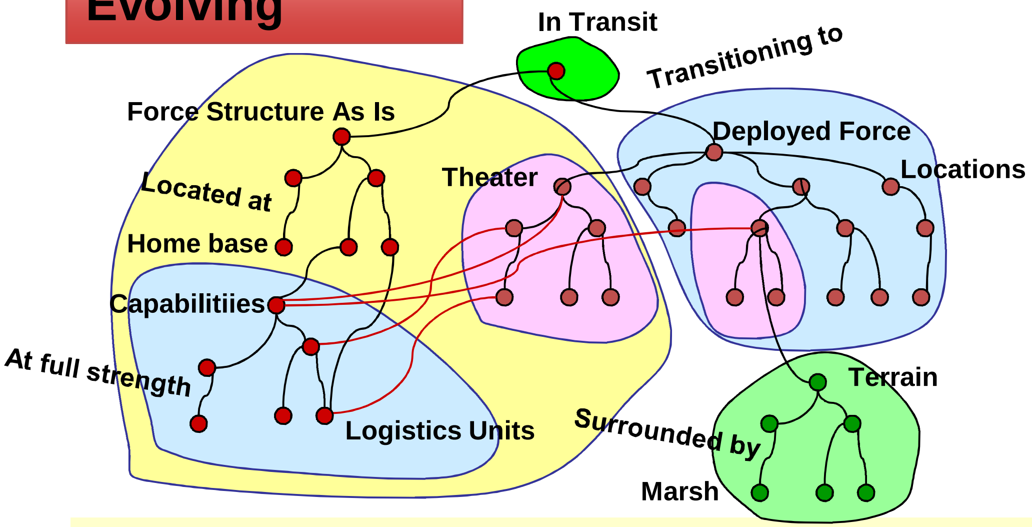
- The Web (2010) is a collection of links and resources
  - Is syntactic & structural only
  - Excludes semantic interoperability at high levels.
  - Google has a linked data structure (keyword) & has no notion of the semantics (meaning) of your query
- Semantic Web extends the Web so information is given well-defined meaning
  - Enables semantic interoperability at high levels
  - Google of tomorrow will be concept based (we are seeing that now)
  - Able to evaluate knowledge in context

Web
2010



Humans have to do the understanding

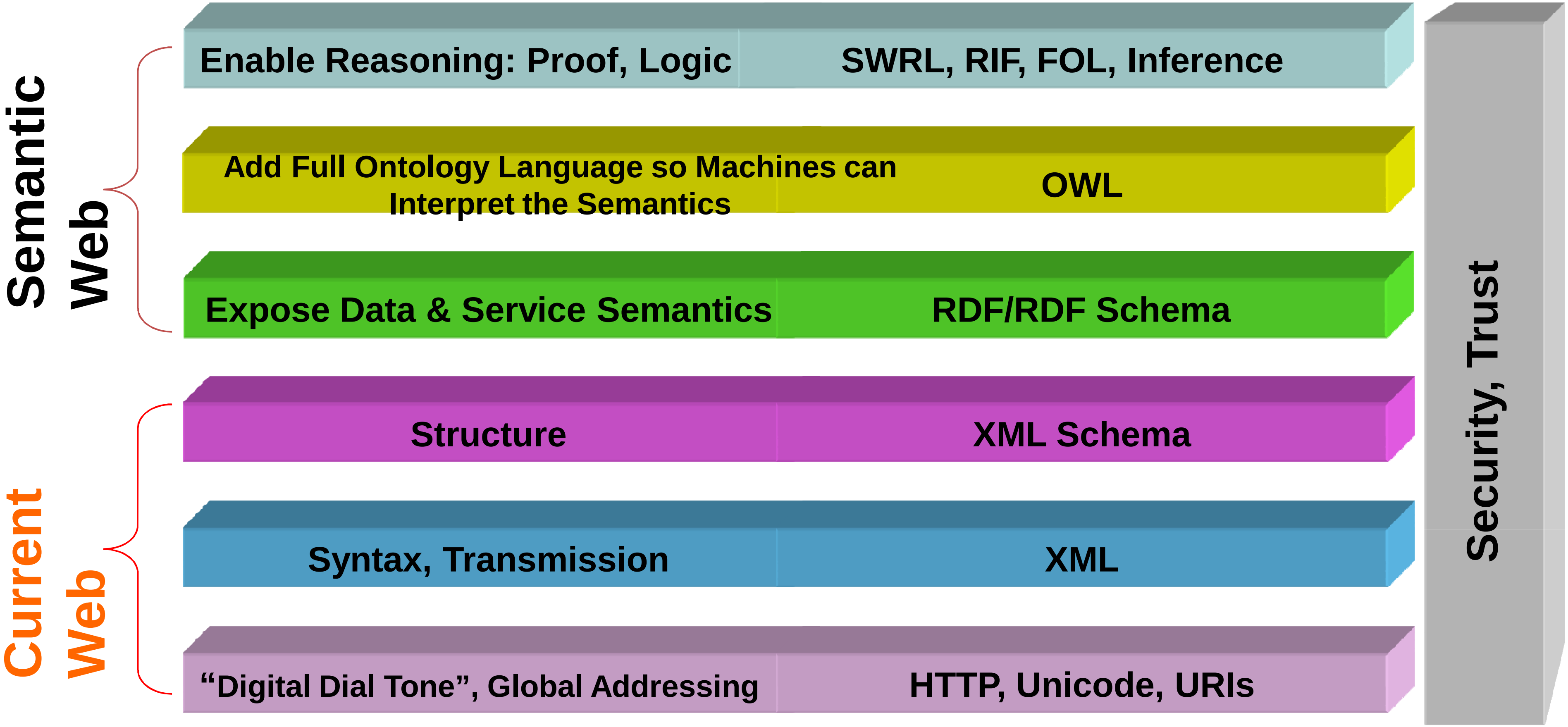
Semantic Web Evolving



Machines partially understand what humans mean

Courtesy: Leo Obrst, MITRE

Semantic Web Context



Semantic Web

Enable Reasoning: Proof, Logic	SWRL, RIF, FOL, Inference
Add Full Ontology Language so Machines can Interpret the Semantics	OWL
Expose Data & Service Semantics	RDF/RDF Schema
Structure	XML Schema
Syntax, Transmission	XML
"Digital Dial Tone", Global Addressing	HTTP, Unicode, URIs

Current Web

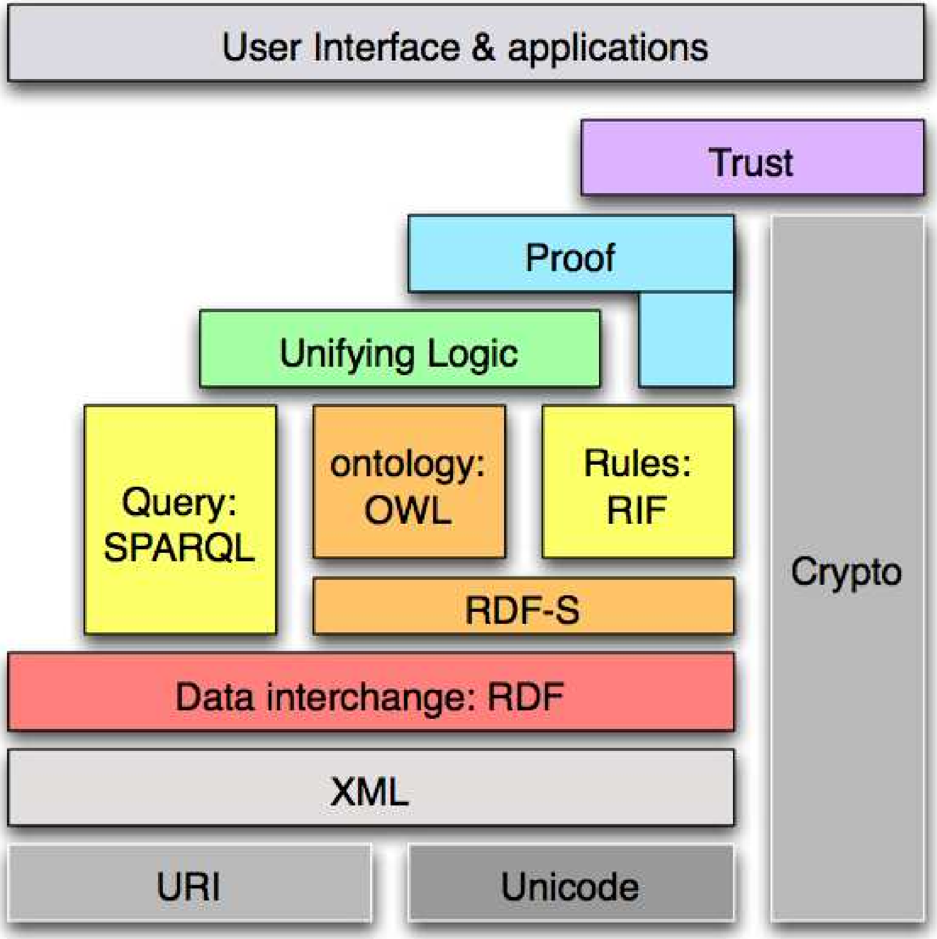
Security, Trust

Anyone, anywhere can add to an evolving, decentralized "global database"

