Ontology Summit 2016 Communique
Ontologies within Semantic Interoperability Ecosystems

Lead Editors: Donna Fritzsche, Michael Gruninger
Co-Editors: Ken Baclawski, Mike Bennett, Gary Berg-Cross, Leo Obrst, Todd Schneider, Ram Sriram, Mark Underwood, Andrea Westerinen

Introduction

In today's digital information environment, interoperability between systems is a ubiquitous need and expectation. Businesses, organizations and research groups seek to create optimal experiences, minimize operational overhead, reduce costs, and drive future innovations utilizing:

- the Internet of Things,
- the Smart Grid,
- Intelligent Agents and Machine Learning,
- Personalized Services and Content Delivery, and
- Data Mining Techniques.

Both syntactic and semantic interoperability across systems and applications are necessary. In practice, however, Semantic Interoperability (SI) is difficult to achieve.

Ontologies and related reasoning systems (both terms used in the broadest sense) are key to the facilitation of semantic integration and interoperability. But several key questions are involved. How do we define the tools, methodologies and frameworks which support this interoperability? What is required to achieve optimal performance across applications and domains? How do we frame the conversation when discussing the role of ontologies in support of interoperability with various stakeholders?

Ontology Summit 2016 explored how ontologies and ontological methods can facilitate semantic Interoperability (SI). In this document we present an overview of semantic interoperability challenges. We discuss the ontological tools and methodologies which enable SI and we summarize findings abstracted from recent work covering 4 broad domains: healthcare/pharma, earth sciences, engineering, and finance/retail. Finally, we present design approaches, strategies and next steps for the field moving forward.

Semantic Interoperability Overview

What is Semantic Interoperability?

Interoperability is the ability for two or more entities to operate together to attain or meet a set of goals or operational objectives; while integration is a tight coupling or binding of entities, again (usually) to meet some operational objectives (e.g., functionality, speed, cost reduction, risk management, etc.). In the context of interoperability, semantics[1] is fundamentally interpretation within a specific context. Semantic interoperability is the ability for a receiver of information to

---

[1] "Semantics (from Ancient Greek: σημαντικός sēmantikós, "significant") is the study of meaning. It focuses on the relationship between signifiers—like words, phrases, signs, and symbols—and what they stand for, their denotation." (Wikipedia)

interpret or understand the contents in a manner relatively consistent with the sender's intended interpretation/meaning to meet (common) operational objectives (i.e., the context for and of the information).

The Role of Ontologies

The use of properly conceived and developed ontologies has been proposed as a key technology to support semantic interoperability. The origins of applied ontology in the Shareable and Reusable Knowledge Base (SRKB) project show how the semantic heterogeneity problem can be addressed. Within that framework, ontologies provide a set of terms and relations together with a computer-interpretable specification of the intended interpretations (meanings) of the terms. They are intended to support the semantic interoperability of information systems and data sources through a consistent use of the terminology in their respective ontologies. The promise held out by ontologies is that semantic interoperability and integration can be enabled by:

- · the establishment of base semantic representations via ontologies (class level) and their knowledge-bases (instance level),
- · the definition of semantic mappings and transformations among ontologies, and
- · the use of algorithmic methods that can determine semantic similarity to facilitate mapping between ontologies.

Taken together the proper use of ontologies and semantic mapping software mitigates loss of semantics (meaning) in information exchange among heterogeneous applications.

The Ontology Summit 2009 proposed an Ontology Usage Framework, which included several areas related to semantic interoperability. One was referred to as "information integration", in which multiple information resources are combined using ontologies to match concepts with similar meaning; examples include information aggregation and data fusion. A second relevant usage of ontologies is software interoperability, in which software systems communicate by exchanging messages that are composed using ontologies. Each system uses an ontology (either its own or a set of shared ontologies) to translate the exchanged messages.

Semantic Heterogeneity

The semantic heterogeneity problem is common in today's engineering and manufacturing industries and is thus a primary consideration when discussing semantic interoperability. A multifaceted notion, it arises from the need to share content and data within and across digital ecosystems. It is evident in every domain considered during Ontology Summit 2016.

Within digital ecosystems, a combination of devices, sensors, software applications and data sources are involved in operational practice. Across systems, domain information is:

- · described in multiple schemas,
- · described using vocabularies with varying and locally developed semantics,
- · implemented in different markup languages, and

·   based on models featuring different conceptualizations.

Data is also represented at different levels of granularity whose semantics are based on different models. As a result, information often cannot be shared between software applications, and data sources cannot be usefully combined. Additionally, the interaction experience of end-users, knowledge engineers, and subject matter experts (SME) is often compromised.

Products are becoming increasingly complex and exist in different digital and physical environments through their lifecycles: concept, design, development, testing, manufacturing/deployment, operation, upgrade, decommissioning. It is critical to understand the entities involved and the various software and physical environments in which they exist or pass through (such as CAD-CAM, testbed or factory), and most importantly, the associated relationships. In each environment or lifecycle stage, decisions are made that impact subsequent (and sometimes previous) stages or environments. Engineering and manufacturing such complex entities requires multiple disciplines to cooperate and share data and information used to make decisions. However, engineering disciplines have developed concepts, relations, terminologies, paradigms, and tools to meet their own needs. In doing so, implicit semantics and particular interpretations have become common, understood by practitioners in the field, and embedded in many standards, specifications, and tools (both software and hardware, e.g., machining). Thus when multiple disciplines need to collaborate and share information, interoperability among terminology, data, and tools becomes a problem. These problems lead directly to delays, errors, and increased costs in development, manufacturing/deployment, or operation.

A lack of interoperability is costly as overall system adoption rates and satisfaction suffer. It has been estimated by the Office of Financial Research that the lack of a common language in finance is a multibillion dollar problem for the industry. According to a report by the Center for Medicare and Medicaid Services, the United States spent nearly $3.0 trillion on healthcare in 2014, and this is expected to nearly double by the end of this decade. It is estimated that the efficient use of health information technology will result in considerable cost savings, in addition to saving saving thousands of lives per year. Improved doctor and patient experiences are an important intangible benefit of highly interoperable systems.

