The Translational Medicine Ontology and Knowledge Base: Driving personalized medicine by bridging the gap between bench and bedside

Joanne S. Luciano*[a,b], Bosse Andersson[c], Colin Batchelor[d], Olivier Bodenreider[e], Tim Clark[f], Christopher Domarew[g], Thomas Gambet[h], Lee Harland[i], Anja Jentzsch[j], Vipul Kashyap[k], Peter Kos[f], Julia Kozlovsky[l], M. Scott Marshall[m,n], James P. McCusker[a], Deborah L. McGuinness[a], James McGurk[o], Chimezie Ogbuji[p], Elgar Pichler[q], Robert L. Powers[b], Eric Prud'hommeaux[h], Matthias Samwald[r,s], Lynn Schriml[t], Peter J. Tonellato[f], Patricia L. Whetzel[u], Jun Zhao[v], Michel Dumontier[w]

[a]Rensselaer Polytechnic Institute, Troy, NY. [b]Predictive Medicine Inc., Belmont, MA, USA. [c]AstraZeneca, Lund, Sweden. [d]Royal Society of Chemistry, Cambridge, UK. [e]National Library of Medicine, Bethesda, MD, USA. [f]Harvard Medical School, Boston, MA, USA. [g]Albany Medical Center, Albany, NY. [h]W3C, Cambridge, MA, USA. [i]Pfizer, Sandwich, UK. [j]Freie Universitat, Berlin, Germany. [k]Cigna, Hartford, CT, USA. [l]AstraZeneca, Waltham, MA, USA. [m]Leiden University Medical Center, Leiden, NL. [n]University of Amsterdam, Amsterdam, NL. [o]Daiichi Sankyo, NJ, USA. [p]Cleveland Clinic, Cleveland, OH, USA. [q]W3C HCLSIG. W3C, Cambridge, MA, USA. [r]Digital Enterprise Research Institute, Galway, Ireland. [s]Information Retrieval Facility, Vienna, Austria. [t]University of Maryland, Institute for Genome Sciences. [u]Stanford University, Stanford, CA, USA. [v]University of Oxford, Oxford, UK. [w]Carleton University, Ottawa, Canada.

Email: jluciano@cs.rpi.edu;

*Corresponding author

Abstract

Translational medicine requires the integration of knowledge using heterogeneous data from health care to the life sciences. Here, we describe a collaborative effort to produce a prototype Translational Medicine Knowledge Base (TMKB) capable of answering questions relating to clinical practice and pharmaceutical drug discovery. A key component of this work is the Translational Medicine Ontology (TMO) which provides a foundation upon which chemical, genomic and proteomic data may be linked to disease, treatments and electronic health records. We demonstrate the use of semantic web technologies in the integration of patient to biomedical data, and how such a knowledge base can aid physicians in providing tailored patient care and facilitate the recruitment of non-responsive patients into active clinical trials. Thus, patients, physicians and researchers may explore the knowledge base to better understand therapeutic options, efficacy and mechanisms of action. The TMKB takes

us a step forward in using semantic web technologies to facilitate integration of relevant external sources and we expect this work to form the basis for future work towards the development of a computational platform that supports personalized medicine.

Introduction

A major element of personalized medicine involves the identification of therapeutic regimes that are safe and effective for specific patients. Personalized medicine aims to focus therapy on an individual or group of individuals with similar characteristics. This is in contrast to the well-known concept "blockbuster" drugs, with the concept of targeted patient groups in-between [1]. The current decline in emphasis on blockbuster therapeutics corresponds to the recent quest for tailored therapeutics. Essential to the realization of personalized medicine is the development of information systems capable of providing accurate and timely information about potentially complex interrelationships between individual patients, drugs and tailored therapeutic options.

The demands of personalized medicine include integrating knowledge across data repositories that have been developed for divergent uses, and do not normally adhere to a unified schema. This paper demonstrates the integration of such knowledge across multiple heterogeneous datasets. We show the formation of queries that span these datasets, connecting the information required to support the goal of personalized medicine from both the research and the clinical perspectives.

Integration of the patient electronic health record (EHR) with publicly accessible information creates new opportunities and challenges for clinical research and patient care. Care must be taken that the information complexity available in the clinic does not impair the clinician's ability to accurately and rapidly prescribe drugs that are safe and effective for a specific patient, and are also covered by the patient's insurance provider. On the other hand, EHRs enable the identification of adverse events and outbreak awareness and provide a rich set of longitudinal data, from which researchers and clinicians can study disease, co-morbidity and treatment outcome. Moreover, the increased desire to rapidly translate drug and gene-based drug therapy to clinical practice depends on the comprehensive integration of the entire breadth of patient data to facilitate and evaluate drug development (Woolf 2008). Thus, EHR integration could create the ideal conditions under which new or up-to-date evidence-based guidelines for

disease diagnosis and treatment can emerge. Although supplying patient data to the scientific community presents both technical and social challenges [2], a comprehensive system that maintains individual privacy but provides a platform for the analysis of the full extent of patient data is vital for personalized treatment and objective prediction of drug response [3]. The impetus to collect and disseminate relevant patient-specific data for use by clinicians, researchers, and drug developers has never been stronger. Simultaneously the impetus to provide patient-specific data to patients in a manner that is accurate, timely, and understandable, has also never been stronger.

This motivation takes specific form in the US where health care providers who want stimulus funded reimbursement from recent EHR funding to implement or expand the use of EMR's in care practices, must achieve "meaningful use". Achieving meaningful use requires both using certified EHR technology and achieving documented objectives that improve the quality, safety, and efficiency of care while simultaneously reducing disparities; engaging patients and families in their care; promoting public and population health; improving care coordination; and promoting the privacy and security of EHRs (CMS 2010)[1]. A 'certified' EHR must meet a collection of regulations and technical requirements to perform the required meaningful use functions (ONCHIT 2010)[2]. Minimum meaningful use requirements include fourteen core objectives, five out of ten specific objectives, and fifteen clinical quality measures (CMS 2010). These criteria, conditions and metric achievements are all delayed and complicated by the typical data fragmentation that occurs between the research and health care settings and will continue until a "translational" ontology is available to bridge activities, transferring data and entities between research and medical systems.

Translational medicine refers to the process by which the results of research done in the laboratory are directly used to develop new ways to treat patients. It depends on the comprehensive integration of the entire breadth of patient data with basic life science data to facilitate and evaluate drug development [4]. In the 1990s, Luciano pioneered the use of heterogeneous data integration, mathematical and computational modeling and simulation to tease apart the underlying dynamics and different individual treatment response patterns clinicians observed in patients diagnosed with Major Depressive Disorder [5] [6]. When information regarding the patient experience (symptoms, pharmacokinetics/pharmacodynamics, outcomes, side effects) can be directly linked to biomedical

[1]Centers for Medicare & Medicaid Services (CMS). Medicare & Medicaid EHR Incentive Program Meaningful Use Web Site. Available at: www.cms.gov/EHRIncentivePrograms/35_Meaningful_Use.asp. Last accessed August 2010.