Explicit semantics enable looser coupling, flexible composition of services and data

Courtesy: Leo Obrst, MITRE

Semantic Web Architecture



Courtesy: Jim Hendler

ONTOLOGIES

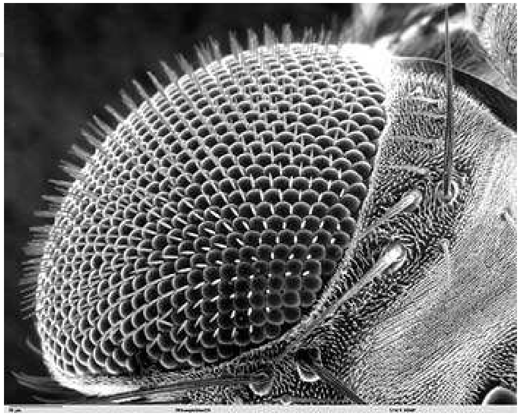
What Is An Ontology

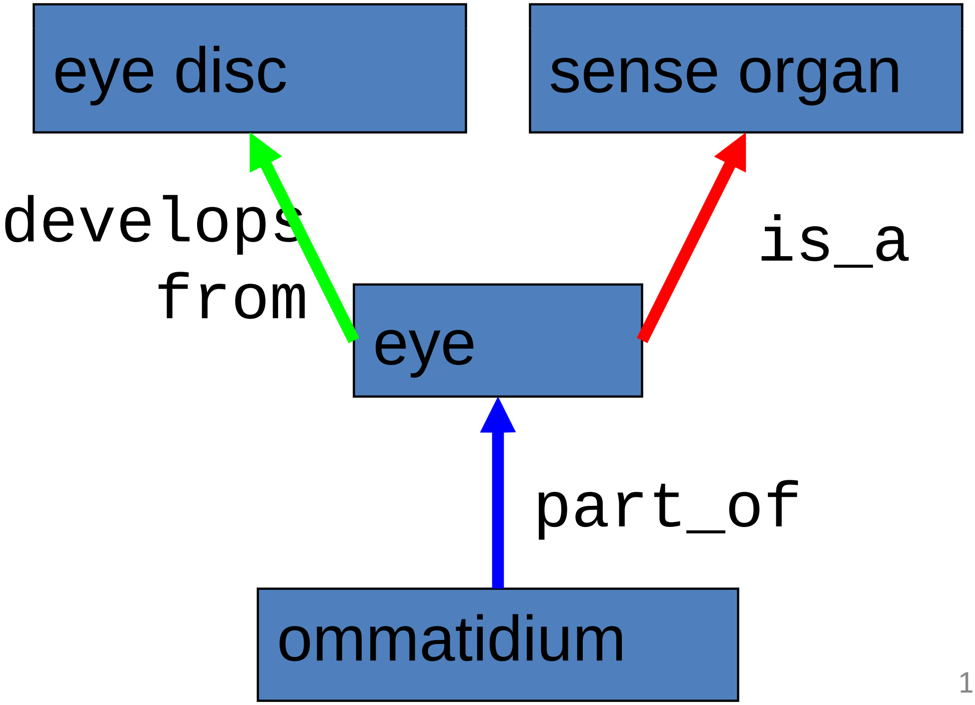
- An ontology is an explicit description of a domain:
  - concepts
  - properties and attributes of concepts
  - constraints on properties and attributes
  - Individuals *(often, but not always)*
- An ontology defines
  - a common vocabulary
  - a shared understanding

Courtesy: Natalya F. Noy

Example: A biological ontology is:

- A machine interpretable representation of some aspect of biological reality
  - what *kinds* of things exist?