Consider a hierarchy of integration necessary to achieve interoperability. Syntactic integration exists at the lower end, structural in the middle, and semantic on top. To a large degree, many industry sectors and communities of practice have been converging on the lower and middle level of this hierarchy with common protocols and data formats to ensure the proper exchange of data. There is some belief that syntactic interoperability can be achieved through standardization such as controlled vocabularies. However, *robust* semantic interoperability relies on a common interpretation of the messaging and exchanged data, i.e., meaning remains invariant during the exchange among different domains and between multiple systems within a coupled ecosystem.

Kinds of Ontologies in the SI Ecosystem

Overview

Ontology Summit 2007 explored the wide range of semantic content (including taxonomies, thesauri, topic maps, conceptual models, and formal ontologies specified in various logical languages) that constitute the artifacts often referred to as ontologies. Although ontologies hold the promise of resolving the semantic heterogeneity problem, the variety of ontologies and semantic content, some in different formal languages, raises the possibility that in an undisciplined use they will only exacerbate the problem. Despite the increasing number and quality of ontologies there still is what has been graphically described as a "semantic mess." This remains a state of affairs because the domain information is heterogeneous and described in:
1. multiple schemas,
2. different vocabularies and markup languages, and
3. ontologies with different levels of data granularity and using different conceptualizations.

There are quite a few ontologies developed along the spectrum of semantic formality, but also along different degrees of comprehensiveness and completeness. How can this semantic content from across the ontology spectrum be used?

An example that illustrates how ontologies with varying levels of specificity can be used and extended is NASA's Semantic Web for Earth and Environmental Terminology (SWEET) ontology with about 6000 concepts in over 200 separate, modular ontologies. Such ontologies build on community efforts to develop standard vocabularies within domains to support data and system interoperability. SWEET can provide some basis for helping with tasks such as semantic tagging, however, it has few axioms to support reasoning and needs to be supplemented whenever used for advanced purposes. It is generally recognized that some early efforts are low on the spectrum of formal semantics and that more richly axiomatized ontologies capturing domain understanding can help to address this limitation. On the other hand, many terms, in biomedicine and geoscience for example, are somewhat semi-scientific. Examples include ideas about river, channel, water body or the relations of symptoms to disease. Scientists or healthcare providers simply don't have a full and clear understanding of these concepts especially in relation to other concepts that make up reality as a system. This reflects in part the complexity of reality and science's current state of limited understanding.

Upper Ontologies

The first Ontology Summit in 2006 was motivated by the need to resolve the semantic differences among ontologies, particularly with respect to their ontological commitments.
Upper ontologies, such as Basic Formal Ontology (BFO), Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), UpperCYC, ISO 15926, COSMO, and Suggested Upper Merged Ontology (SUMO) are tools that may aid in solving the semantic heterogeneity problem, since they explicitly axiomatize those concepts which are shared by a wide variety of more specialized

ontologies. The complete or partial adoption of an upper ontology and its modeling practices by lower ontologies may help to minimize the introduction of additional semantic inconsistencies to an ecosystem.

The use of upper-level ontologies can help mitigate semantic interoperability problems, but there remain several serious challenges.
1. There are several competing upper ontologies.
2. Many of the upper ontologies are difficult to understand, with complicated axioms, and abstractions that remain too far from real data.
3. They need to be harmonized with domain, reference, application and local ontologies.
4. An upper ontology, especially one developed in a top-down manner, may impose ontological commitments that may not be acceptable by all interested parties who have 'local' vocabularies and meanings.
5. Like artificial intelligence systems in the past, upper ontologies may also be too brittle, meaning that small changes are not easily incorporated or compromise the semantics.

Reference Ontologies

Reference ontologies reflect the base-level knowledge of a broad domain or the semantic consensus an industry sector. Growing out of efforts to represent deep knowledge of basic science in a principled way, they are intended to be reused and are not rigidly tied to an application's specific use cases and requirements. By design, they are created to facilitate integration across systems, repositories and data sources. Rather than serving as an upper ontology that helps *mediate* between other ontologies, a reference ontology serves as a means for mapping the terminology of multiple information systems and data to a common set of shared concepts. A classic example is The Foundational Model of Anatomy (FMA). Comprised of roughly 75,000 classes, 120,000 terms, and 168 relationship types, the FMA is a reference ontology that represents the structure of the human body. Properly conceived, a collection of reference ontologies can be viewed as orthogonal (non-overlapping), interoperable resources. This is roughly the model adopted by the Open Biological and Biomedical Ontologies (OBO) Foundry (comprised of ontologies from the molecular to organism level covering both biological structure and function).

A reference ontology approach has recently been adopted in finance where an increasing number of financial applications broadly rely on the Financial Industry Business Ontology™ (FIBO) which is an ontology created by Enterprise Data Management Council and the Object Management Group (OMG). FIBO specifies the definitions, synonyms, structure, and contractual obligations of financial instruments, legal entities and financial processes.

Application and Local Ontologies

An application ontology is "an ontology engineered for a specific use or application focus and whose scope is specified through testable use cases" (Malone & Parkinson 2010). The key point is

that they are created to meet project requirements. They may be applied to a local domain or cross several related domains. When application and local ontologies are created without a reference to broader, more principled approaches and concepts and/or reference ontologies, they may not be readily linked to other ontologies. As a result, some integration is supported by mapping or bridging concepts between different application ontologies. In turn, systematizing such bridging may provide some basis for formation of a reference ontology.