[2]Office of the National Coordinator for Health Information Technology (ONCHIT). Standards & Certification Criteria Web Site. Available at: http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov_standards_ifr/1195. Last accessed August 2010.

knowledge (genetics, pathways, enzymes, chemicals, brain region activity), clinical research can gain new insights in causality and potential treatments. Detailed recordings of clinical encounters are a crucial component of this approach [7] [8] and devices such as personal electronic diaries[3] aid both patient and clinician in capturing accurate patient data of these accounts.

Semantic Web technologies enable the integration of heterogeneous data using explicit semantics, the expression of rich and well-defined models for data aggregation, and the application of logic to gain new knowledge over the raw data. Also, semantic technologies may be used to encode metadata such as provenance, e.g. where the data came from and how it was generated. Indeed, ontologies, which formalize the meaning of terms used in discourse are expected to play a major role in the automated integration of patient data with relevant information to support basic discovery and clinical research, drug formulation, and drug evaluation through clinical trials. The four main Semantic Web standards for knowledge representation are: Resource Description Framework (RDF); RDF Schema (RDFS); Web Ontology Language (OWL); and SPARQL as a query language. Already, OWL ontologies have been developed to support drug, pharmacogenomics and clinical trials [9] [10] [11] and are increasingly used in the health care and life sciences applications [12]. Collectively, these next generation semantic web technologies provide the resources required to systematically re-engineer both EHR and research data warehouse systems so that it becomes easier and more practical to integrate, query and analyze the full spectrum of relevant laboratory and clinical research data, as well as EHRs, in supporting the development of cost effective and outcome-oriented systems.

In this paper, participants in the Translational Medicine task force of the World Wide Web Consortium's Health Care and Life Sciences Interest Group (HCLSIG) present the Translational Medicine Ontology (TMO) and the Translational Medicine Knowledge Base (TMKB). The TMKB consists of TMO, mappings to other terminologies and ontologies, and data in RDF format across discovery research and drug development, of therapeutic relevance to clinical research and clinical practice. The TMO provides a foundation for types declared in Linking Open Drug Data (LODD) [13] and electronic health records (EHRs). The TMO captures core, high level terminology to bridge existing open domain ontologies and provides a framework to relate and integrate patient-centric data across the knowledge gap from bench to bedside. Using the framework of the TMO, we demonstrate with the TMKB how to bridge the gap and how to develop valuable translational knowledge pertinent to clinical research, and thence to clinical practice.

[3] http://www.symtrend.com

The remainder of the paper is structured as follows: we describe the use case for the TMKB, which involves Alzheimer's Disease, then describe the methods used to build the TMKB, the ontology design process, data sources and mappings. We then explore pertinent questions that the TMKB can answer in the results and discuss our findings, and conclude with future directions at tantalizing and yet unsolved problems.

Use Case

Alzheimer's Disease (AD) is an incurable, degenerative, and terminal disease with few therapeutic options [14] [15]. It is a complex disease influenced by a range of genetic, environmental, and other factors [15]. Recently, Jack *et al.* [16] demonstrated the value of shared data in AD biomarker research. A New York Times article quotes John Trojanowski at U Penn Medical School: "It was unbelievable, ...[we] parked our egos and intellectual-property noses outside the door and agreed that all of our data would be public immediately."[4] Efficient aggregation of relevant information improves our understanding of disease and significantly benefits researchers, clinicians, patients and pharmaceutical companies. By aggregating semantically annotated Alzheimer's Disease (AD) data from multiple data sources, we demonstrate the value of linked open data published using Semantic Web technologies to answer questions about AD diagnosis and therapeutic options.

Methods

As part of its requirements analysis, The HCLSIG Translational Medicine task force identified seven use cases against which its activities would be measured. These include scenarios involving chemogenomics, animal models, pharmacogenomics, therapeutic development, patient care, and integrative informatics. The full description of the details of each use cases can be found on the wiki site [5]. This work presented here follows questions asked in the patient care scenario, and are summarized in Table 1.

Ontology Design

The scope of the Translational Medicine Ontology (TMO) is defined by the use case terminology and their respective data sources. Each term and corresponding data source was analyzed for its conceptual and representational and reasoning capability as required by the use case requirements. TMO terms were obtained from a lexical analysis of sample research questions from 16 types of users, all of whom were

[4] "Sharing of Data Leads to Progress on Alzheimer's", Kolata, G. New York Times, August, 2010. http://www.nytimes.com/2010/08/13/health/research/13alzheimer.html

[5] http://esw.w3.org/HCLSIG/PharmaOntology/UseCases

involved in aspects of research, clinical care and or business (Table 1). Terms that refer to real world entities are were then represented as classes, relations or individuals in the ontology. Terms that appear in statements that hold in general (e.g. "patients participate in consultations" and "active ingredient is a role played by a molecular entity") form key background knowledge, refer to types that can be instantiated in the real world and are represented as classes in the ontology. 80 classes were created to represent material, processual, qualitative, attributive and informational entities of relevance to our study. By contrast, particulars (e.g. "a patient with a given name" and "a blister package of a pharmaceutical product with a particular identifying code on it") refer to individuals and these are represented as instances of classes in the ontology. Consequently, a particular consultation at a given time and day, the particular patient role in that consultation, and the physician role in that consultation can be represented as instances of classes in the ontology.

Figure 2 shows a portion of the TMO and illustrates selected types, subtypes and existential restrictions that hold between types. For instance, all chemical substances are chemical entities that are composed of molecular entities. Relations were specified using the relation ontology. A key part of designing the ontology laid in disambiguating polysemous terms e.g. "drug". A drug can refer to the whole pharmaceutical product or to the active ingredient. The TMO differentiates these meanings as a "molecular entity" (TMO_0034) for individual molecules, "active ingredient" (TMO_0000) for biologically active chemicals in formulated pharmaceuticals, "formulated pharmaceutical" (TMO_0001) for a substance that may or may not have been approved by a regulatory authority, and "pharmaceutical product" (TMO_0002) for a drug approved by a regulatory authority.

Given the prevalence of the terms defined in the ontology and the desire to establish the TMO as a global ontology, we also created 223 class equivalence mappings (using *owl:equivalentClass*) from 60 TMO classes to 201 target classes from 40 ontologies (see Table 2; Figure 3). These mappings were manually identified and verified using the NCBO Bioportal[6] and UMLS[7].

The TMO was built using Protégé 4.0.2 and represented as an OWL2 compliant ontology. TMO Terms are defined in the http://www.w3.org/2001/sw/hcls/ns/transmed/ namespace. The ontology is available from the TMO Google Code project[8].

[6]http://bioportal.bioontology.org
[7]http://www.nlm.nih.gov/research/umls/
[8]http://code.google.com/p/translationalmedicineontology/

Data Sources

The data sources used in this study include formulary lists, pharmacogenomics information, clinical trial lists, and scientific data about marketed drugs (Table 3). Clinicaltrials.gov is a registry of clinical trials, AD diagnostic refers to a formalized version of the diagnostic criteria for Alzheimer's Disease (AD), DailyMed contains marketed and FDA approved drugs, Diseaseome contains information about the genetic basis of disease, DrugBank contains detailed drug and drug target data, Medicare contains Medicare D approved drugs, Patient contains the synthetic patient data created for use in this study, PharmGKB contains data about drug response associated with genetic variation associated, and SIDER identifies side effects associated with marketed drugs.