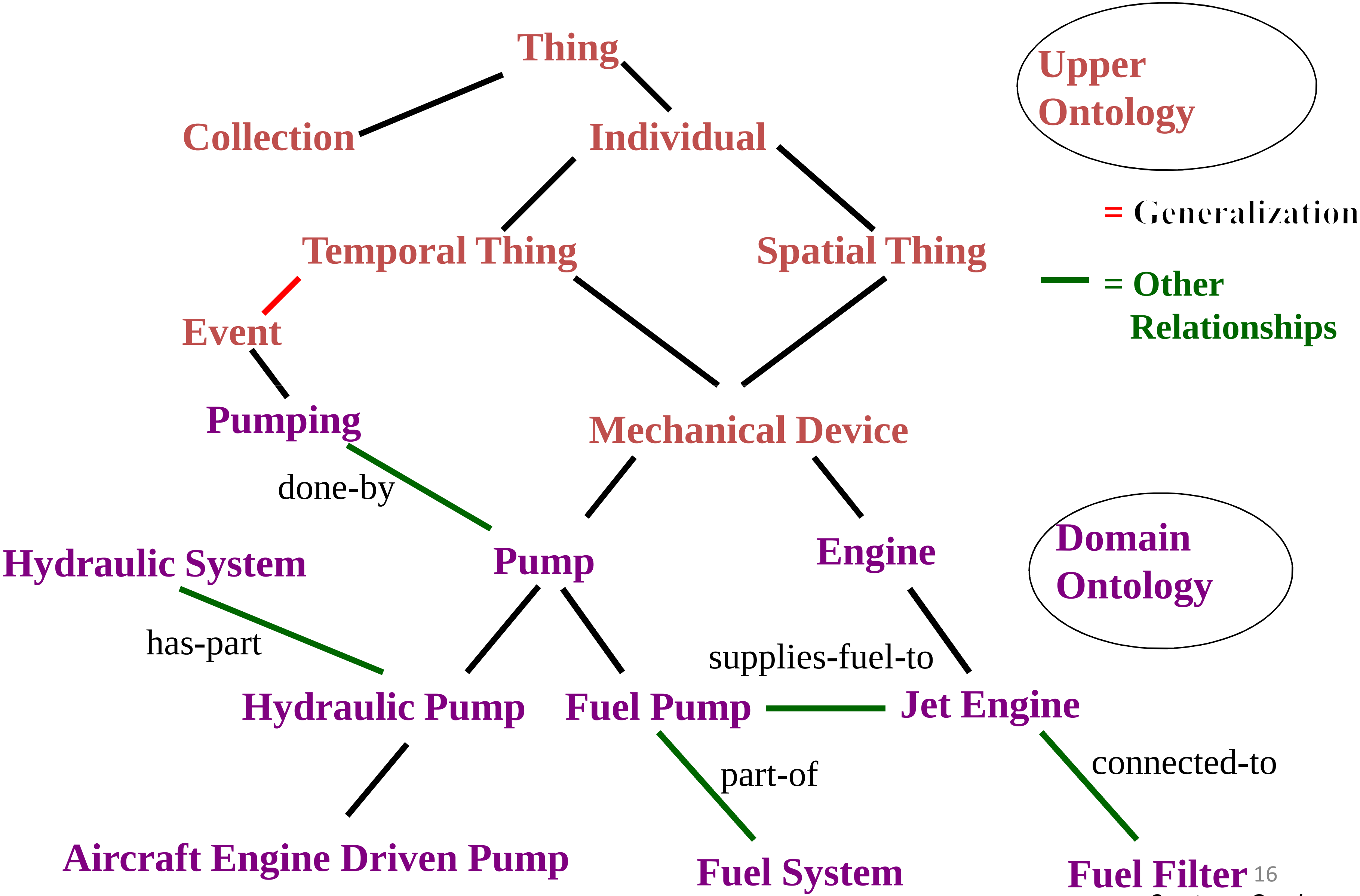




Courtesy: Musen

The Foundational Model of Anatomy

Engineering Ontology



Thing

Collection

Individual

Temporal Thing

Spatial Thing

Event

Pumping

Mechanical Device

done-by

Hydraulic System

Pump

Engine

has-part

supplies-fuel-to

Hydraulic Pump

Fuel Pump

Jet Engine

part-of

connected-to

Aircraft Engine Driven Pump

Fuel System

Fuel Filter

Upper Ontology

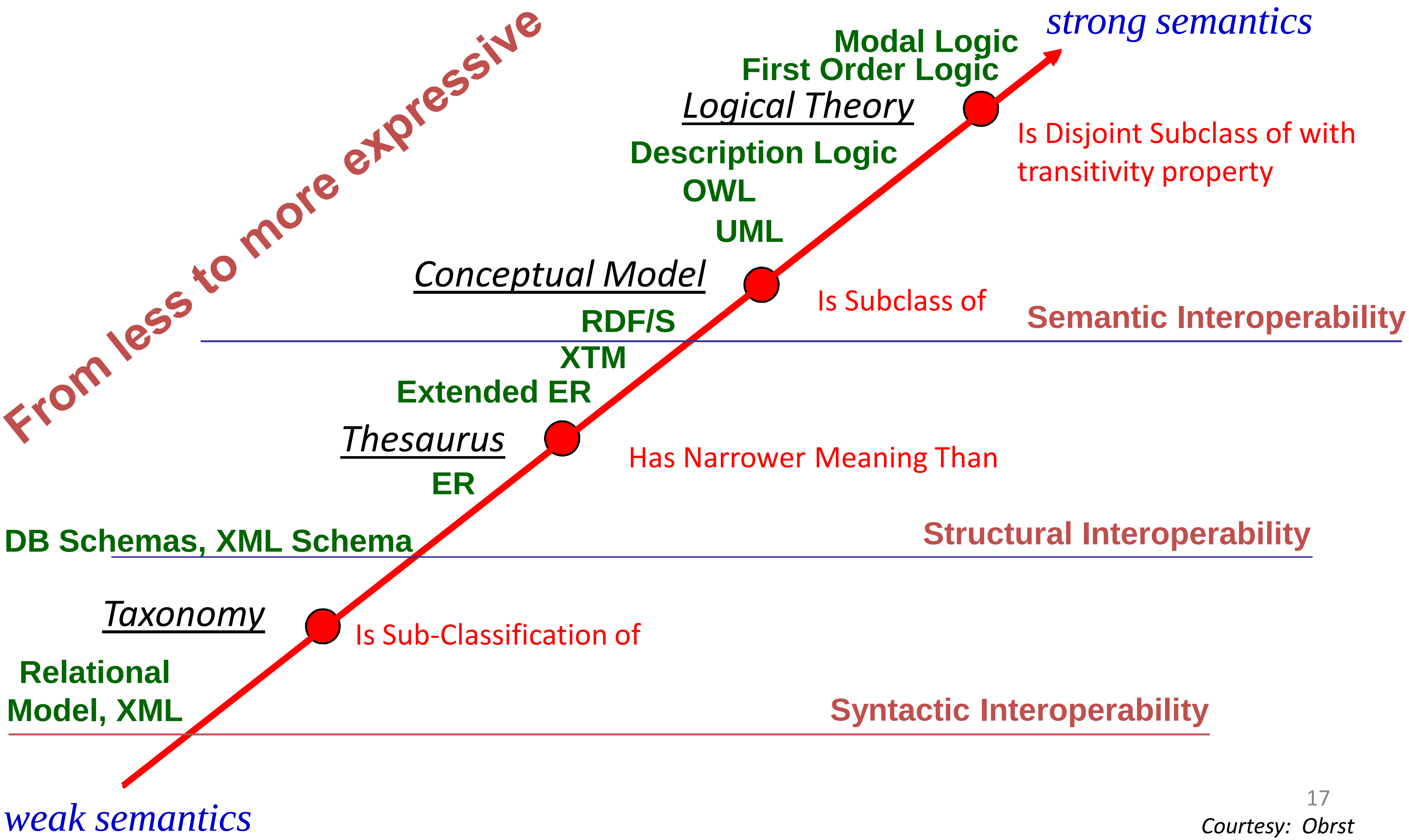
Domain Ontology

🟥 = Generalization

🟩 = Other Relationships

Courtesy: Gruninger

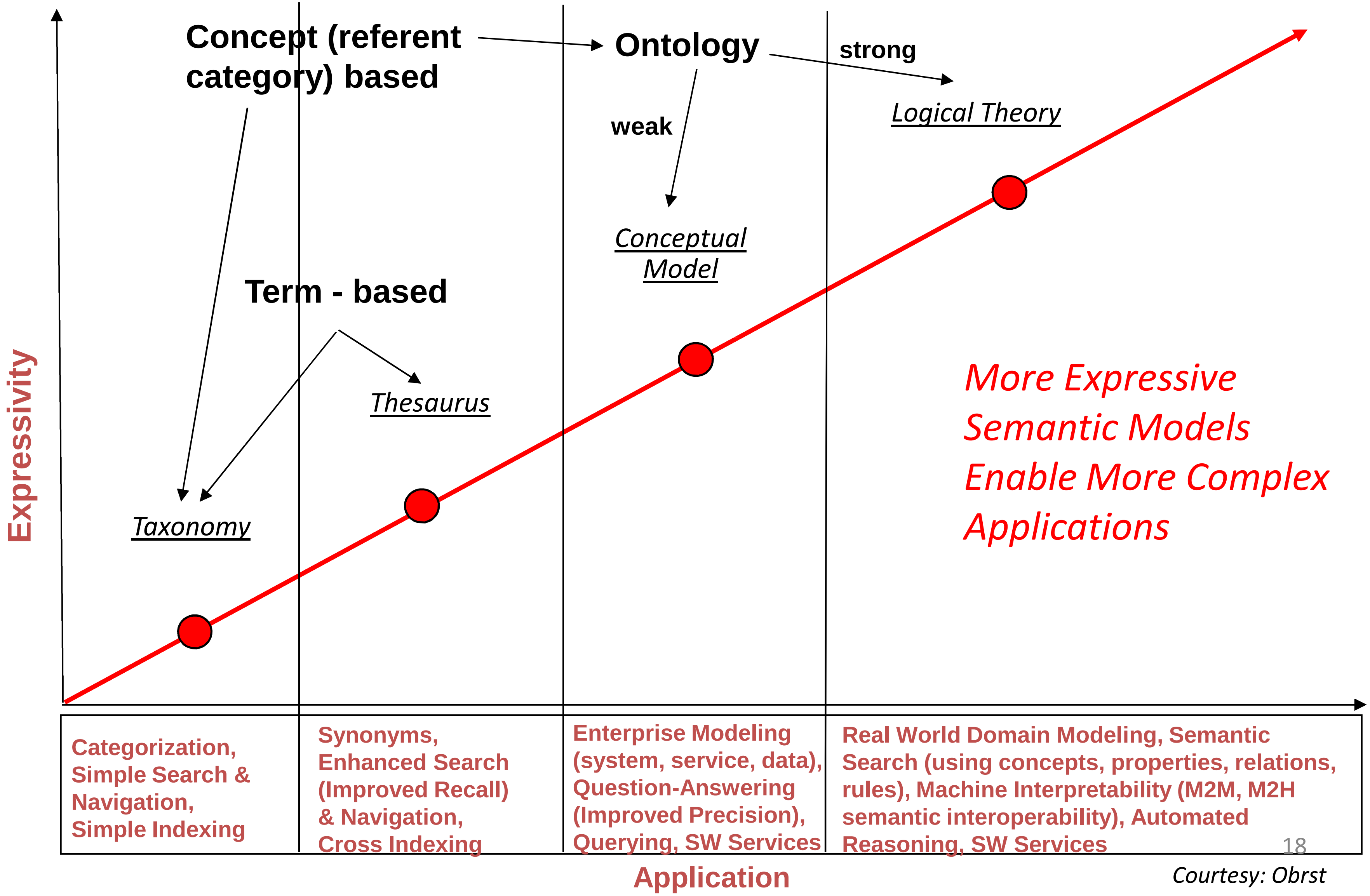
Ontology Spectrum: One View



Courtesy: Obrst

Ontology Spectrum: Application

see also http://vimeo.com/11529540



Concept (referent category) based

Ontology

strong

Logical Theory

weak

Conceptual Model

Term - based

Thesaurus

Taxonomy

Expressivity

More Expressive Semantic Models Enable More Complex Applications

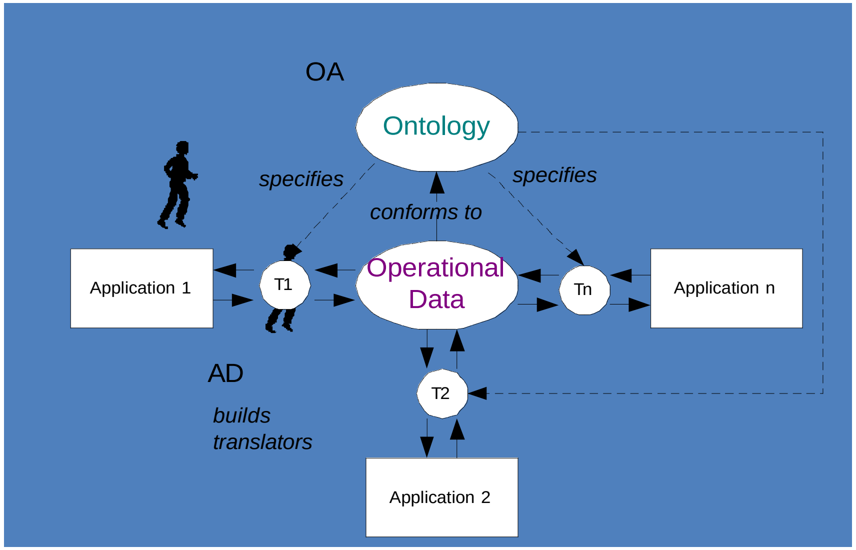
