Ontology Summit 2016: Healthcare Track Synthesis
April 14, 2016

Co-Champions:
Ram D. Sriram (NIST)
Leo Obrst (MITRE)
Levels of Biological Information

HEALTH CARE

- Ecologies
- Societies/Populations
- Individuals
- Organs
- Tissues
- Cells

BIOSCIENCES

- Protein and gene networks
- Protein interaction networks
- Protein
- mRNA
- DNA
Health IT Interoperability
EHRs: Key Issues

- Input (user interfaces)
- Store (representation and persistency)
- Manipulate (search, mining, knowledge creation)
- Exchange (syntactic and semantic interoperability)
Interoperability

Electronic Health Record (EHR) System

Electronic Health Record (EHR) System

Laboratory Information systems (LIS)

Public Health
Interoperability

Electronic Health Record (EHR) System

Pharmacy

Medical Device

Electronic Health Record (EHR) System

Medical Device

Medical Device

Medical Device
### Semantic Mapping Techniques
The Medical “Tower of Babel”

Some terms for “Hypersomatotrophic Gigantism”:

<table>
<thead>
<tr>
<th>Vocabulary</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>UMLS Metathesaurus</td>
<td>Hypersomatotrophic Gigantism</td>
</tr>
<tr>
<td>ICD-9-CM</td>
<td>No direct translation</td>
</tr>
<tr>
<td>MeSH</td>
<td>No direct translation</td>
</tr>
<tr>
<td>DXplain</td>
<td>Pituitary Gigantism</td>
</tr>
<tr>
<td>Read Codes</td>
<td>No direct translation</td>
</tr>
<tr>
<td>SNOMED</td>
<td>Hypersomatotrophic Gigantism</td>
</tr>
</tbody>
</table>
True EHR Interoperability

Syntax  Context  Semantics  Semantics  Context  Syntax

Conformance to raw message specifications in the Standard

Conformance to Specific Clinically Relevant Test data to exercise all options

Conformance to Specified Vocabularies in clinically relevant test data
Interoperability

- **Standards** are essential to achieving conformance and interoperability

- Implementation guidelines are needed

- **Rigorous** testing is critical to achieving **conformance** and enabling **interoperability**

- Conformance CAN NOT be definitively determined\(^*\) - but gives a level of confidence based on quality and quantity of test(s) performed

  \(^*\)unless specification is very basic

- Conformance ≠ Interoperability

  – A is Conformant, B is Conformant
  * The above does not say anything about interoperability between A and B
Standards That Link to an EHR
Standards in HealthCare

• Terminology
  – Systematized Nomenclature for Medicine – Clinical Terms (SNOMED-CT), Logical Observation Identifiers Names and Codes (LOINC), RxNorm (normalized names for clinical drugs)

• Classification Systems

• Devices
  – IEEE 11073

• EHR-Related
  – Digital Imaging and Communications in Medicine (DICOM), Health Level Seven (HL7) Clinical Document Architecture (CDA)

• Interoperability
  – DICOM, HL7 Messaging, Health Insurance Portability and Accountability Act (HIPAA) Transactions, National Council for Prescription Drug Programs

• Language Formats
  – XML, X12

• Internet Protocols
  – HTTP/HTTPS ⊆ TCP/IP

• De Facto Ontology Standard: Open Biological & Biomedical Ontologies (OBO)
Biomedical Ontologies: Categories of Use

• Knowledge Management
  – Indexing and retrieval of data and information, access to information, mapping among ontologies

• Data Integration
  – Exchange and semantic interoperability

• Decision Support and Reasoning
  – Data selection and aggregation, decision support, natural language processing, knowledge discovery
Semantic Interoperability/Integration Definition

- **To interoperate** is to participate in a *common purpose*
  - operation sets the context
  - purpose is the intention, the end to which activity is directed
- **Semantics is fundamentally interpretation**
  - within a particular context
  - from a particular point of view
- **Semantic Interoperability/Integration is fundamentally driven by communication of purpose**
  - participants determined by interpreting capacity to meet operational objectives
  - service obligations and responsibilities explicitly contracted
Towards Semantic Interoperability & Integration