All datasets except for PharmGKB, diagnostic criteria and patient records are available through the LODD[9] project [13]. Alzheimer's diagnostic criteria were formalized from the criteria panel described in Dubois et al. [17].

Seven synthetic patient records were manually created to capture typical medical record data: demographic information, contact information, family history, life style data, allergies, immunizations, information on conditions, procedures, prescriptions, and encounters with members of the medical community. Our records were to a large extent built upon the XML-based Indivo specification for personally-controlled health care records[10]. The Indivo initiative[11] offers simple user interfaces to store their records and to grant others controlled access to them. Archiving systems like i2b2's database records and Indivo's XML records can generically record data such as test results in tuples that include a coding system, a code, a tested value and the units of the value. For example, a systolic blood pressure measurement might use a SNOMED-CT code and mmHg units:

```
<VitalSign>
  <dateMeasured>2010-11-12T18:03Z</dateMeasured>
  <name type="http://...umls-snomed" value="271649006"
      abbrev="BPsys">Blood Pressure Systolic</name>
  <value>130</value>
  <unit type="http://codes.indivo.org/units/" value="31"
      abbrev="mmHg">millimeters of mercury</unit>
...</VitalSign>
```

[9] http://esw.w3.org/HCLSIG/LODD/Data
[10] http://wiki.indivohealth.org/index.php/Main_Page
[11] http://indivohealth.org/

We used GRDDL/XSLT[12] to define an RDF representation for Indivo patient records. A straightforward RDF representation of the above XML is:

```
_:X a  :VitalSign ;
 :dateMeasured \2010-11-12T18:03Z"^^xsd:dateTime ;
 :type <http://...umls-snomed#_BPsys> ;
 :value \130"^^<http://codes.indivo.org/units/#mmHg> .
```

Where possible, this representation instantiates types in the TMO ontology. However, this representation leaves the consumer having to normalize (e.g. MPa[13] to mmHg) before comparing or reporting values of potentially different units. Representing frequently needed and commonly used vital signs in a normalized form simplifies the effort needed to reuse these data:

```
_:X :systolicBPpascals \173322"^^<http://...#Pascals> .
```

Including the generic and the "standardized" forms allows us to meet a wide range of use cases[14]. Given that an XSLT stylesheet converts the XML-based Indivo data to instances of TMO classes, the mapping process should also perform this normalization. Currently, we normalize only a small set of vitals, but this is expected to expand as we draw on more diverse data.

Unit Testing

In order to keep our queries synchronized with the data model, we developed a simple test mechanism based on a practice of incremental development and testing. When changes are made to the data, incremental testing provides an efficient way to test all the known queries when changes are made to the data they match. Practically, this means critiquing the accuracy of the RDF representation, deciding whether it should be modeled differently, making changes (in our case to the XSLTs which generate the RDF), and finally invoking the unit testing system to determine whether queries can still be answered. The advantages of this workflow are increased accountability, increased agility/confidence, and error messages tied to recent edits[15].

[12] http://www.w3.org/2004/01/rdxh/spec

[13] In France, blood pressure (BP) values are reported in SI units (MPa). Pa = Pascal, MPa = megapascals

[14] This tension between flexibility and predictability is the crux of the art of standards.

[15] Our testing strategy could be described as "Extreme Ontology Development" a term derived from a programming methodology called "Extreme Programming" which incorporates regular and automated testing of essential application features into the development cycle and increases vigilance to the inadvertent errors that are typically introduced during development

Data Mapping

The questions in Table 1 are related to the patient scenario use case described on the wiki, detailed in 14 steps. The first step in the mapping was to work through each step, identifying key terms and a standard ontology that contains that term. For example, in Patient Scenario Step 9[16], we map the word "patient" to the "patient role" in the Ontology for Biomedical Investigation (OBI) ontology [OBI:0000093] and Physician to the NCI Theasaurus term Physician].

In the absence of identical matches on the labels, the Linkage Query Writer tool was used to create mappings between LODD datasets [18], along with Silk [19], which employs similarity metrics including string, numeric, data, URI, and set comparison methods. Entity identity was asserted using *owl:sameAs*. The mappings were augmented by those provided for PharmGKB via Bio2RDF [20]. Mappings between LODD dataset types and the TMO types were established using *owl:equivalentClass*.

Translational Medicine Knowledge Base

The Translational Medicine Knowledge Base (TMKB) is an RDFS-reasoning-capable Semantic Web knowledge base composed of the TMO, RDFized datasets, and equivalence mappings (Figure 1). The TMO, dataset, and mapping files were loaded into OpenLink Virtuoso 6 open source community edition, which provides a SPARQL endpoint[17] and a faceted text search interface[18]. To check the consistency of the knowledge base with more advanced reasoning, the ontology, equivalence mappings and datasets were loaded into a BigOWLIM repository[19]. BigOWLIM is a highly scalable reasoner that supports OWL2 RL reasoning[20].

Results and Discussion

Translational medicine requires the full extent of patient data to be accessible so that questions that span the multiple data sources such as those discussed herein can be asked and answered. For example, a physician within clinical practice would like to easily ask the criteria for the diagnosis of a disease and the prescription of personalized medicines. However, TMKB has the potential to be equally relevant to

[16] 9. In follow up, patient [patient role OBI:0000093] later reports nausea from Donepezil, and the physician [NCI Thesaurus: Physician] is aware of this common side effect (other side effects reported include bradycardia, diarrhea, anorexia, abdominal pain, and vivid dreams etc. . .) re-consults literature to ensure this is acceptable and agreeable with patient [patient role OBI:0000093]. If not, revisit loop above. Document side effect for post marketing adverse event pick up MedWatch, and future study. Change medication if necessary or add another medication to alleviate side effects. Micromedex, Facts and Comparisons. Consider moving patient to a trial.

[17] http://tm.semanticscience.org/sparql

[18] http://tm.semanticscience.org/fct

[19] http://www.ontotext.com/owlim/BigOWLIMFactSheet.pdf

[20] http://www.w3.org/TR/owl2-profiles/#OWL_2_RL

scientists developing new pharmaceutical products. While simple questions may be answered by queries on a single data set, other scientific questions may be addressed only when diverse data sets are fully integrated [21]. Importantly, answering more sophisticated questions may require inference i) over the subclass hierarchy of TMO types or ii) through equivalence mappings. Examples of queries that can now be executed with SPARQL[21] are listed in Table 4.

SPARQL Queries

To demonstrate the utility of the TMO and TMKB we created a set of twelve questions that are typical of the kinds of questions that arose in the requirements analysis when applied to the use cases. The wiki site contains the questions, the SPARQL source code and a clickable link that runs the query against the TMKB and displays the results[22]. The twelve queries are printed below. The SPARQL source and results are presented for two selected queries. Recall that the seven records created for patient data are fictitious and were created manually for the sole purpose of the demonstration. To run the queries, click on the link if provided, or copy the text of the SPARQL query and paste into the query text box at http://tm.semanticscience.org/sparql and clicking on "Run Query" button.