Categorization, Simple Search & Navigation, Simple Indexing	Synonyms, Enhanced Search (Improved Recall) & Navigation, Cross Indexing	Enterprise Modeling (system, service, data), Question-Answering (Improved Precision), Querying, SW Services	Real World Domain Modeling, Semantic Search (using concepts, properties, relations, rules), Machine Interpretability (M2M, M2H semantic interoperability), Automated Reasoning, SW Services

Application

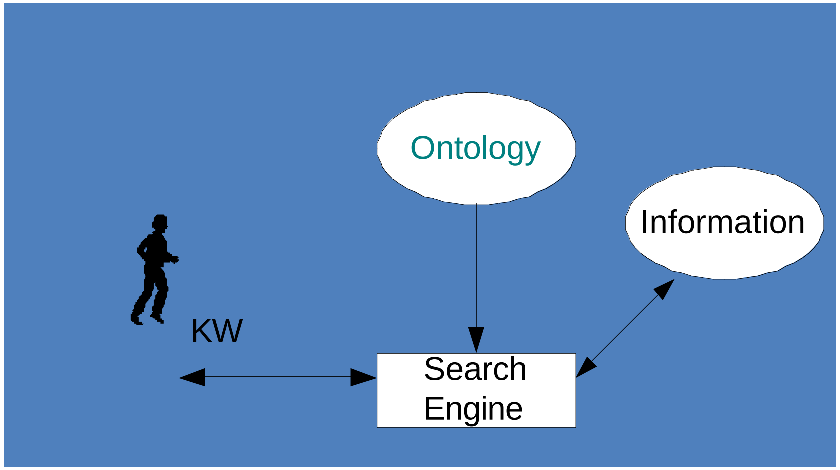
Courtesy: Obrst

Ontology Application Scenarios

- Common Access to Information
  - information required by multiple agents
  - expressed in wrong terms/format
  - ontology used as agreed standard, basis for converting/mapping
  - ***Benefits*: interoperability, more effective use/reuse of knowledge**



Ontology-Based Search
  - Ontology used for concept-based structuring of information in a repository
  - ***Benefits: better information access***



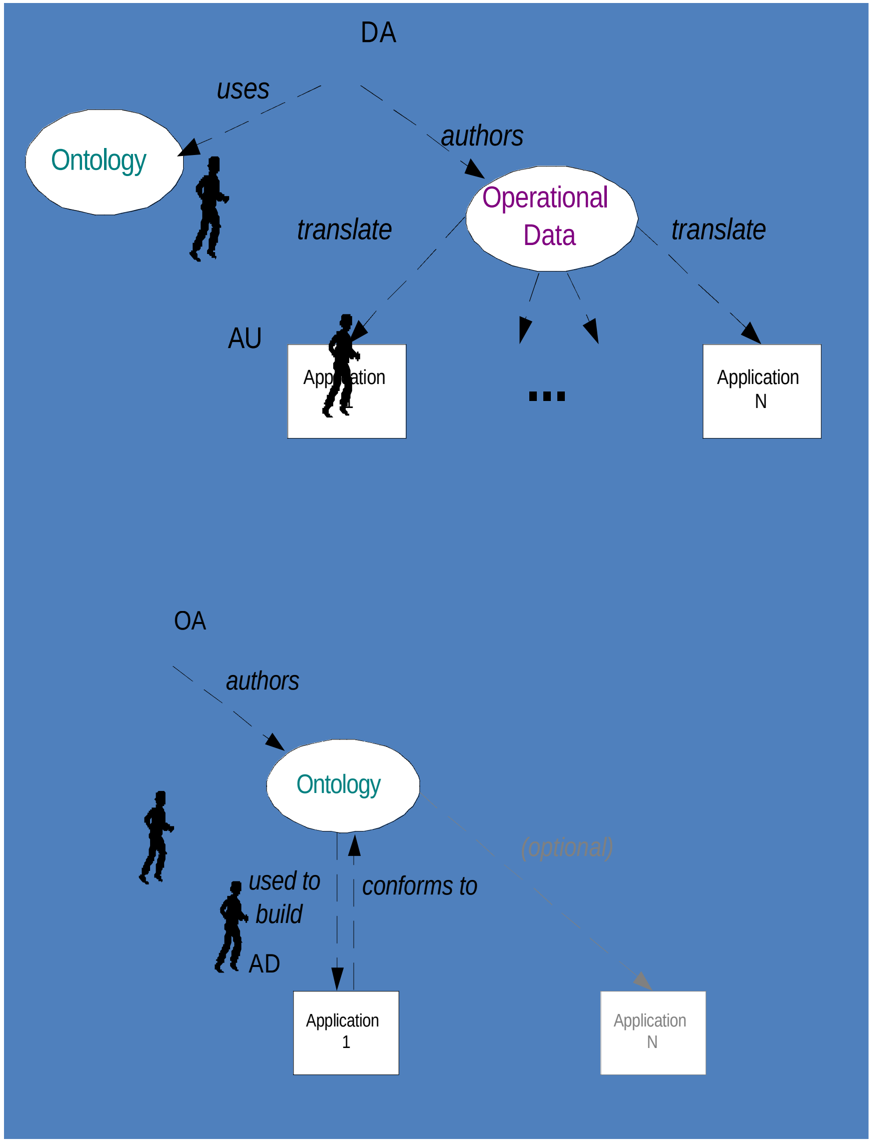
Courtesy: Gruninger

More Application Scenarios

- Neutral Authoring
  - artifact authored in single language, based on ontology
  - converted to multiple target formats
  - ***Benefits*: knowledge reuse, maintainability, long term knowledge retention**