• Semantic Integration is enabled through:
  – Establishing base semantic representation via ontologies (class level) and their knowledge bases (instance level)
  – Defining semantic mappings & transformations among ontologies (and treating these mappings as individual theories just like ontologies)
  – Defining algorithms that can determine semantic similarity and employing their output in a semantic mapping facility that uses ontologies

• The use of ontologies & semantic mapping software can reduce the loss of semantics (meaning) in information exchange among heterogeneous applications
What is Needed

• Multiple contexts, views, application & user perspectives
• Multiple levels of precision, specification, definiteness required
• Multiple levels of semantic model verisimilitude, fidelity, granularity, dynamicity
• Multiple kinds of semantic mappings, transformations needed:
  – Entities, Relations, Properties, Ontologies, Model Modules, Namespaces, Meta-Levels, Facets (i.e., properties of properties), Units of Measure, Conversions, Theories, Interpretations, etc.
• Upper/Foundational, Mid-level, and Utility Ontologies are important to be able to interrelate domain ontologies.
Healthcare Track: **Goal and Mission**

- **Healthcare and biomedical domains:** a very large ecosystem of vocabularies and ontologies
- **Goal:** To provide insight into the nature of semantic interoperability issues and approaches in the Healthcare ecosystem
- **Mission:** Present issues of semantic interoperability and integration in domain of healthcare
  - Discuss approaches in semantic interoperability and integration in healthcare: precoordination, postcoordination, mapping, anchoring, foundational ontologies and architectures, hybrid approaches
  - Discuss approaches for achieving semantic interoperability of vocabularies and ontologies in healthcare
  - Discuss gaps in current approaches in semantic interoperability in healthcare
  - Discuss current/future challenges and prospects in semantic interoperability in healthcare
  - Propose best methods for achieving semantic interoperability in healthcare
Healthcare Track: March 10

• Kincho Law (Department of Civil and Environmental Engineering, Stanford University, USA):
  An Ontology-Based Approach for Facilitating Information Retrieval from Disparate Sources: Patent System as an Exemplar

• Olivier Bodenreider (US National Library of Medicine, USA):
  Ontologies Support Semantic Interoperability in Healthcare

• Eric Little (OSTHUS, Inc. USA):
  Integrating the Laboratory for Improved Analytics: How Semantics Is Driving an Evolution of 21st Century Labs

• Parsa Mirhaji (Clinical Research Informatics, Montefiore Medical Center, Albert Einstein College of Medicine, USA):
  The Semantic Data Lake in Healthcare
Healthcare Track: April 7, 2016

• Keith Campbell (US Veterans Health Administration (VHA) Knowledge-Based Systems, USA):
  – SOLOR: Building a coherent foundation for clinical data by leveraging the SNOMED/LOINC Cooperative Agreement, and integrating RxNorm

• Barry Smith (U-Buffalo, NY, National Center for Ontological Research, USA):
  – An Ontology Ecosystem Approach to Electronic Health Record Interoperability

• Amit Sheth (Kno.e.sis, Wright State University, USA):
  – Ontology-enabled Healthcare Applications exploiting Physical-Cyber-Social Big Data

• Leo Obrst (MITRE, USA)
  – Semantic Interoperability in General and in Healthcare: Clinical Care
Key Ideas

• Retrieving Information from Disparate Sources (Kincho Law)
• Facilitating Semantic Interoperability (Leo Obrst, Olivier Bodenreider)
• Aiding in Medical Decision Making and Data Analytics (Parsa Mirhaji, Olivier Bodenreider)
• Integrating Laboratory Data From Multiple Sources (Eric Little)
• Integrating and Post Coordinating Standardized Nomenclature (Keith Campbell)
• Building Ontologies with Basic Formal Ontology (Barry Smith)
• Developing Smart Healthcare Applications (Amith Sheth)
Brief Summary of Talks

- Need ontologies from multiple sources for any practical application
- Current ontologies, such as SNOMED, have inconsistency issues
- Proposed methodologies like FHIR show some promise, but need further work
- Methodologies for ontology development were presented
- Several useful use cases presented
- Electronic Health Record (EHR): coherent, semantic representation sorely needed
BACK UP SLIDES
Retrieving Information Disparate Sources (Law)
Patent Validity and Infringement/Enforcement Questions involves analysis of documents in various domains – Patents, USPTO File Wrappers, Court Documents, Scientific/Technical Publications, and Technical Product Literature