Bridge Ontologies

Related to the notion of reference ontologies (which inform and mediate between the terminologies of multiple systems), bridge ontologies are typically used to mediate between specific concepts of multiple ontologies. The mediation may be straightforward and can be accomplished by mapping between similar concepts. Often, though, the mediation requires the addition of a new concept that may be missing in the ontologies under consideration. The newly added concept provides a means to link two or more related concepts across the ontologies. An example is a concept of a process that 2 entities may participate in but which may not have been modeled in either ontology. An intelligent broker, for example, could know that the concept of "conductivity" is related to "salinity" and "seawater". When an intelligent broker is queried for salinity data it will display data on the conductivity of seawater if the user wants related terms to be included in the query results.

Although upper ontologies are often proposed to play this bridging role, there are other cases in which bridge ontologies capture the commonalities between various application and local ontologies within the same domain. Existing GeoScience standards, ontologies, models and associated vocabularies, for example, were typically developed in isolation and major problems exist when attempts are made to combine them. Some glue, such as bridge ontologies, is needed to integrate and harmonize these. The coverage of a bridge ontology may not be as broad as a reference ontology and may be lightweight- without extensive use of axioms. The use of a bridge ontology often makes clear the need for a more comprehensive model - for instance, a reference ontology or an application ontology with broader requirements.

Common Metadata Templates and Metadata Schema

A variety of efforts have made it possible to create, maintain, and promote common metadata attributes and schemas for annotating structured and unstructured content and data across websites, emails and repositories. The use of metadata attributes and schemas increases the likelihood that the content is findable, accessible and reusable. In consumer domains, such annotation facilitates personalization and the delivery of targeted content. One of the goals of these efforts is to provide a flexible framework to minimize the barriers to adoption and to maximize reusability.

Historically, this formal practice started with the Dublin Core set of 15 metadata elements (properties) to be used for annotation to facilitate search and findability. It has now grown into a

variety of efforts. Based on the Dublin Core Metadata Initiative (DCMI), the Darwin Core is an extension of the Dublin Core for biodiversity information (and has evolved into a more full-fledged reference ontology). This reflects growing efforts to add more semantics to metadata and structure to these efforts. More recently, Schema.org has addressed general use and the retail domain, specifically targeting search engine functionality. Under the Schema.org framework, schemas are represented as 'types' (e.g event, place, product, offer), each associated with a set of properties. Schema.org consists of 571 Types and 832 Properties. Within e-commerce, online retailers using markup information implement Schema.org's Offer and Product Entities. Efforts are underway to incorporate other standards into the Schema.org framework including FIBO. One current limitation in Schema.org utility is that it does not cover all domains including engineering.

Also of note, is the recent establishment of the Center for Expanded Annotation and Retrieval (CEDAR) which has the goal of providing a unified framework to create consistent and easily searchable metadata. CEDAR's target audience is scientific researchers and the initial target use case (from the Human Immunology Project Consortium) employs a standard template to annotate and describing laboratory studies and the associated biomedical data. Metadata templates created using CEDAR technology are stored in an openly accessible community repository and researchers can access that library to look for appropriate templates to annotate their studies. CEDAR's focus will be to address the full lifecycle of the metadata creation and entry process with the long-term goal of creating sophisticated, semantically enabled metadata ecosystems for target audiences and related communities.

Architectural Developments

Semantic Data Lakes

A data lake is a type of storage repository that holds a vast amount of raw data in its original (native) format until it is needed. While many repositories, such as hierarchical data warehouses store data in files or folders, a data lake uses a flat architecture to store data. Each data element in a lake is assigned a unique identifier and tagged with a set of extended metadata tags. A data lake can be queried for relevant data, using the ID. Like many storage and integration approaches, standard data lakes lack a shared approach to semantic interoperability, thus the depth and breadth of functionality that they facilitate is limited.

Solutions require integration of both concepts and data. Shared semantics are often required to build knowledge bases which pull data from a variety of sources. "The Semantic Data Lake extends the state of the art in big-data management to account for large scale integration of data, metadata, knowledge, and linked open data to support analytics across the spectrum of applications from precision medicine to accountable and learning healthcare systems." As described by Parsa Mirhaji, the approach taken at Montefiore Medical Center (a Montefiore, Franz, Intel and Cisco collaboration) utilizes healthcare-specific semantic elements such as medical vocabularies, taxonomies and ontologies to underpin and improve evidence generation and avoid analytic silos.

In related efforts, the Common Workflow Group is developing Semantic Annotations for Linked Avro Data (SALAD), designed to "provide a bridge between the record oriented data modeling supported by Apache Avro and the Semantic Web."

Information Modeling for Federation

Federation facilitates the use of independently conceived information sets for purposes beyond those for which the individual information sets were originally defined. The Semantic Information Modeling for Federation (SIMF) project's goal is to "help federate information across different authorities, vocabularies and formats." Current conceptual and logical information modeling approaches tend to be focused on a particular information modeling problem, using a particular technical approach. Examples of such technical approaches include object modeling, DBMS modeling and exchange schema modeling. SIMF seeks to address the problem of information federation by specifying standards for conceptual domain modeling, logical information modeling and model bridging relationships.

The Cloud Presents New Opportunities

Cross-cutting influences across the discipline of software engineering can limit or expand potential approaches chosen by system architects. Almost without exception, these are trends that are initiated outside of the ontology or semantic web communities. Two important consequences of the "cloudification" of computing are DevOps and an API-first design philosophy (espoused by Intel's Brian Krzanich). While Service Oriented Architecture (SOA) and "composable services" introduced many of the same concepts in earlier generations (both DevOps and API-first steal from well-burnished concepts), the level of adoption of the newer processes across software and data providers is unprecedented.

"API-first" is part buzzword, part design pattern. The concept is described in the trade press as "The Future of Coding." API-first is partly driven by entrepreneurial considerations and partly by the Internet of Things. Also framed as "microservices," the approach has been characterized by Chet Kapoor in *Techcrunch*:

> As we move into a more interconnected world, amazing new possibilities emerge. Developers like to consume "bite-sized" stuff. Amazon popularized this approach — they told developers what the system does and got out of the way. For tech companies, the 'telling' will be handing over APIs. It's no wonder we've moved toward microservices that enable best-of-breed platforms to thrive.