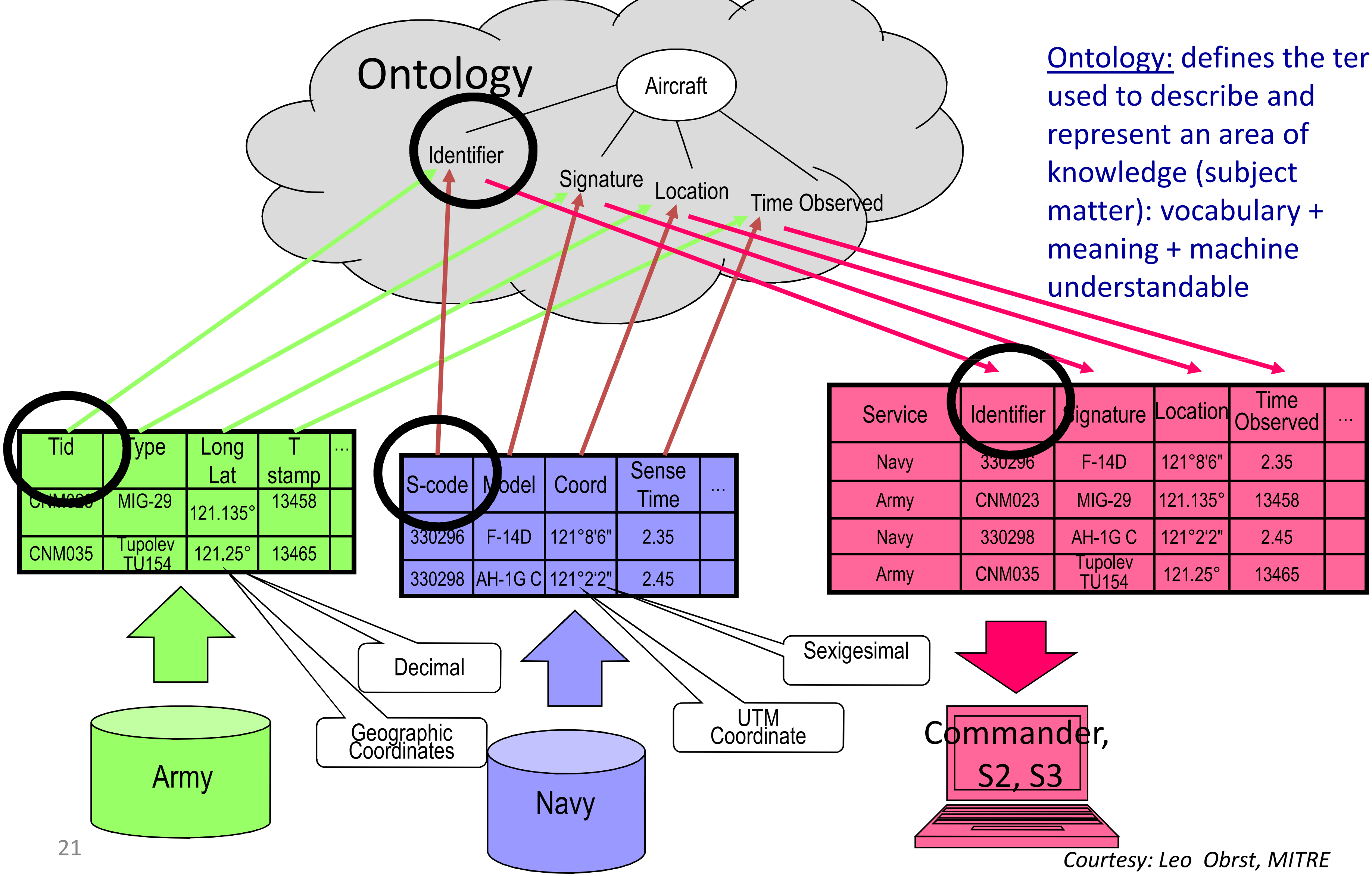
Ontology as Specification

  - build ontology for required domain
  - produce software consistent with ontology
    manual or partially automated
  - ***Benefits: documentation, maintenance, reliability, knowledge (re)use***



Courtesy: Gruninger

A Military Example of Ontology for Data Integration



Tid	Type	Long Lat	T stamp	…
CNM023	MIG-29	121.135°	13458	
CNM035	Tupolev TU154	121.25°	13465	

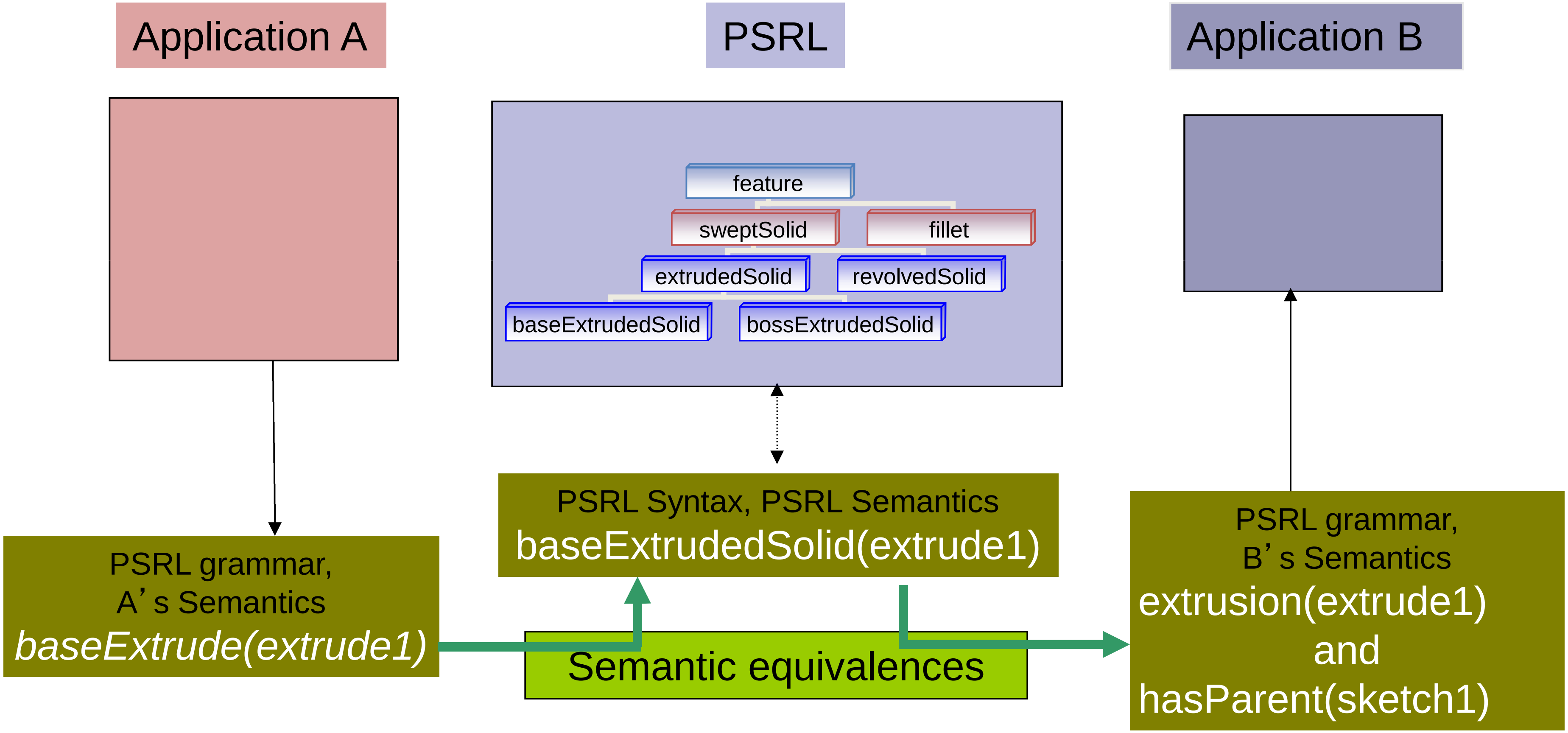
S-code	Model	Coord	Sense Time	…
330296	F-14D	121°8'6"	2.35	
330298	AH-1G C	121°2'2"	2.45	