The significance of the SPARQL queries we present is to demonstrate that several different types of investigation, spanning information from different disciplines, can be carried out from the same query interface. In the hospital or clinic, the often fragmented information systems don't interoperate, requiring analogous investigations to coordinate between different specialists with access to different types of information. The combination of disparate types of information sources such as EHRs with clinical trial information, information about drugs and adverse reactions, as well as information about genetic variants, is crucial to reaching the goals of personalized medicine. It is precisely this type of information integration that is enabled by linked data approaches such as the one described here.

1. *How many patients experienced side effects while taking Donepezil?*

2. *What are the diagnostic criteria for AD?*

3. *Is Donepezil covered by Medicare D?*

[21] http://esw.w3.org/topic/HCLSIG/PharmaOntology/Queries

[22] At the time of this writing, fourteen exemplar questions are presents on the wiki site with corresponding SPARQL source code. Clickable results are presented for the first ten. http://esw.w3.org/HCLSIG/PharmaOntology/Queries

4. *Have any of my AD patients been treated for other neurological conditions as this might impact their diagnosis?*
5. *Are there other clinical trials that my patient may participate in for AD which have a different*
6. *Are there any AD patients without the APOE4 allele as these would be good candidates for the clinical trial involving Bapineuzumab?*
7. *What active trials are ongoing that would be a good fit for Patient 2?*
8. *Do I have suitable patients for an AD trial where they are looking for females who are aged over 55 years, have the APOE variant, and low ADAS COG scores?*
9. *What genes are associated with or implicated in AD?*
10. *What biomarkers are associated with or implicated in AD?*
11. *An APOE variant is strongly correlated with AD predisposition. Are there drug classes and drugs target APOE?*
12. *Which existing marketed drugs might potentially be re-purposed for AD because they are known to modulate genes that are implicated in the disease?*
13. *What are the results of Georg Steffen Möller's lipid panel?*
14. *What is Monica Mary Mall's platelet count over time?*

The following query demonstrates the ability to perform patient eligibility studies when the appropriate information is accessible. Finding eligible patients can be a costly endeavor for clinical trials so this query can save significant costs, as well as increase the effectiveness of treatment. The APOE allele can be identified from a blood test. Next generation sequencing is expected to bring more specific genetic information to bear and make medicine even more personalized:

Q. 6 Are there any AD patients without the APOE4 allele as these would be good candidates for the clinical trial involving Bapineuzumab?

The SPARQL query:

```
PREFIX trans: <tag:eric@w3.org:2009/tmo/translator#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT distinct (?name) ?patient ?testname ?apoe4

WHERE {
 ?encounter trans:patient ?patient .
 ?patient foaf:name ?name .
 ?patient trans:hasCondition ?condition .
 ?condition trans:diagnosedWith ?diagnosis .
 FILTER (regex (?diagnosis, "alzheimer", "i"))

OPTIONAL {
 ?encounter trans:test ?test .
 ?test trans:testName ?testname .
 ?test trans:result ?result .
 ?result trans:variant_Synonyms ?apoe4
 FILTER (regex (?apoe4, "APOE4"))
}
}
```

Results:

name	patient	testname	apoe4
Benny Smith	http://tag:eric@w3.org:2009/pchr/3#me		
Edward Quesada	http://tag:eric@w3.org:2009/pchr/4#me		
Edward Quesada	http://tag:eric@w3.org:2009/pchr/4#me	ADmark Alzheimer's Evaluation	APOE4, NG_007084.2:g.7903T>C
Julianne Sarah Christopherson	http://tag:eric@w3.org:2009/pchr/7#me		
Julianne Sarah Christopherson	http://tag:eric@w3.org:2009/pchr/7#me	ADmark Alzheimer's Evaluation	APOE4, NG_007084.2:g.7903T>C
Georg Steffen Möller	http://tag:eric@w3.org:2009/pchr/5#me		
George Andrew Tour	http://tag:eric@w3.org:2009/pchr/1#me		
George Andrew Tour	http://tag:eric@w3.org:2009/pchr/1#me	ADmark Alzheimer's Evaluation	APOE4, NG_007084.2:g.7903T>C
Monica Mary Mall	http://tag:eric@w3.org:2009/pchr/2#me		
Monica Mary Mall	http://tag:eric@w3.org:2009/pchr/2#me	ADmark Alzheimer's Evaluation	APOE4, NG_007084.2:g.7903T>C

This next query presents an example of repurposing existing marketed drugs. We understand this to be of interest to the pharmaceutical industry because of the huge savings in time and money for development and clinical trials. The benefits also translate to physicians and patients because it means that medicines may be available sooner to help manage medical conditions. This query takes advantage of the information in PharmGKB, in which the relations between genes, drugs, and diseases are tracked.

Q. 12 Which existing marketed drugs might potentially be re-purposed for AD because they are known to modulate genes that are implicated in the disease?

The SPARQL query:

```
PREFIX pharmgkb: <http://bio2rdf.org/pharmgkb:>
```

```
SELECT distinct ?drug_name ?disease2_name
WHERE {
 GRAPH <pharmgkb> {
  ?association rdf:type pharmgkb:DrugGeneVariantInteraction .
  ?association pharmgkb:description ?description .
  ?association pharmgkb:disease ?disease .
  ?association pharmgkb:variant ?variant .
  ?disease rdfs:label ?disease_name
  FILTER regex(?disease_name,"alzheimer","i") .
  ?association pharmgkb:gene ?gene .
  ?gene dc:identifier ?gene_name .

  ?a2 a pharmgkb:Association .
  ?a2 pharmgkb:gene ?gene .
  ?a2 pharmgkb:disease ?d2 .
  ?d2 rdfs:label ?disease2_name .
  ?a2 pharmgkb:drug ?drug .
  ?drug rdfs:label ?drug_name
 }
}
order by asc(?drug_name)
```

Results[23]:

[23]We present the first 25 lines of the results. The full result can be viewed by pasting the query into the query text box at http://tm.semanticscience.org/sparql and clicking on "Run Query" button.

drug_name	disease2_name
(s)-rolipram	Schizophrenia
(s)-rolipram	Autistic Disorder
(s)-rolipram	Bipolar Disorder
(s)-rolipram	Depression
ACE INHIBITORS, PLAIN	Angioneurotic Edema
ACE INHIBITORS, PLAIN	Hypertension
ACE INHIBITORS, PLAIN	Hypertrophy, Left Ventricular
ACE INHIBITORS, PLAIN	Coronary Disease
ACE INHIBITORS, PLAIN	Alzheimer Disease
ACE INHIBITORS, PLAIN	nondiabetic proteinuric nephropathy
ACE INHIBITORS, PLAIN	Alcoholism
ACE INHIBITORS, PLAIN	Abnormalities
ACE INHIBITORS, PLAIN	Fetal Death
ACE INHIBITORS, PLAIN	Cardiovascular Abnormalities
ACE INHIBITORS, PLAIN	Cardiovascular Diseases
ACE INHIBITORS, PLAIN	Cough
ACE INHIBITORS, PLAIN	Heart Failure
ACE INHIBITORS, PLAIN	Kidney Diseases
ANGIOTENSIN II ANTAGONISTS AND CALCIUM CHANNEL BLOCKERS	Cardiovascular Diseases
ANGIOTENSIN II ANTAGONISTS AND CALCIUM CHANNEL BLOCKERS	Hypertension
ANTIPSYCHOTICS	Schizophrenia
BETA BLOCKING AGENTS	Abnormalities
BETA BLOCKING AGENTS	Fetal Death
BETA BLOCKING AGENTS	Cardiovascular Abnormalities
atenolol	glomerulosclerosis
...	...