On a related note, firms like PBS, VMware, Cisco and Spotify say they have adopted the RESTful AI Modeling Language (RAML), which offers the promise of supporting the full API lifecycle "in a human readable format with code and design pattern reuse."

Methods which Facilitate Semantic Interoperability

Ontology Reuse and Modular Design

From the Ontology Summit 2014 Communique, reuse "can be defined as the ability to include content from one source in another, or simply to be inspired by the content in a source". Reuse of an ontology's concepts and semantics is one way to achieve interoperability because the content is the same across all the uses. We are seeing movement in this direction with the previously mentioned vocabularies like Schema.org, which is extended by both the GoodRelations and FIBO ontologies. For instance, the full expressivity of GoodRelations can be accessed directly from the Schema.org namespace in Microdata syntax. In general, however, there is the perennial problem that *new* ontology development is preferred over reuse. The 2014 Communique (Section 3.1.) addressed factors limiting reuse.

- · Existing ontologies are too large and complex, not sufficiently granular and may not be sufficiently documented.
- · It is difficult for a potential user to determine what aspects of an ontology are valuable, since those aspects must first be found and then they must be understood.
- · In addition, semantic languages such as OWL do not support partial import of an ontology and therefore force the inclusion of many more concepts than may be deemed appropriate or (worse) correct.

Creating ontologies from more granular, cohesive and self-contained modules and design patterns would go a long way to improve ontology evolution and reuse. Smaller modules and patterns could be combined together and extended to form larger ontologies. For example, FIBO has employed modularization to create a set of ontologies building up to the complete ontology for financial instruments. But, this is not the norm. To make this approach work with today's ontologies, we may first need to decompose them into more manageable building blocks.

Vocabulary Harmonization

Naming conventions and vocabulary terms that can be understood by data providers and consumers are important when considering the knowledge management aspects of facilitating research, scientific, industry and government projects. Increasingly, vocabulary harmonization is needed since terms are usually collected during different activities or projects, in isolation from one another. This results in vocabularies that have the same conceptual scope, but are represented with different terms, use different formats and formalisms, and are published and stored with alternative access methods. For example, most water quality vocabularies combine multiple concepts into a single term. Thus a term for a given observation may conflate the substance (or taxon) with the medium (e.g. water) observed, with the procedure used for observation and the units used for measurement. The lack of orthogonal elements, makes vocabulary harmonization difficult as does the lack of definitions and poor maintenance of vocabularies.

Increasingly controlled vocabularies (standardized sets of terms) and their human readable definitions are accessible via Linked Data and consist of concepts locatable via URIs. This practice makes access to terms and their definitions uniform. In addition, some tools facilitate the storage of various vocabularies in a common repository. The National Environment Research Council (NERC) Vocabulary Server, for example, is a tool that provides access to lists of standardized terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider Hydrology and Earth Sciences communities.

To a modest degree metadata from such tools can be used to identify and label data which helps mitigate some of the problems of ambiguities associated with data markup. A growing practice is to add at least some basic thesaurus metadata using broader, narrower, related relations and SKOS properties such as: prefLabel, altLabels, text definition, source.

This remains, however, work largely at the lower level of SI. Despite these improvement we need to provide richer metadata which would ideally use generalized semantic relationships beyond thesaurus relations. Without such foundational, formal semantics there remains limited ability for the vocabularies to be interpreted by computers for harmonization with other vocabularies.

Vocabulary harmonization requires conceptualization and development of uniform schema to show the relations of concepts underlying the definitions of terms. Examples of such an effort are the development of the OGC Observations and Measurements ontology and the W3C's Semantic Sensor Net ontology.

In summary, despite some progress, vocabulary harmonization remains a bottleneck and is impeded by lack of such things as bridge, core and reference ontologies.

Ontology Mapping

In the absence of a single upper or reference ontology, many approaches to semantic interoperability rely on the specification of mappings between local ontologies that are used by the various information systems. There is an extensive and still-growing body of research on ontology mappings with OWL ontologies. Although there has been less work in the context of more expressive languages such as Common Logic, recent standards such as the Distributed Ontology, Modelling and Specification Language (DOL) are expected to provide the foundations for ontology mappings not only between ontologies that are axiomatized in the same language, but also between ontologies that are axiomatized within different languages (e.g. between OWL ontologies and Common Logic ontologies).

Even with metalanguages and techniques for expressing ontology mappings, the automatic generation of ontology mappings remains a substantial challenge, particularly for the more expressive ontology representation languages.

Ontology Design Patterns

The artifacts collectively referred to as design patterns were introduced by Christopher Alexander in the 1970's. Design patterns are comprised of abstract solutions and shared guidelines to specific types of design problems and associated use cases within specific contexts. Design patterns have been used as a tool in a variety of disciplines including architecture and software development. The more recent Ontology Design Patterns (ODPs) grew out of this tradition. An ODP is a modular, self-contained building block which represents a reusable solution to a frequently occurring modeling and ontology development problem. An ODP can be used to name, organize, and conceptualize highly related pieces of design knowledge. ODPs have been applied to a diverse set of domains including the Earth Sciences and Biomedical fields.

Consider the following design approaches and best practices when creating ODPs.

1. A well-designed ODP is grounded in bottom-up definitions and data, but abstracted to a level that can be applied more generally. A pattern formed this way can then be used as a building block of a more complex "core" ontology. Overtime, the general community can add required axiomatized commitments to solidify the base pattern at the core level and make it more specialized as needed. Subtyping general patterns is sometimes required by use cases and local complexities. But a pattern's relatively straightforward formalization goes noticeably beyond the typical subsumption/taxonomic hierarchies with minimal semantics that reduce ontologies to mere structured nominalization of terms. Instead, a pattern is simultaneously understandable to humans and formalizable in languages such as OWL to support a useful range of logical inferences.