Service	Identifier	Signature	Location	Time Observed	…
Navy	330296	F-14D	121°8'6"	2.35	
Army	CNM023	MIG-29	121.135°	13458	
Navy	330298	AH-1G C	121°2'2"	2.45	
Army	CNM035	Tupolev TU154	121.25°	13465	

Ontology: defines the terms used to describe and represent an area of knowledge (subject matter): vocabulary + meaning + machine understandable

Courtesy: Leo Obrst, MITRE

Interoperability Example



Application A

PSRL

Application B

feature

sweptSolid

fillet

extrudedSolid

revolvedSolid

baseExtrudedSolid

bossExtrudedSolid

PSRL grammar,
A’s Semantics
baseExtrude(extrude1)

PSRL Syntax, PSRL Semantics
baseExtrudedSolid(extrude1)

Semantic equivalences

PSRL grammar,
B’s Semantics
extrusion(extrude1)
and
hasParent(sketch1)

Courtesy: Lalit Patil, Deba Dutta &Ram D. Sriram

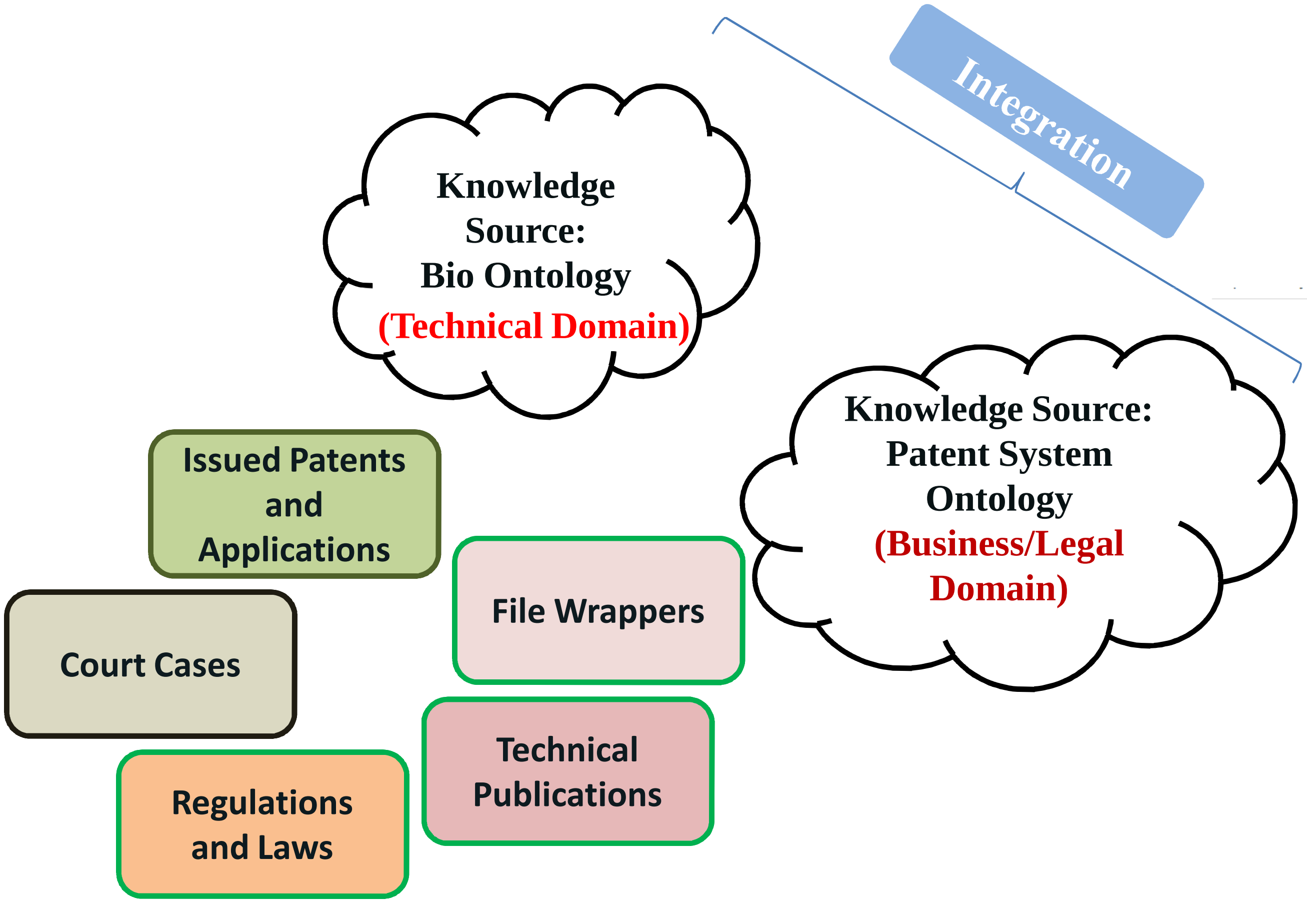
BIO-REGENT

Bioportal
(bioportal.bioontology.org)
NCI Thesaurus

Patent Document

Scientific Publication

Court Case



Integration

Knowledge Source:
Bio Ontology
(Technical Domain)

Knowledge Source:
Patent System
Ontology
(Business/Legal
Domain)

Issued Patents and Applications

Court Cases

File Wrappers

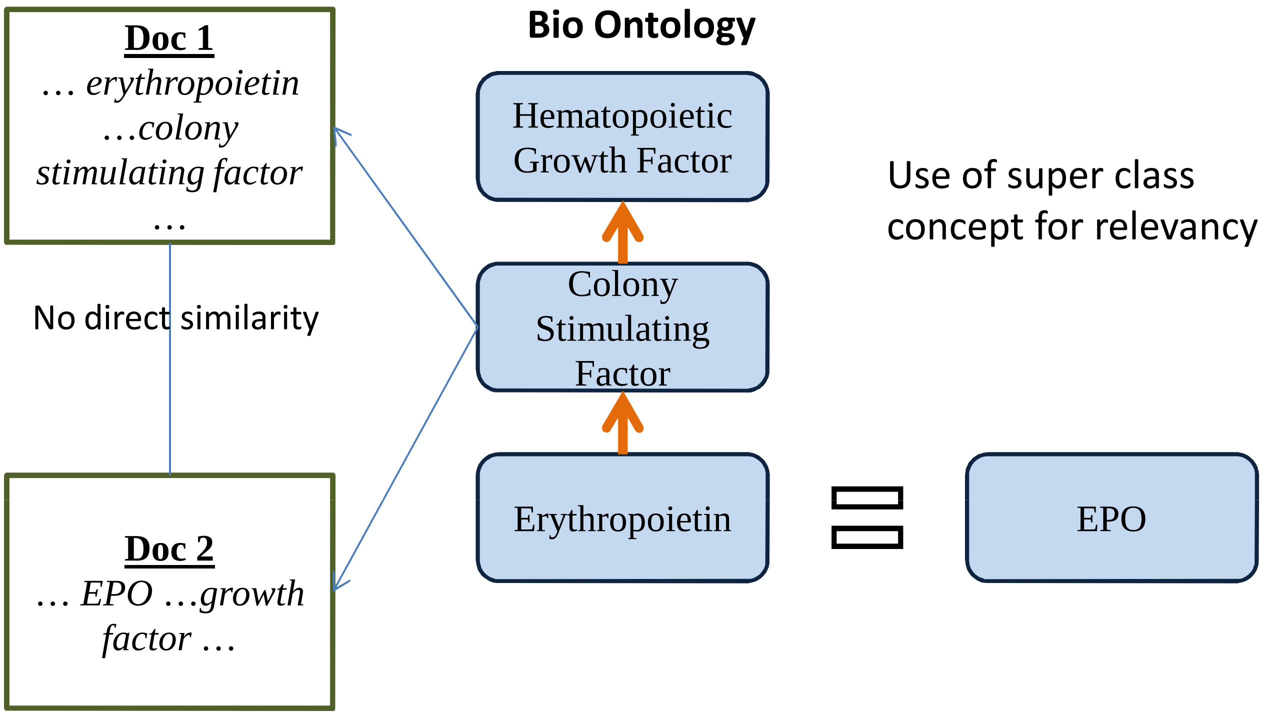
Technical Publications

Regulations and Laws

Siloed Patent System Information

Courtesy: Kincho Law (partial support from NIST)