- Owned by disparate public (government) and private sectors
- The information is often available online, but siloed into several diverse information sources
- Today, the analysis is done manually and poorly by companies offering various patent research and strategy services
Summary: BIO-REGNET

Knowledge Source:
Bio Ontology
(Technical Domain)

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

Bioportal
(bioportal.bioontology.org)
NCI Thesaurus

Issued
Patents and
Applications

Court Cases

File
Wrappers

Technical
Publications

Regulations
and Laws

Siloed Patent System Information

Integration
Summary and Discussion

- IP informatics: from research/development, patent filings to infringement and IP protection
- Knowledge-Driven Ontology-Based Approach
  - Technological ontologies
  - Patent system ontology
- Generalization – Linking to other information sources – technical/scientific publications, product literature
- User Interface – Efficient presentation of relevant (semantic) information
- Comparative analysis of documents
- Scalability (Graph Database?)
- Experiment with more use cases in other technical domains outside of the biomedical domain
Facilitating Semantic Interoperability (Obrst, Bodenreider)
Methodology for Clinical Use (Obrst)

- Analyze current ontologies and terminologies
- Develop clinical use cases
- Create or adapt a Clinical Care Ontology
- Evaluate Clinical Care Ontology using clinical use cases
- Evaluate applicability of Semantic Web technology in context of clinical use cases

Goal: support patient-centered, care plan-centered visualization and reasoning
ARRA & The Meaningful Use Program (CMS)

- The American Reinvestment & Recovery Act (ARRA) was enacted on February 17, 2009. ARRA includes many measures to modernize our nation's infrastructure, one of which is the "Health Information Technology for Economic and Clinical Health (HITECH) Act". The HITECH Act supports the concept of electronic health records - meaningful use [EHR-MU], an effort led by Centers for Medicare & Medicaid Services (CMS) and the Office of the National Coordinator for Health IT (ONC). HITECH proposes the meaningful use of interoperable electronic health records throughout the United States health care delivery system as a critical national goal.

- Meaningful Use is defined by the use of certified EHR technology in a meaningful manner (for example electronic prescribing); ensuring that the certified EHR technology is connected in a manner that provides for the electronic exchange of health information to improve the quality of care; and that in using certified EHR technology the provider must submit to the Secretary of Health & Human Services (HHS) information on quality of care and other measures.
Proposed Stages of Meaningful Use

Stage 1
- Electronically capturing health information in a coded format
- Using that information to track key clinical conditions
- Communicating that information for care coordination purposes
- Initiating the reporting of clinical quality measures and public health information

Stage 2
- Stage 1 objectives
- Disease management
- Clinical decision support
- Medication management
- Support for patient access to their health information
- Quality measurement and research
- Bi-directional communication with public health agencies

Stage 3
- Stage 1 and 2 objectives
- Improvement in quality, safety and efficiency
- Decision support for national high priority conditions
- Access to self management tools
- Access to comprehensive patient data and improving population health outcomes
Reference value sets

- $\text{MedicationVTEProphylaxis} = \text{Union of:}$
  - "Medication, Administered: Low Dose Unfractionated Heparin for VTE Prophylaxis"
  - "Medication, Administered: Low Molecular Weight Heparin for VTE Prophylaxis"
  - "Medication, Administered: Injectable Factor Xa Inhibitor for VTE Prophylaxis"
  - "Medication, Administered: Warfarin"

Value Set Information
- Name: Low Dose Unfractionated Heparin for VTE Prophylaxis
- OID: 2.16.840.1.113762.1.4.1045.39
- Code System: RXNORM

https://vsac.nlm.nih.gov/
Aiding in Medical Decision Making and Data Analytics (Mirhaji, Bodenreider)
Ontologies provide a standard vocabulary for biomedical entities, helping standardize and integrate data sources. A system for drug allergies must be able to resolve drug names into standard codes and map between drug coding systems and the allergy knowledge base.

Ontologies are a source of computable domain knowledge that can be exploited for decision support purposes, often in combination with business rules. In an alert system for drug allergies, allergy to betalactams can be represented efficiently if the system can access a classification of drugs (as opposed to direct links to specific drugs).