2. As a best practice, the domains and ranges of relations within an ODP should be carefully considered so as to not over constrain the formal semantics of the general pattern. A lightly axiomatized pattern can then be related as a complex of ODPs which can then be used as a skeleton to create larger ontologies addressing broader problems.

3. As noted, modular patterns are flexible and extensible which allows them to serve an integration role. Integration across an expanse of datasets and local domains can be accomplished by careful and consistent application of ODPs. In response to targeted needs, local applications can extend the pattern by adding or changing axioms and aligning these to local data and needs. In this process local ontology development is the glue, and a local view of the pattern allows the connection of a data source to the patterns via a specific and explicit mapping. To handle different and/or more refined interpretations which serve targeted requirements, an appropriate mapping/alignment between a "local" vocabulary and the pattern is made. This "local view" also employs a very minimalistic schema (class names, property names, simple domain and range axioms). To facilitate this approach, the core conceptualization is separated from "nomenclature" issues, because local view vocabulary terms may be data repository-specific and need not be the same as the terms used to label concepts in a pattern. As a practice, mapping from data to the pattern can be expressed in rules that help populating the patterns. Data providers can then populate the global schema (pattern collection) by simply populating a local view.

In summary, a well-designed set of ODPs provides knowledge engineers and SME's with a foundational tool for capturing, developing, and refining the semantics (e.g. axiomatic enrichment) within a local domain.

Towards the Right Mix

In the following sections we discuss the use of ontologies and related methodologies within the context of each track of this year's summit.

Engineering and Manufacturing

Manufacturing and engineering are related but different. Manufacturing not only uses and creates information, but must also deal directly with physical entities, which many engineering disciplines only address abstractly or indirectly. This additional need imposes greater requirements on ontologies developed for this area. They must be able to support units of measure, measurements, temporal relations, and constraints.

The perennial issue of the expressiveness of ontology representation languages also impacts the use of ontologies for semantic interoperability. OWL may be insufficient to meet the needs of semantic interoperability in the manufacturing domain, both because of its limited expressiveness (in contrast to other ontology representation languages e.g., Common Logic) and its (operational) Open World assumption. In some areas of engineering and manufacturing a (operational) Closed World assumption is valid and can simplify and make more acceptable the use of ontologies by allowing reuse of existing techniques and tools. For some domains and use cases, less expressive ontology representation may be useful and recommended. It is hypothesized, but not all agree, that a continuum can be built across the different ontology representation languages. Agreement amongst communities at the term, natural language definition and synonym level may be a first step towards interoperability, but requires follow up with richer relations. Simple agreement might be expanded to formal axioms and additional constraints as requirements and use cases dictate

Given the complexity of the engineering and manufacturing domains, some domain or application specific ontologies are too constrained (for adaptation or reuse) and foundational ontologies too unconstrained to be used directly. However, areas of commonality can be found so that reference ontologies can be developed that would sit between foundational and domain/application specific ontologies. Such reference ontologies could then be used as the basis for new domain/application specific ontologies and revising existing ones (for greater consistency).

Healthcare and Bioinformatics

Biomedical informatics can be broadly classified into clinical informatics or health information technology (covering healthcare) and bioinformatics (covering biosciences). The discipline of clinical informatics addresses the efficient and accurate use of medical knowledge and information in patient care settings (e.g., Electronic Health Records (EHR), its various interfaces, and decision support). The field of bioinformatics covers topics which deal with computational biology, including all "omics" (genomics, proteomics, metabolomics, etc.). Ontologies, such as those found at the National Center for Biomedical Ontologies (NCBO) and the Open Biological and Biomedical Ontologies (OBO) Foundry, have been used for a number of applications in the bioinformatics field (e.g., patent search described in the virtual panels). However, in the clinical informatics field there has not been much emphasis on the delivery of medical services. Although some attempts, such as the previously mentioned Semantic Data Lake are making considerable progress in certain areas, there are still issues with ontology use by the physicians and associated personnel. In addition, there is activity in some large organizations (e.g., the US Veterans Health Administration) to enrich future EHR systems with clinical ontologies (e.g., the Clinical Care Ontology, the Ontology for General Medical Sciences (OGMS) [1], and SOLOR), as well as proposed extensions to particular healthcare standards such as HL7 that include these. But many of these efforts are internal and not yet externally exposed. Further, there are fundamental problems with the structures of the ontologies in the healthcare domain. When compared side-by-side, these structures are vastly different and at higher levels, similar terms may not match up at all. This makes analyzing the terms and structural similarity of the ontologies very difficult to determine.

[1] https://github.com/OGMS

Earth Sciences

In the Earth Sciences domains, there is a view that the "right" mix of ontologies does not yet exist to achieve interoperability (although some progress has been made with reference ontologies in a small portion of the domain space). The following areas of concern have been pointed out:
- the informal use of ontologies and vocabularies,
- the lack of adequate mappings among concepts, and
- the lack of agreement on well founded integration techniques.

Given the collective magnitude and complexity within the Earth Sciences set of domains and preexisting semantic work, it is not practical to stitch and glue together a very large, all encompassing, master ontology. Instead, some effort is being made to build ODPs in specific areas that can be extended over time and integrated with other ODPs. For example, within a small area of surface and subsurface hydrology, efforts are underway to develop a reference ontology covering previously inconsistent efforts.

Finance and Retail

Common themes in finance and retail involve using and analyzing data to understand customers and to operate a successful business. But, these concerns are more about big data than interoperability. More aligned with interoperability are the requirements for financial institutions to comply with national and international reporting regulations and to manage financial risk. An important standard in the financial space is the Financial Industry Business Ontology (FIBO). It is designed to be a reference ontology providing common concepts and language for financial instruments, processes and legal entities, and was designed with broad input from subject-matter experts to address the multiple needs of banks, financial services firms and regulators. This makes it a logical choice for mapping and integrating financial data.