Using Concept Hierarchy to determine Relevancy


Doc 1
… erythropoietin …colony stimulating factor …
No direct similarity
Doc 2
… EPO …growth factor …
Bio Ontology
Hematopoietic Growth Factor
Colony Stimulating Factor
Erythropoietin
EPO
Use of super class concept for relevancy

- Direct term based matching cannot relate the two documents
- Bio-ontology reveals that EPO and erythropoietin are synonymous
- Class hierarchy provides concepts (such as colony simulating factor) useful for determining relevance between documents (with appropriate weighting scheme)

Courtesy: Kincho Law

Goal of 2014 Summit

- Provide an opportunity for building bridges between the Semantic Web, Linked Data, Big Data, and Applied Ontology communities.
  - How are ontologies actually being used in Semantic Web and Big Data applications, and what are the challenges that these communities are encountering while developing ontologies?
  - How can the Semantic Web and Big Data communities share and reuse the wide array of ontologies that are currently being developed?
  - To what extent can automation and tools help overcome ontology engineering bottlenecks?

Ontology Summit 2014 Symposium Overview

- Virtual Symposium (Seminar) + 2 Day Workshop at NCO_NITRD (Arlington, Virginia)
- Virtual symposium: Every Thursday from 12:30pm-2:30pm EST (9:30am-11:30am PST), started on 2014-01-16.
- Dates for physical workshop were April 28th and 29th, 2014
- All talks were recorded and available on the Ontolog forum
- Summit results summarized and a communiqué was published (see website for previous reports)
- URL: http://ontolog.cim3.net/OntologySummit/2014/about.html

Overall Organization

- Summit General Co-chairs
  - *Michael Gruninger & Leo Obrst*
- Symposium Co-chairs
  - *Tim Finin & Ram D. Sriram*
- Communique and Publications
  - *Lead-Editors: Michael Gruninger & LeoObrst - Co-champions: Todd Schneider, Francesca Quattri*
- Community Resources (Library, Data Collection, Ontology Repository, etc.)
  - *Co-champions: (Amanda Vizedom), Oliver Kutz*
- Outreach (includes Sponsor Relations & Website Development)
  - *Co-champions: Amanda Vizedom (outreach and sponsor relations), Marcela Vegetti (website), Simon Spero (psmw-site),(Matthew West - adv)*
- Program management (includes operations, logistics, production)
  - *Co-champions: Peter Yim, Christi Kapp*
- Co-organizers
  - Ontolog, NIST, NCOR, NCBO, IAOA, NCO_NITRD

Tracks (or Themes) & Champions

- Track A: Common Reusable Semantic Content
  - *Mike Bennett, Gary Berg Cross, Andrea Westerinen*
- Track B: Making use of Ontologies: Tools, Services, and Techniques
  - *Christoph Lange and Alan Rector*
- Track C: Overcoming Ontology Engineering Bottlenecks
  - *Pascal Hitzler, Matthew West, Krysztof Janowicz*
- Track D: Tackling the Variety Problem in Big Data
  - *Ken Baclawski and Anne Thessen*
- Track E: Hackathon
  - *Dan Brickley and Anatoly Levenchuk (Adv: Ken Baclawski)*

Keynote Speakers & Panel Participants

- Dr. Farnam Jahanian, Assistant Director, CISE, NSF
- Mr. Daniel Kaufman, Director, Information Innovation Office, DARPA
- Dr. Philip Bourne, Associate Director for Data Sciences, NIH
- Panel Participants: Carol Bean (NCBO), Tim Finin (UMBC), Mark Fox (Univ. Toronto), Frank Olken (NSF), Ashit Talukder (NIST)

SUMMARY

Ontology Summit 2014 – Statistics

- Co-organizers: 6
- Organizing committee Members: 28
- Advisory Committee Members: 93
- Co-sponsors: 10
- [ontology-summit] list subscribers: 716
- Twitter followers: 97 *(new!)*
- Communique co-editors: 22
- Virtual org sessions: 12

Electronic Messages exchanged:
604(disc) + 456(org) = 1060

Virtual community sessions: 21

Hackathon-Clinic projects: 6

Two-day Symposium
registrants: 82(o) 63(v)
attendees: ~42(o) 28pk(v)

Presentations made: 111

Communique endorsements:
84 (as at end-day 2015.05.14-5:00pm PDT)

Courtesy: Yim

Lessons Learned

- Using ontologies with Big Data and the Semantic Web raises questions about scalability and the expressiveness of the underlying ontology representation languages.
- Reusability of semantic content is a critical challenge
- The Semantic Web and Big Data provide great opportunities for ontology-based services, but also pose challenges for tools for editing, using, and reasoning with ontologies, as well as techniques that address bottlenecks for the engineering of large-scale ontologies.
- For a summary read the communiqué (available at http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2014_Communique)

Back Up Slides

Track A: Common Reusable Semantic Content

- Focused on issues related to reuse and possible solutions such as:
  - Improving ontology repositories and tools
  - Building on smaller, more accessible semantic components
  - Discussing modularization and various exemplary ontologies and vocabularies
  - Identifying design patterns and best practices
- Defining metadata information to enable use/reuse
- Inputs:
  - 2 Track A presentation sessions, Jan. & March 2014
  - Email dialogs and track community page

Speakers & Their Presentations

1. MikeBennett **(EDM Council)** Overview of the track
2. Dr. GaryBergCross (SOCoP) - "Use and Reuse of Semantic Content: The Problems and Efforts to Address Them - An Introduction"
3. Professor PascalHitzler (Wright State U) - "Towards ontology patterns for ocean science repository integration"
4. Ms. AndreaWesterinen (Nine Points Solutions) - "Reuse of Content from ISO 15926 and FIBO"
5. Ms. MeganKatsumi & Professor MichaelGruninger (U of Toronto) - "Reasoning about Events on the Semantic Web"
6. Dr. JohnSowa (VivoMind Intelligence) - "Historical Perspectives: On Problems of Knowledge Sharing"
7. Professor MichelDumontier (Stanford BMIR) - "Tactical Formalization of Linked Open Data"
8. Mr. KingsleyIdehen (OpenLink Software) - "Ontology Driven Data Integration & Big Linked Open Data"

Sharing and Reuse

- Reuse versus sharing …
  - Re-use: What does it take to make use of the work of others instead of having to re-invent?
  - Share-ability: How do you create an artefact in order for someone else to be able to re-use it?
- *How to* re-use versus What *makes* something re-usable?
- Reuse issues are not unique to ontologies/schemas
  - Parallels and differences with software reuse
  - Requires that the concepts (+ relationships, axioms and rules), assumptions and expression(s) of the included content meet a need, and can fit into the re-user’s implementation
- Why reuse?
  - Reduce the development effort (by developing less)
  - Expand the benefit (improve the ROI) of the original content
  - Improve the quality of the original content (by identifying and eliminating errors)

Reuse

- For successful reuse of semantic content it is important to understand how content is being used, with what methods to coordinate reuse are available and what tools are helpful.
- Tooling for modularity, documentation, etc. is critical
  - Broader use by mainstream efforts including Big Data is bottlenecked in part by the paucity of semantic tools integrated into mainstream tools along with the inherent learning curve of understanding semantics.
- In practice reuse is dependent on both the availability of well-documented content AND tooling that supports finding and incorporating this range of content.

Tracks (or Themes) & Champions

- Track A: Common Reusable Semantic Content
  - *Mike Bennett, Gary Berg Cross, Andrea Westerinen*
- **Track B: Making use of Ontologies: Tools, Services, and Techniques**
  - ***Christoph Lange and Alan Rector***
- Track C: Overcoming Ontology Engineering Bottlenecks
  - *Pascal Hitzler, Matthew West, Krysztof Janowicz*
- Track D: Tackling the Variety Problem in Big Data
  - *Ken Baclawski and Anee Thessen*
- Track E: Hackathon
  - *Dan Brickley and Anatoly Levenchuk*

Research Questions

- How can tools and techniques scale to the Web?
- How can services benefit from tapping into the Web?
- How can they help to make Big Data manageable?