Related Work

Translational medicine, the idea of integrating the research pipeline from bench to bedside and back, has been a high priority for national biomedical research programs around the world. NIH's Clinical and Translational Science Awards (CTSAs), set forth by Zerhouni [22], provide leadership in translational research and have been fruitful in producing semantic translational informatics projects [23]. Additionally, a European Union joint undertaking, introduced by Kamel *et al.* [24], created the Innovative Medicine Initiative (IMI). Translational informatics has long been a use case for biomedical semantics. Use cases such as those described in Kashyap *et al.* [25] are being addressed through a number of projects, such as the BRIDG model, a joint project between the Clinical Data Interchange Standards Consortium (CDISC), the HL7 Regulated Clinical Research Information Management Technical Committee (RCRIM TC), the

National Cancer Institute (NCI), and the US Food and Drug Administration (FDA). Its goal is to produce a shared view of the dynamic and static semantics for protocol-driven research.[24] Other efforts have included development of large-scale terminologies, such as the NCI Thesaurus [26] and the Systematized NOmenclature of MEDicine-Clinical Terms (SNOMED-CT) [27]. The Informatics for Integrating Biology and the Bedside (i2b2) [28] project has developed a platform to integrate data from diverse sources, including free text and stuctured databases.

Conclusions

The TMO supports translational medicine by providing a model that facilitates interoperability of data from bench to bedside. Our AD-focused use case demonstrates the use of TMKB in translational research in the context of a well known disease. The TMKB has also been shown as a good candidate for providing more personalized information for patient treatment. While the medical history of our sample patients is not extensive, it reflects the reality of incomplete medical records in practice today within many institutions. Consistency and completeness of EHRs will be increasingly important in collaboration between researchers and physicians. More effective integration of data, as we have demonstrated here through the use of applied ontological methods, should enable data mining in a clinical setting to identify superior efficacy of certain drugs over others in specific sections of the population. "Patterns" detected in large data repositories can only be accurately detected if the form and consistency of data is assured. "Noisy" or contaminated data can generate false patterns or generate sufficient noise that true patterns are undetected. A clinician should be able to efficiently obtain a list of safe, effective, evidence-based therapies for administration to a specific patient while considering what payers can afford.

Since our work specifically focused on integrating existing datasets using a common vocabulary, we inevitably acquired terms that are either difficult to define within the context of the TMO or cannot be found in an existing community ontology. For example, the term "side effect" is particularly challenging because side effects in themselves are so varied. Nightmares are processes, but tender gums are dispositions that are realized in processes (sensation of pain in gums when palpated). While the TMO has "adverse drug event" (TMO_0043), it will take time and effort to correctly assign the full set of side effects listed in SIDER. In addition to the significant health related need for a uniform ontology, in the US, there are now approximately 40 Clinical and Translational Science Centers with approximately 20 more centers to be funded. Each center provides a robust informatics core supporting the entire spectrum of translational

[24] http://www.bridgmodel.org

science activity. At present, approximately half of the funded centers and some additional 20 research and commercial biomedical research groups around the world use Harvard Medical School's i2b2 platform. The i2b2 system provides a tremendous opportunity to test TMO's impact in a broad collection of translational medicine programs and projects. We intend to incorporate the current release of TMO into the i2b2 platform and design a set of pilot projects using TMO to accelerate the research and clinical efforts.
Future work will focus on the addition of entities related to drug discovery and drug development in order to increase its utility for the pharmaceutical industry. We aim to incorporate pathway references [29], to support additional pharmaceutical industry use cases. A broader goal is to enable interoperability with large scale e-Science work [30] [31]. In order to do this, we would need to extend the underlying representation to include provenance. Encodings could be done in a provenance interlingua such as the Proof Markup Language [32] or the Open Provenance Model [33]. Many interdisciplinary eScience efforts find that they need to provide services to access information such as the sources relied on to generate a conclusion or the transformations applied to the data or assumptions embodied in the data. Further we hope to support deeper semantic scientific knowledge integration [34]. Additionally, we hope to engage in additional evaluation of data to identify potential inconsistencies and readiness for use. We have utilized logical consistency checking such as the services available by state of the art OWL reasoners, but we may expand to either utilize or build additional evaluation services that may, for example, check instance data for possible problems such as those encountered when at the border between open and close-world reasoning [35]. In addition, given the project's reliance on equivalence links, we may explore using additional types of equivalence or similarity relationships such as those in [36], [37].
Another key goal is the development of a role-based user interface that would encourage vendors of EHRs to use ontologies such as the TMO and ontology-enhanced services not only to guide question answering, but also to improve representation and integration of data [38]. The TMKB is intended to provide a first step towards normalizing the sharing and integration of research and clinical artifacts. We wish to enable scientists to capitalize on the benefits derived from open data, communities of practice, and semantic web technology for reasoning across vast amounts of health care and life science data. The TMO can also be used to power a set of ontology-enhanced services such as ontology-enhanced search, provenance, and verification services, thus helping to improve accuracy, trust, and accountability of scientific information.
And lastly, we would like to support semantic publishing, referencing and authoring efforts such as SPAR

[25][26][27] by including refrences to terms in those ontologies.

Authors contributions

BA contributed to discussions and provided pharma perspectives on use cases.

CB was involved in the original development of the OWL2 ontology and contributed to the formal ontological structure of TMO and the discussion of the ontological challenges of side effects.

TC reviewed the manuscript and provided information on the AD use case.

CD is a clinician and helped to develop each of the exlempar patient records while providing clinical guidance and support.

AJ did the work using LinQuer for mapping the data sources.

JK participated in telecons and contributed to the definition of terms.

PK participated in TMO use case development and transformed narrative descriptions into medical coding; created xml structure for family history, immunizations, lifestyle, and encounters with medical coding designations; contributed to draft and final manuscripts.

JSL is a computational and life scientist with expertise in semantic web applications to the life sciences; contributed to the development of the TMO and LODD mapping; provided expertise and guidance and contributed to the writing, editing, and scoping of the manuscript.

MD is the current lead for the HCLSIG Translational Medicine task force; he provided guidance and expertise, implemented the ontology, mappings, data (PharmGKB), SPARQL queries, and contributed to the writing and editing of the manuscript.