Less clear is the choice of models and ontologies for retail. In the area of online retail, businesses rely on customers finding their products and services via online search. This, in turn, requires that the search engines "understand" the details of the businesses' products and services. One of the main sources for providing such information to search engines is rich snippets and Schema.org.

In 2012, Schema.org integrated the GoodRelations e-commerce ontology. This begins to address shared semantics and interoperable data in retail. But, GoodRelations has been in existence since 2002, and there is no clear consensus on models or ontologies. There are also modeling solutions and schemas from organizations such as the National Retail Foundation (https://nrf.com) which provide transaction-oriented views of retail data and support data warehousing. NRF's schemas are designed for interoperability, while also supporting enterprise data management. And, regarding use of Schema.org, a 2014 study done by Searchmetrics (http://www.searchmetrics.com/news-and-events/schema-org-in-google-search-results/) revealed that only 0.3% (of the over 50 million domains analyzed) actually use Schema.org on their web pages. So, while the creation of ontologies like GoodRelations has opened the door for more interoperability, there is no clear solution in the retail space.

Designing for Semantic Interoperability

Which Tools and Environments are Necessary?

There was broad consensus within the Summit that better software tools are required to support the integration of concepts or data. Almost all of the methods discussed in the preceding section lack adequate development and maintenance environments. Thus, there is a clear call for new tools and the ongoing improvement of existing tools which support the development and maintenance of semantic data lakes, federated systems and ontology design patterns. Similarly, there is an ongoing need for tools and techniques which facilitate: ontology reuse & modular design, vocabulary harmonization and ontology mapping. In the following sections, we provide a detailed description of additional functionality that will aid in the design process.

Semantic Enablement Layer/Knowledge Infrastructure

More mature conceptual modeling and knowledge engineering tools are required to support the design, development and storage of ontologies and vocabularies. To be effective, these should be integrated into software development and data management tools that provide support across the IT lifecycle and related operations. This view is compatible with the original formulation of the Semantic Web and may be called a knowledge infrastructure vision or *Semantic Enablement Layer*. Such an infrastructure layer would transparently mediate between both existing data and metadata infrastructures and, allow access to reasoning services, Linked Data, and ontology repositories from the Semantic Web as well as the other way around. That is, it would enable Semantic Web applications to access traditional infrastructures. Such a layer needs to be transparent and non-disruptive to ensure that changes to existing and well-standardized infrastructures are not required.

Semantic Brokering

An improvement to agent brokering is also required as part of such a knowledge infrastructure. Agent brokering employs central mechanisms which:
· help resolve disparate vocabularies,
· support data and information distribution requests,
· enforce translatable standards, and
· enable uniformity of search and access in heterogeneous operating environments.

Intelligent mediators/semantic brokering supports better search by using mapped concepts together with a more semantic form of search—reasoning across concepts and domains according to their characteristics, often using inferred concepts.

Among tools employed to support such search and find related resources or to bridge vocabularies are a class of middleware, such as by EuroGEOSS, called semantic brokers. These augment search for resources and go beyond simple "semantic tagging" of resources using vocabularies. Semantic brokers work by expanding textual terms/tagged metadata contained in operations such as "traditional" query using a collection of different controlled vocabularies, thesauri, ontologies together with semantic mapping engines. The goal is that every search should be semantically enabled to the degree appropriate and possible by current tools and domain understanding. Ontologies to facilitate such seamless access and discovery of all data types will require several enhanced semantic technologies, including ontology and vocabulary mapping, and development of ontologies for data/entities at various levels of granularity. Modest improvement in the current state of the semantic technology suite will open up data/metadata to a wider set of options for computer aided manipulation, distribution and long term reuse.

Enhanced Ontology Repositories

To serve their intended purpose, ontologies and ODPs need to be stored, managed, clearly documented and accessible. While improvements are still required, some ontology repositories

have been developed and are available. Some build on earlier experiences with metadata catalogs and registries. Examples include:

- BioPortal,
- OntoHub (GitHub based),
- COLORE (GitHub based),
- Marine Metadata Interoperability Ontology Registry and Repository (MMI ORR) (BioPortal based),
- ESIP Ontology Repository (currently under development & BioPortal based), and
- Center for Expanded Data Notation (CEDAR).

GitHub is playing a major role in collaboration and serves as a repository for some ontologies (as we are seeing with COLORE and OntoHub).

Ontologies and Communities

There are many socio-technical issues in the use of ontologies for semantic applications, including how to:
- facilitate communities to contribute and discuss use cases and requirements which drive the development of a shared ontology,
- develop a governance process to facilitate and coordinate efforts regarding the design, evaluation and validation of ontologies including both the natural text definitions and axioms,
- provide quality assurance, evaluation and validation processes and get wide agreement of validity. (If an ontology is created to provide functionality x; who verifies it actually does x?),
- establish provenance, security, and administration policies,
- provide ongoing maintenance as required since evolution is expected,
- develop bridging disciplines, a key challenge, especially with the earth sciences and healthcare domains, and
- engage communities in developing and sharing materials.

Over the course of the 2016 Summit, a variety of community models were considered. Several distinct and successful community models have been identified in the course of discussions in this Summit.
1. The BFO/OBO Foundry ontology ecosystem is an example of a successful approach. Its mission is to "develop a family of interoperable ontologies that are both logically well-formed and scientifically accurate".
2. Both FIBO and Allotrope Foundation are examples of standards formed by multiple companies within one industry. Their ontology creation efforts were driven by common goals. Formed in 2012, the Allotrope Foundation has brought together 12 pharmaceutical and biotech companies which are working toward the goal of building shared semantic tools which address common data acquisition, archiving and management problems within analytic laboratory environments.