First Session (2014-01-30)

- **TillMossakowski**: scaling an ontology **tool** suite (Hets/Ontohub) from “reasoning in the small” to the Web
- **ChrisWelty**: the potential of linking Big Data to ontological reasoning, as demonstrated by the IBM Watson natural language question answering **service**
- **AlanRector**: OWL and alternative modeling **techniques**, reviewed from the perspective of engineering knowledge-rich systems.

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

Second Session (2014-03-13)

- **MikeBergman**: OSF, an enterprise platform that integrates and enhances several well-known ontology **tools**
- **JoseMariaGarcia**: combining linked data technology with web **services**
- **MariaPovedaVillalon**: a **technique** for engineering linked data vocabularies, i.e. lightweight ontologies that scale to the Web

http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2014_03_13

Should Ontologies Cover Everything?

- Traditional ontology languages assume universal knowledge. OWL is good for this
- In the real world, knowledge is often contingent, accidental or particular.
- Template formalisms such as frames, UML or rules are good for this.
- Translations across formalisms not yet well understood
- RDF(S) + SPARQL usage outnumbers OWL usage

  . . . but users are often ignorant of formal semantics.

  Still it copes well with heterogeneous data (variety)

Is OWL still useful?

Yes!

- E.g., in the OSF, using OWL allows for
  - duplicate names
  - incomplete information (thanks to open world assumption)
  - extensibility to multiple schemas
- Lots of tools and techniques (but most date back to small, hand-made ontologies):
  - limited to single or few formalisms
  - similar to knowledge silo-ing
- Can use OWL more creatively
  - e.g. take inspiration from template formalisms
- OntoIOp translates between OWL and other formalisms

Beyond a Single Ontology Language

- OntoIOp supports alignments and reasoning across ontology languages.
- Not yet “big” w.r.t. volume and velocity
  . . . but w.r.t. variety
- OntoIOp retrofits linked data conformance (e.g. IRI identifiers) into pre-Web languages
- Growing tool support: Ontohub (→ Hackathon)

RDF as a Knowledge Representation Foundation

RDF is the “native language” of Linked Data:

- enforces a low ontological commitment

. . . but still allows to link to complex descriptions

E.g., the Open Semantic Framework (OSF) uses a single, internal, canonical data model (RDF and some OWL):

- representing structured, semi-structured, unstructured data
- data structures translate into web widgets; ontologies
  - inform interface displays
  - define component behaviors
  - guide visualization template selection and content

Linked Web Services

Web services:

- Service provider registers service in central registry
- Service consumer finds service . . .

. . . and communicates with it to execute it

Semantic web services go beyond syntactic descriptions (e.g. WSDL) - previous state:

- web services exchanging heavy XML messages over SOAP
- semantics-first modeling using expressive WSMO or OWL-S ontologies

Face the reality:

- lightweight REST interfaces much more popular
- describe their semantics bottom-up in a linked data
- style: **Linked Services** (e.g. Linked USDL lightweight ontology)

Engineering Vocabularies

“Vocabulary” = “Lightweight Ontology”

Linked Open Terms, an agile engineering technique:

- determine the terms needed to describe your dat
- look for them in existing vocabularies (a lot exist on the Web!)
- create your own when necessary, but link to existing ones
- continuous evaluation

Conclusion

- Lightweight means Scalable
  - Heavyweight semantic web services have failed
  - A little RDF goes a long way
  - Even vocabularies can be engineered systematically
- Remaining Challenges
  z
  - Visualization
  - Scalability of reasoners
  - Requirements for ontology-based tools, services and techniques in a big data world still unclear.

Tracks (or Themes) & Champions

- Track A: Common Reusable Semantic Content
  - *Mike Bennett, Gary Berg Cross, Andrea Westerinen*
- Track B: Making use of Ontologies: Tools, Services, and Techniques
  - *Christoph Lange and Alan Rector*
- **Track C: Overcoming Ontology Engineering Bottlenecks**
  - ***Pascal Hitzler, Matthew West, Krysztof Janowicz***
- Track D: Tackling the Variety Problem in Big Data
  - *Ken Baclawski and Anne Thessen*
- Track E: Hackathon
  - *Dan Brickley and Anatoly Levenchuk*

Mission and Scope of Track C

The mission of track C is to **identify bottlenecks that hinder the large-scale development and usage of ontologies and identify ways to overcome them**.

BOTTLENECKS:

- Ontology engineering processes that are time consuming,
- Social, cultural, and motivational issues
- Modeling axioms or knowledge representation language fragments that cause difficulties in terms of an increase in reasoning complexity or reducing the reusability of ontologies
- The identification of areas and applications that would most directly benefit from ontologies but have not yet considered their use and development.

Report from Track C Session I (2014/02/06)

Session I title: Strategies and Building Blocks

Speakers:

Prof. Werner Kuhn (University of California, Santa Barbara)
"Abstracting behavior in ontology engineering"

Prof. Aldo Gangemi (University Paris 13 and ISTC-CNR Rome)
"Knowledge Patterns as one means to overcome ontology design bottlenecks"

Mr. Karl Hammar (Jönköping University)
"Reasoning Performance Indicators for Ontology Design Patterns"

Ontology Engineering Bottlenecks – Session II

Oscar Corcho (Universidad Politecnica de Madrid)

10 basic rules to overcome ontology engineering deadlocks in collaborative ontology engineering tasks

Dhaval Thakker (University of Leeds)

Modeling Cultural Variations in Interpersonal Communication for Augmenting User Generated Content

Peter Haase (Fluid Operations)

Developing Semantic Applications with the Information Workbench – Aspects of Ontology Engineering

Reflections

- Bottlenecks and barriers to the use of ontologies in Big Data and the Semantic Web are many and various – there is no clear pattern
- Reuse (rather than reinvention) of ontologies and ontology patterns offers promise in overcoming development bottlenecks, but comes with its own bottlenecks and barriers
- Automation of tedious and repetitive tasks is demonstrated to be effective, but there is a need for more tools that deliver this automation

Tracks (or Themes) & Champions

- Track A: Common Reusable Semantic Content
  - *Mike Bennett, Gary Berg Cross, Andrea Westerinen*
- Track B: Making use of Ontologies: Tools, Services, and Techniques
  - *Christoph Lange and Alan Rector*
- Track C: Overcoming Ontology Engineering Bottlenecks
  - *Pascal Hitzler, Matthew West, Krysztof Janowicz*
- **Track D: Tackling the Variety Problem in Big Data**
  - ***Ken Baclawski and Anee Thessen***
- Track E: Hackathon
  - *Dan Brickley and Anatoly Levenchuk*

Handling Variety

- Development of new storage and indexing strategies for handling volume and velocity
  - “Map Reduce” was developed in 1994. [2]
- Development of techniques for handling variety
  - Schema mapping
  - Controlled vocabularies
  - Knowledge representations
  - Ontologies and semantic technologies
- Connection between these two?
  - Surprisingly little collaboration and communication.
  - A notable exception is the early work starting in 1992 on representing biological research papers. [3]

Track D: Speakers

- Eric Chan - **Enabling OODA Loop with Information Technology**
- Nathan Wilson - **The Semantic Underpinnings of EOL TraitBank**
- Ruth Duerr - **Semantics and the SSIII Project**
- Mark Fox - **Variety in Big Data: A Cities Perspective**
- Malcolm Chisolm - **Data Governance to Manage Variety in Big Data**
- Dan Brickley - **Schema.org, FOAF and Linked Data: Lessons for Web-scale vocabulary deployment**
- Rosario Uceda-Sosa - **Big Data, Open Data and the Smart City**

Track D: Challenges Posed

- Little collaboration between the communities
- Big Data focus on volume and velocity, assuming someone else will handle variety
- Tool incompatibility
- Incompatibility between statistical and logical techniques (hybrid reasoning gap)