MSM co-chaired the HCLS IG, participated in teleconferences, and contributed to the writing and editing of the manuscript.

JPM is a bioinformaticist and computer scientist, and wrote the related work section. He also provided proofreading, overall review, formatting, and coordination between authors.

DLM is an expert in semantic web languages and environments and wrote a future work segment, consulted on the semantic approach, and reviewed and edited the paper and its semantic approach.

EP contributed to the development of TMO and to the mapping of TMO classes to other reference ontologies and source vocabularies.

[25]SPAR: Semantic Publishing and Referencing http://esw.w3.org/HCLSIG/SWANSIOC/Actions/RhetoricalStructure/meetings/20101115
[26]http://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies/
[27]SALT http://salt.semanticauthoring.org/ontologies/sro

ericP and TG expressed patient data in Indivo, created the query testing framework and produced XSLT to map Indivo to TMO.

RLP was involved in discussions of TMO.

MS gave advice during ontology development and worked on knowledge base consistency checks.

SS co-chaired the HCLS IG, coordinated the task force prior to MD, created the patient records and contributed to the development of the the TMO.

PJT participated in TMO use case development; confirmed use case consistency and accuracy; contributed to draft and final manuscript.

PLW contributed to the development of TMO, participated in teleconferences, and reviewed the paper.

Other authors participated in conference calls or made other noteworthy contributions to the use cases, TMO or TMKB development effort.

Acknowledgements

We would like to thank the National Center for Biomedical Ontology (NCBO) for their contribution to this work. We also thank the entire Translational Medicine Ontology Task Force, and more generally the support provided by the World Wide Web Consortium (W3C) to the W3C Health Care and Life Science (HCLS) Interest Group. The authors would also like to thank Christine Denny for her contributions to the development of the TMO and participation in the conference calls.

References

1. Trusheim M, Berndt E, Douglas F: **Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers**. *Nature Reviews Drug Discovery* 2007, **6**(4):287–293.
2. Rodwin M: **The case for public ownership of patient data**. *JAMA* 2009, **302**:86.
3. Roses A: **Pharmacogenetics in drug discovery and development: a translational perspective**. *Nature Reviews Drug Discovery* 2008, **7**(10):807–817.
4. Woolf S: **The meaning of translational research and why it matters**. *Jama* 2008, **299**(2):211.
5. Luciano JS, Negishi M, Cohen MA, Samson JA: **Depression Research: Modeling to Illuminate Darkness**. In *Neural Modeling of Cognitive and Brain Disorders*. Edited by Reggia J, Ruppin E, Berndt R, World Scientific Publishing Company 1996.
6. Luciano JS: **Neural Network Modeling of Unipolar Depression: Patterns of Recovery and Prediction of Outcome**. *PhD thesis*, Boston University 1996.
7. Levine M, Calvanio R: *The Recording of Personal Information as an Intervention and as an Electronic Health Support*. Springer 2007.
8. Calvanio R, Buonanno F, Levine D, Levine M: **Neuropsychiatric sequelae and life events: Analysis and management**. In *6th World Stroke Congress* 2008.
9. Dumontier M, Villanueva-Rosales N: **Towards pharmacogenomics knowledge discovery with the semantic web**. *Briefings in bioinformatics* 2009, **10**(2):153.

10. Coulet A, Smail-Tabbone M, Napoli A, Devignes M: **Suggested Ontology For Pharmacogenomics (SO-Pharm): Modular Construction And Preliminary Testing**. *Lecture Notes in Computer Science* 2006, **4277**:648–657.

11. Arikuma T, Yoshikawa S, Azuma R, Watanabe K, Matsumura K, Konagaya A: **Drug interaction prediction using ontology-driven hypothetical assertion framework for pathway generation followed by numerical simulation**. *BMC bioinformatics* 2008, **9**(Suppl 6):S11.

12. Shah N, Jonquet C, Chiang A, Butte A, Chen R, Musen M: **Ontology-driven indexing of public datasets for translational bioinformatics**. *BMC bioinformatics* 2009, **10**(Suppl 2):S1.

13. Jentzsch A, Zhao J, Hassanzadeh O, Cheung K, Samwald M, Andersson B: **Linking open drug data**. In *Triplification Challenge of the International Conference on Semantic Systems*, Citeseer 2009.

14. Patterson C, Feightner J, Garcia A, Hsiung G, MacKnight C, Sadovnick A: **Diagnosis and treatment of dementia: 1. Risk assessment and primary prevention of Alzheimer disease**. *Canadian Medical Association Journal* 2008, **178**(5):548.

15. Minati L, Edginton T, Grazia Bruzzone M, Giaccone G: **Reviews: Current Concepts in Alzheimer's Disease: A Multidisciplinary Review**. *American Journal of Alzheimer's Disease and Other Dementias* 2009, **24**(2):95.

16. Jack C, Wiste H, Vemuri P, Weigand S, Senjem M, Zeng G, Bernstein M, Gunter J, Pankratz V, Aisen P, et al.: **Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease**. *Brain* 2010, **133**(11):3336.

17. Dubois B, Feldman H, Jacova C, DeKosky S, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, et al.: **Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria**. *The Lancet Neurology* 2007, **6**(8):734–746.

18. Hassanzadeh O, Kementsietsidis A, Lim L, Miller R, Wang M: **A framework for semantic link discovery over relational data**. In *Proceeding of the 18th ACM conference on Information and knowledge management*, ACM 2009:1027–1036.

19. Volz J, Bizer C, Gaedke M, Kobilarov G: **Silk–a link discovery framework for the web of data**. In *Proceedings of the 2nd Linked Data on the Web Workshop* 2009.

20. Belleau F, Nolin M, Tourigny N, Rigault P, Morissette J: **Bio2RDF: Towards a mashup to build bioinformatics knowledge systems**. *Journal of biomedical informatics* 2008, **41**(5):706–716.

21. Stephens S, LaVigna D, DiLascio M, Luciano J: **Aggregation of bioinformatics data using Semantic Web technology**. *Web Semant.* 2006, **4**:216–221, [http://portal.acm.org/citation.cfm?id=1222219.1222307].

22. Zerhouni E: **Translational and clinical science–time for a new vision**. *New England Journal of Medicine* 2005, **353**(15):1621.

23. Mirhaji P, Zhu M, Vagnoni M, Bernstam E, Zhang J, Smith J: **Ontology driven integration platform for clinical and translational research**. *BMC bioinformatics* 2009, **10**(Suppl 2):S2.

24. Kamel N, Compton C, Middelveld R, Higenbottam T, Dahlén S: **The Innovative Medicines Initiative (IMI): a new opportunity for scientific collaboration between academia and industry at the European level**. *European Respiratory Journal* 2008, **31**(5):924.

25. Kashyap V, Hongsermeier T: **Can semantic web technologies enable translational medicine?** *Semantic Web* 2007, :249–279.

26. Sioutos N, Coronado S, Haber M, Hartel F, Shaiu W, Wright L: **NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information**. *Journal of biomedical informatics* 2007, **40**:30–43.