3. The European Skills, Competences, Qualifications and Occupations (ESCO) project has low semantic complexity, but is based on a strong and consistent model. It provides a multi-country, multi-language reference model with targeted goals addressing requirements in the HR domain.
4. The VIVO semantic web-application was created as a solution driven by a common set of requirements that are not industry specific. A diverse set of organizations, including federal agencies, academic institutions, professional societies and members of the semantic web and ontology communities contributed to the creation and build-out of the VIVO open-source semantic web application and associated tool sets.
5. To meet specific business requirements, some organizations adopt an enterprise-wide, centrally managed vocabulary. The centralization of the management and governance of vocabularies is viewed as critical to ensure harmonization, coordinated development, quality and continuous improvement. This internal governance model may be especially important in highly regulated environments and for high risk applications, as discussed in the application of an industry standard (FIBO) to a specific corporation. Another application is to enable corporate best practices or mandates such as branding.

Lessons Learned

Approach to Complicated Domains

Semantic heterogeneity is introduced when a problem-space spans diverse content and communities, such as those represented by the EarthCube effort. For a given domain, improved SI can be achieved with the substantial study of interconnected use cases which expose requirements that address data, models, and tools. The stated use cases and requirements should have clear implications for data interoperability, ontology, and semantic infrastructure.

A solution strategy is to provide methods that enable users to flexibly load and combine different ontologies instead of hardwiring data to particular ontologies and, thus, hinder their flexible reusability. This would include work from modular building blocks with microtheories of locally valid semantics. It is easier to manage multiple, small, internally consistent ontologies and focus on interrelations as needed for inter-operation.

A blended bottom-up, top-down approach is recommended. Hybrid SME/ontology engineer teams are needed to engineer this properly. That is, projects must be structured so domain experts are active participants in building semantic models with help from ontological engineers. Work should proceed from use cases through conceptualization to validating final products supported by good documentation.

Pilot Projects

The deployment of SI techniques as part of pilot projects can kick-start elements of community participation including: conversations, engagement and education with respect to the use of

semantic technologies and related ontologies. Early successes drive interest, motivation and adoption.

For short-term proof of concept projects, teams should emphasize lightweight approaches that are opportunistic efforts which leverage existing work but are not tightly bound to it. Focus on requirements which address low hanging fruit and leverage existing vocabularies and conceptual models/ontologies to ensure that a semantics-driven infrastructure is available for use in early stages of work. This helps reduce entry barrier for domain experts/SMEs and motivates them to contribute. These efforts can be incremental, starting with domain vocabularies from which richer schema can be developed with formal semantics.

On Complexity and Learnability

Clear, well-defined terminology is an essential component of efficient and useful communication between individuals in a community of interest. When formalized, these sets of terminologies are sometimes called controlled vocabularies or thesauri. They can provide a base level of semantic interoperability. When use cases and requirements call for more complete interoperability, formal ontologies offer a systematic method for specifying and automating the use of controlled vocabularies, their definitions and their synonyms.
The creators of ontologies should provide adequate documentation, examples and training material. The reuse, adoption and integration of ontologies (and data) requires good and easily comprehensible documentation and examples. For example, in the financial domain, FIBO and GoodRelations are "not intuitive" but require learning and experimentation. Documentation is available, but it is sometimes difficult to understand what to use, where to link and what to extend.

Evaluation

Ontologies should be evaluated against their requirements and within the context of their intended use. Several summit presentations addressed evaluation directly. As part of his presentation, Lalit Patil introduced an interesting discussion on markers for successful Semantic Interoperability within digital ecosystems. He posited that systems that have successfully met SI challenges will:
· require fewer user decisions,
· require less contextual intelligent information, and
· exhibit increased structured integration with other domains.

The detailed description of his reasoning can be found in his presentation.

The NCOIC's Systems, Capabilities, Operations, Programs and Enterprises (SCOPE) model offers a structured approach to identifying and cataloging the capabilities and top-level requirements needed to establish an effective solution. SCOPE is designed to characterize interoperability relevant aspects and capabilities of systems.

A Roadmap for the Future

A Vision for Semantically-Aware Ecosystems

In practice, interoperability is difficult to achieve, semantic interoperability even more so. Applications across domains utilize information in different ways, and the knowledge/ontology conceptualizations and representation formalisms inherent in these applications (either implicitly or explicitly) may also be different. There is general agreement in the domain communities (addressed during the Summit) that there is a lack of appropriately-sized, ontological building blocks and that ODPs are underutilized. Efforts are required to extend the usefulness and adoption of ODPs, including the definition of integration patterns. Modular, incremental development approaches are needed which utilize these ontological building blocks and ODPs, with the end goal of creating well-designed reference and bridge ontologies based on them.

Building a semantic infrastructure should be both community-driven and vision-based. The infrastructure must be designed for long-term use and evolution. Part of the vision is to converge on the reference and bridge ontologies mentioned above. This would enable interoperability of cross-domain systems and their data in an open, transparent and inclusive manner. Formal semantics in the ontologies would promote integration and flexibility by connecting the several layers of data and information management, from the resource layer with access to data and information, to the data curation and management layer. However, new approaches and technologies are required to make this happen, and/or we need to combine existing tools and solutions in different ways.

To achieve this vision, we recommend ongoing education (about semantic interoperability), expanded evaluation efforts, improved design and engineering practices, along with continued innovation.

Communication Strategy

To address the interdisciplinary nature of building, extending and improving SI Ecosystems, it is suggested that the Ontolog Community and related organizations look to communicate outward to a diverse set of audiences. We need to understand and leverage the drivers of semantic technology and communicate the value proposition of semantic technologies using non-technical language understood by the various domains including the IT profession. One tool to help accomplish this goal is set of common communication artifacts comprised of messaging and documentation which addresses recurrent and prevalent bottlenecks.

We propose the creation of artifacts and documentation to address the following:
- characterization of use cases and requirements (potentially build on SCOPE model),
- characterization of semantic models (what functionality do they offer consuming applications/ecosystem participants),
- characterization of semantic interoperability ecosystems as a whole,
- measures of appropriate SI, and
- conditions and requirements for successful SI projects.

A proposed goal is to have a set of artifacts for review and discussion within a 1-year timeframe.