Besides the clinical decision support, ontologies can also be used to support reasoning in applications. The Foundational Model of Anatomy was used in one application to predict the consequences of a penetrating injury.
Analytics (Bodenreider)

• Clinical data warehouses
  – Distinct from EHR systems

• “ETL” (extract – transform – load) processes
  – Data normalized to “standards”
  – Local data mapped to ontologies
    • Facilitated by ontology integrations systems (e.g., UMLS)

• Analysis leverages hierarchical and other relations
  – Transitive closures

• Facilitates analysis of large-scale data repositories
  – Including distributed repositories across institutions

See http://www.ohdsi.org
Semantic Data Lake (Mirhaji)

• Platform for Big Data Analytics and Cognitive Computing
• Based on principles of HDFS/Spark/GraphX, and Semantic Web
• Knowledge Based (metadata, terminology, ontologies)
  - Enables linking Data -> Knowledge -> Linked Open Data
    • URI level security, authorization, provenance
    • Optimized for integration of heterogeneous health and well-ness data
  - Structured and non-structured data
    • Sophisticated temporal and geospatial analytics, Machine Learning
    • To support basic science, translational research, operational care coordination, population health management (ACO)
• A Montefiore, Franz, Intel, Cisco Collaboration
The Semantic Data Lake

- Registries
- Financial
- Free Text
- PRO
- Devices
- Bio Bank
- EMR
- LIMS
- HL7 Feed
- Web Services
- Legacy
- Omics
- CTMS
- Claims
- Annotation Engine

SDL Loader

- ML-Lib
- SPARQL
- Prolog
- R
- Java API

Allegro Graph

Spark

Hadoop

HDFS

Spark

Hadoop

HDFS

Spark

Hadoop

HDFS

Spark

Hadoop

HDFS
SDL Knowledgebase
(UMLS Vocabularies and Semantic Network Linked to OMOP Terminology)
Ongoing projects supported by SDL at Montefiore

Personalized Medicine

Predictive Modeling

Translational Research

Fraud Detection

Risk Assessment

Business Intelligence

Public Health

Mobile Health

Decision Support
Integrating Laboratory Data From Multiple Sources (Little)
Integrating the Laboratory for Improved Analytics

• Discuss the current status of laboratories – data is siloed, machines are not integrated, numerous file types exist, and information is difficult to retrieve and utilize over time with any consistency.

• Show how utilizing semantics can provide researchers with improved ways to query and understand their data from multiple perspectives by providing a common reference layer – show specific examples of how the Allotrope Foundation is solving this problem.

• Discuss how semantic technologies such as those developed under Allotrope can be extended to provide advanced analytics capabilities that bring together both logical and mathematical approaches to graph theory, knowledge discovery and big data analytics.

• Briefly discuss how this can impact the entire Life Sciences industry and drive real business value to a wide variety of customers.
The Current Lab Situation

Many challenges exist for data to be captured, integrated and shared

- Data Silos
- Incompatible instruments and software systems
- Legacy architectures are brittle and rigid
- SME knowledge resides in people’s heads
- Data schemas are not explicitly understood
- Lack of common vision between business units and scientists
Allotrope Foundation Taxonomies (AFT)
Applying the Semantic Spectrum

Allotrope Focus

Code (Lists)  Terms (Instrument, drug, etc.)

Taxonomy (Hierarchy)  Controlled Vocabulary (Agreed Upon Terms)

Thesaurus (Preferred Labels, Synonyms, etc.)  RDF Models (Triples as Graphs)

Reasoning (Rule-based Logics: Discover New Patterns)

Allotrope can be extended to include analytics and automated reasoning
Integrating and Post Coordinating Standardized Nomenclature (Campbell)
• Use description logic to integrate SNOMED, LOINC and RxNorm
• SNOMED/LOINC co-operative agreement is already achieving the integration of SNOMED and LOINC
• RxNORM raw data is mapped into OWL EL and then integrated into above.
• SNOMED + LOINC + RxNorm + post-coordination provides comprehensive coverage for typical clinical data representation requirements
Postcoordination in SNOMED

• The meaning of a *postcoordinated expression* is defined by the way the *concepts* are combined in the *expression*, and the defining *relationships* of the referenced *concepts*. [From Snomed Manual]

• Postcoordination can be done by Refinement, Qualification, and Composition
Ontology Ecosystem Approach to EHR Interoperability

• Even a total victory for EPIC (which is becoming the standard EHR in many organizations) will not solve the interoperability problem.