27. Stearns M, Price C, Spackman K, Wang A: **SNOMED clinical terms: overview of the development process and project status.** In *Proceedings of the AMIA Symposium*, American Medical Informatics Association 2001:662.

28. Murphy S, Weber G, Mendis M, Gainer V, Chueh H, Churchill S, Kohane I: **Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2)**. *Journal of the American Medical Informatics Association* 2010, **17**(2):124.

29. Luciano J, Stevens R: **e-Science and biological pathway semantics**. *BMC Bioinformatics* 2007, **8**(Suppl 3):S3, [http://www.biomedcentral.com/1471-2105/8/S3/S3].

30. Hey T, Trefethen A: **Cyberinfrastructure for e-Science**. *Science* 2005, **308**(5723):817–821, [http://dx.doi.org/10.1126/science.1110410].

31. Hey T, Trefethen A: **e-Science and its implications**. *Philos Transact A Math Phys Eng Sci* 2003, **361**(1809):1809–1825, [http://dx.doi.org/10.1098/rsta.2003.1224].

32. McGuinness D, Ding L, da Silva P, Chang C: **Pml 2: A modular explanation interlingua**. In *Proceedings of AAAI, Volume 7* 2007.

33. Moreau L, Clifford B, Freire J, Futrelle J, Gil Y, Groth P, Kwasnikowska N, Miles S, Missier P, Myers J, et al.: **The open provenance model core specification (v1. 1)**. *Future Generation Computer Systems* 2010.

34. McGuinness D, Fox P, Brodaric B, Kendall E: **The Emerging Field of Semantic Scientific Knowledge Integration**. *IEEE Intelligent Systems* 2009, :25–26.

35. Tao J, Ding L, McGuinness D: **Instance data evaluation for semantic web-based knowledge management systems**. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*, IEEE 2009:1–10.

36. Halpin H, Hayes P, McCusker J, McGuinness D, Thompson H: **When owl: sameAs isn't the Same: An Analysis of Identity in Linked Data**. In *Proc. 9th Int. Semantic Web Conf* 2010.

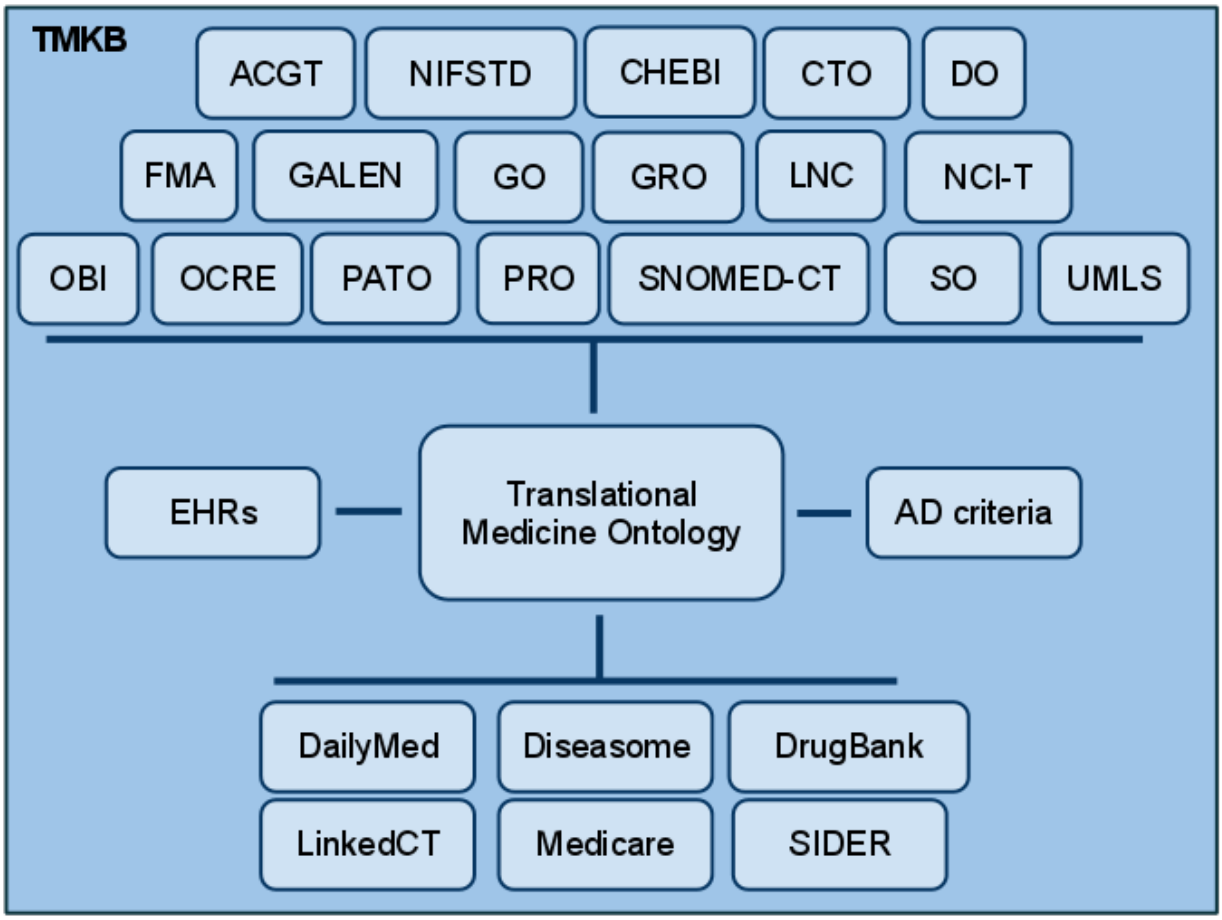
37. Ding L, Shinavier J, Shangguan Z, McGuinness D: **SameAs Networks and Beyond: Analyzing Deployment Status and Implications of owl: sameAs in Linked Data**. In *Proc. 9th Int. Semantic Web Conf* 2010.

38. Goble C, Pettifer S, Stevens R, Greenhalgh C: **Knowledge Integration: In Silico Experiments in Bioinformatics**. *The Grid: Blueprint for a New Computing Infrastructure* 2003, :121–134.

39. Scheuermann R, Ceusters W, Smith B: **Toward an ontological treatment of disease and diagnosis**. *Proceedings of the 2009 AMIA Summit on Translational Bioinformatics* 2009, :116–120.