The collective understanding of the elements which comprise Semantic Interoperability Ecosystems needs to be extended. Over the course of the past 3 summits, we touched on the various elements at play, but there often are not readily available and clear descriptions of the relevant methodologies and software artifacts. By detailing these elements and their points of intersection, we can gain a greater appreciation of the semantic problems that arise from varying configurations of digital ecosystems. Additionally, this material and the insights gleaned, can be leveraged to refine existing standards for ontologies and related SI efforts and to create new standards where required.

Within this summit we saw several promising ecosystem approaches including but not limited to:
- BFO,
- SIMF,
- Use of grounded vocabularies derived from FIBO,
- FIHR effort,
- Semantic Data Lake approach, and
- SCOPE.

Testbeds for Semantic Interoperability

Establish an interoperability test-bed. A testing facility would allow real-world scenarios and facilitate the testing of various ontological approaches to semantic interoperability. Additionally, the test-bed will have representative datasets from various domains, for example in the healthcare domain this could be simulated clinical notes, billing, MRIs, and labs. This could provide a much needed benchmark for evaluating different approaches.

Educational Workshops

Domain communities need better understanding of the role and value of semantics as well as opportunity to practice developing simple models. There is a lack of clarity regarding what can be accomplished with current semantic technologies and what the limitations are. Additional material describing how to incorporate current semantic technologies into common software development lifecycles is required. It is not enough, for example, to slavishly convert current metadata into RDF or OWL without suitable conceptual analysis. Likewise, it is often wrong to quickly pick terms and/or relations for reuse from an existing ontology for a new purpose. In these situations, additional conceptualization, axiom revision and bridging may be required. Educational workshops could address these concerns. They also could provide hands-on practice reusing existing ontologies and ODPs.

Funding and Ongoing Support

There is little incentive for coordinated development. Building out semantic architectures, ontologies and Infrastructures is foundational work. Such architectures are costly to build and maintain. Mapping semantics to existing infrastructures also requires sustained support. These efforts must be funded, governed and supported.

Community participation by scientists and organizations such as data centers must be motivated to contribute data and expertise to semantic-focused efforts. The recent establishment of the CEDAR project and success of the VIVO project are prime examples.

Given that the benefits of semantic-based frameworks are often realized incrementally, as a community of practitioners, we recommend continued support, especially over the course of the development lifecycle. The source of this funding is an open question and is determined by a number of factors. In the community section, we have provided several successful efforts.

In Closing

Semantic heterogeneity poses a major challenge across a wide range of domains, including healthcare, earth sciences, engineering, manufacturing, finance and retail. Throughout Ontology Summit 2016, solutions were presented which addressed how ontologies (in forms which cut across the ontology spectrum) are being used to target the semantic heterogeneity problem. In particular, specific ontologies (such as upper, reference, and bridge ontologies) are being developed to support mediation between the terminologies of multiple systems. At the same time, complementary methods which facilitate semantic interoperability are being developed or enhanced; these methods range from ontology mapping and vocabulary harmonization to the design of modular reusable ontologies and ontology design patterns. Related architecture developments include the semantic data lake. Even with recent progress, major obstacles still remain -- interoperability is still difficult to achieve.

There was broad consensus within the Summit that improved software tools and environments are required to support the integration of concepts and data. We discussed the variety of socio-technical issues that hinder the use of ontologies for supporting semantic interoperability. To address these challenges, several proposals were identified in a roadmap for the future, including testbeds for semantic interoperability, enhanced communication efforts, educational workshops, and ongoing support for the coordinated development of ontologies and ontology mappings.

References

https://financialresearch.gov/frac/files/FRAC_DTS_Instrument_Database_Recommendation.pdf

http://www.cms.gov

https://en.wikipedia.org/wiki/Semantics

https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.html

https://sweet.jpl.nasa.gov/

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839690/

http://si.washington.edu/projects/fma

http://www.edmcouncil.org/financialbusiness

http://www.omg.org/hot-topics/finance.htm

http://ontogenesis.knowledgeblog.org/295

http://schema.org/docs/about.html

http://metadatacenter.org/

http://www.omgwiki.org/architecture-ecosystem/doku.php?id=semantic_information_modeling_for_federation_rfp

https://s3.amazonaws.com/ontologforum/OntologySummit2016/2016-03-10_Healthcare/The-Semantic-Data-Lake-in-Healthscience--ParsaMirhaji.pdf

http://www.commonwl.org/draft-3/SchemaSalad.html

http://www.omgwiki.org/architecture-ecosystem/lib/exe/fetch.php?media=semantic_information_modeling_for_federation_overview.pdf

http://ontolog.cim3.net/OntologySummit/2014/communique.html

http://wiki.goodrelations-vocabulary.org/GoodRelations_and_schema.org

http://www.edmcouncil.org/semanticsrepository/index.html

http://linkeddata.org/

http://www.opengeospatial.org/standards/om

https://www.w3.org/2005/Incubator/ssn/ssnx/ssn

http://www.omg.org/spec/DOL/1.0/Beta1/index.htm

http://ceur-ws.org/Vol-1081/womo2013_invited_paper_1.pdf

https://global.oup.com/academic/product/a-pattern-language-9780195019193?cc=us&lang=en&

https://nrf.com/

http://www.searchmetrics.com/news-and-events/schema-org-in-google-search-results/

http://www.obofoundry.org/

http://www.allotrope.org/#!organization/c20k4

https://ec.europa.eu/esco/portal/escopedia/European_Skills%252C_Competences%252C_Qualifications_and_Occupations_%2528ESCO%2529

http://vivoweb.org/info/about-vivo

http://ontologforum.org/index.php/ConferenceCall_2016_04_21

http://ontologforum.org/index.php/ConferenceCall_2016_02_18

https://www.ncoic.org/technology/technical-products/scope-workshops

Arp, Robert; Smith, Barry; Spear, Andrew D. (2015-08-28), Building Ontologies with Basic Formal Ontology, The MIT Press, Kindle Edition.