• MU Stage 2 requires use of SNOMED CT for value lists (see Olivier’s talk)
  – Human coding with SNOMED is unreliable and inconsistent
  – e.g. **Solitary leiomyoma (disorder)** has Concept ID: 254769006 and **Leiomyoma (morphologic abnormality)** has Concept ID: 4459800. Are these same or different?

• FHIR (Fast Healthcare Interoperability Resources) is proposed as a solution and still people are trying to figure this out
Personalized medicine needs large cohorts with rich phenotypic data conforming to common standards. How to get there?

THREE CHOICES

1. Everyone uses Epic
2. Government enforces common standards
3. Let’s start again from scratch, using the same approach we should have used from the beginning: rigorous testing-based development of EHRs by leading medical research institutions until we see what technologies will work
Building Ontologies with Basic Formal Ontology

By Robert Arp, Barry Smith and Andrew D. Spear

Overview

In the era of “big data,” science is increasingly information driven, and the potential for computers to store, manage, and integrate massive amounts of data has given rise to such new disciplinary fields as biomedical informatics. Applied ontology offers a strategy for the organization of scientific information in computer-tractable form, drawing on concepts not only from computer and information science but also from linguistics, logic, and philosophy. This book provides an introduction to the field of applied ontology that is of particular relevance to biomedicine, covering theoretical components of ontologies, best practices for ontology design, and examples of biomedical ontologies in use.

After defining an ontology as a representation of the types of entities in a given domain, the book distinguishes between different kinds of ontologies and taxonomies, and shows how applied ontology for biomedicine is a natural development of formal ontology. It presents the core features of the Basic Formal Ontology (BFO)
Developing Smart Healthcare Applications (Sheth)
Ontology-enabled Healthcare Applications Exploiting Cyber-Physical-Social Data

• Discussed several projects at Kno.e.sis (See http://knoesis.org/amit/commercialization/index.php?page=2)

• Ontologies developed at Kno.e.sis
  – Drug Abuse Ontology
  – Depression Insight Ontology
  – Healthcare Ontology
  – Human Performance and Cognition Ontology (HPCO)
  – Ontology for Parasite Lifecycle (BioPortal)
  – Parasite Experiment Ontology (BioPortal)
  – Provenir Ontology (Provenir)
I was sent home with 5 x 2 mg Suboxones. I also got a bunch of phenobarbital (I took all 180 mg and it didn't do shit except make me a walking zombie for 2 days). I waited 24 hours after my last 2 mg dose of Suboxone and tried injecting 4 mg of the bupe. It gave me a bad headache, for hours, and I almost vomited. I could feel the bupe working but overall the experience sucked.

Of course, junkie that I am, I decided to repeat the experiment. Today, after waiting 48 hours after my last bunk 4 mg injection, I injected 2 mg. There wasn't really any rush to speak of, but after 5 minutes I started to feel pretty damn good. So I injected another 1 mg. That was about half an hour ago. I feel great now.
Smarter Data Generated with Ontologies

Ontology

Lexicon

Lexico-ontology

Rule-based Grammar

ENTITIES

TRIPLES

EMOTION

INTENSITY

PRONOUN

SENTIMENT

DRUG-FORM

ROUTE

OF ADM

SIDE EFFECT

DOSAGE

FREQUENCY

INTERVAL

Suboxone, Kratom, Herion, Suboxone-CAUSE-Cephalalgia

disgusted, amazed, irritated more than, a, few of I, me, mine, my Im glad, turn out bad, weird

ointment, tablet, pill, film

smoke, inject, snort, sniff

Itching, blisters, flushing, shaking hands, difficulty breathing

DOSAGE: <AMT><UNIT>
(e.g. 5mg, 2-3 tabs)

FREQ:
<AMT><FREQ_IND><PERIOD>
(e.g. 5 times a week)

INTERVAL:
<PERIOD_IND><PERIOD>
(e.g. several years)