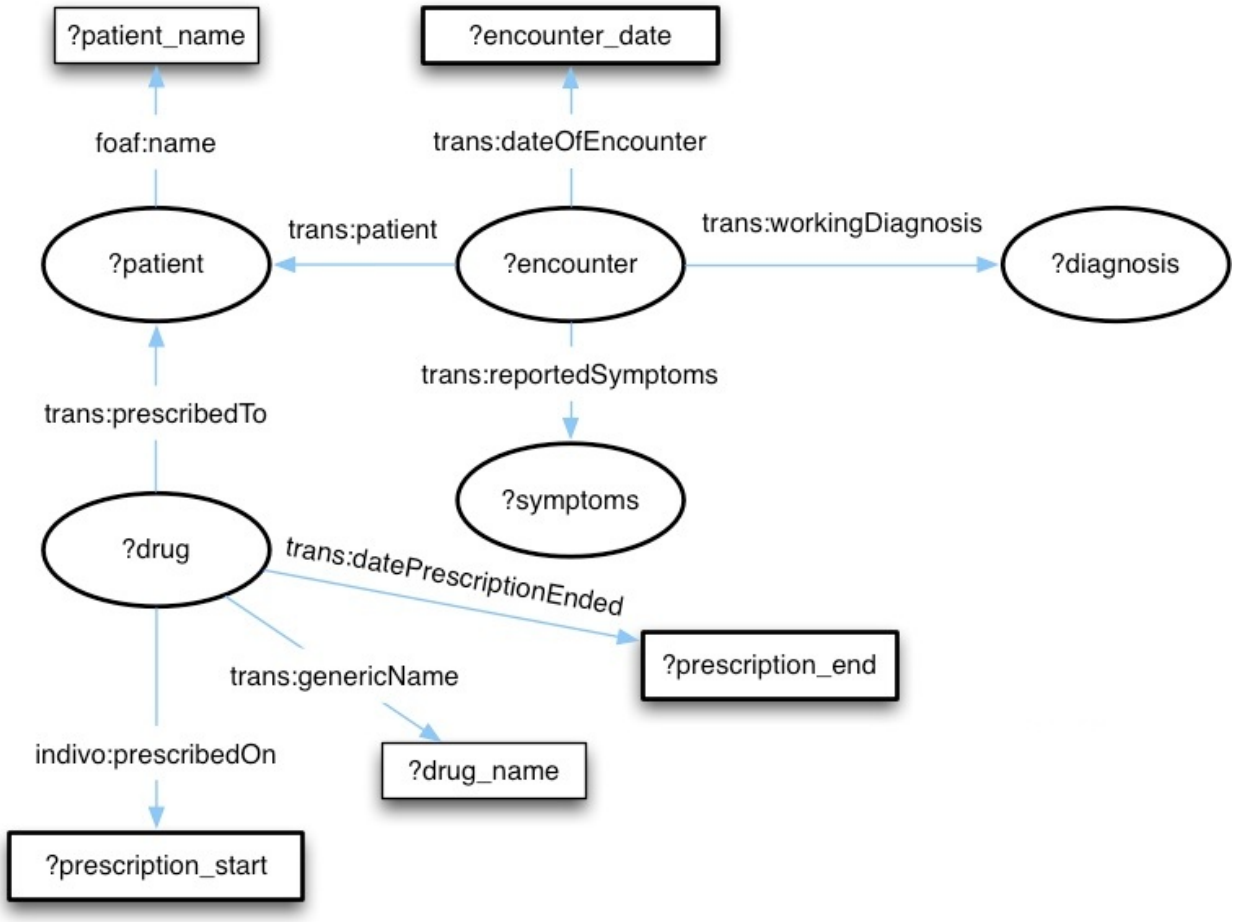
Figures

Figure 1 - TKMB Overview



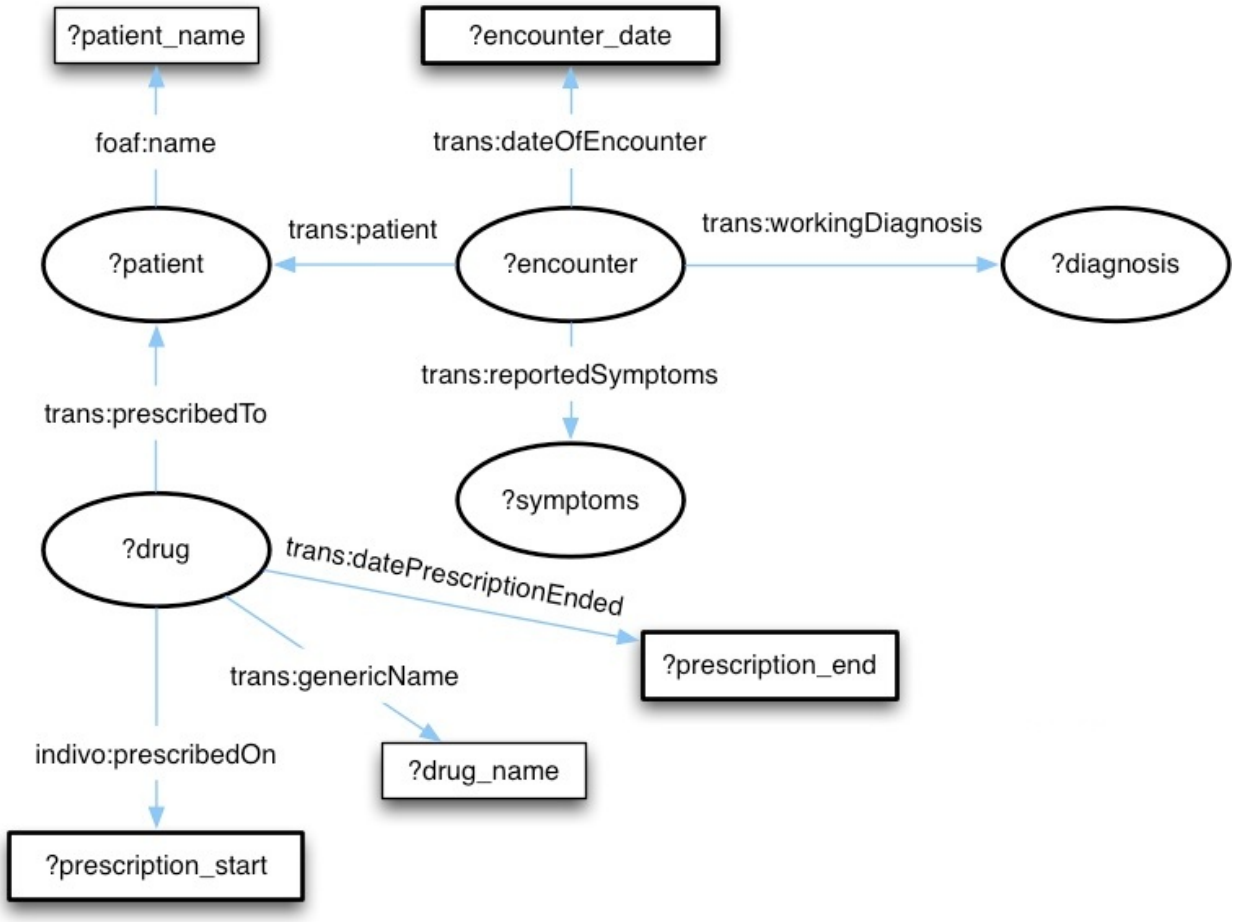
Overview of the contents of the Translational Medicine Knowledge Base (TMKB). TMKB is composed of the Translational Medicine Ontology with mappings to ontologies and terminologies listed in the NCBO bioportal. The TMO provides a global schema for Indivo-based electronic health records (EHRs) and can be used with formalized criteria for Alzheimer's Disease. The TMO maps types from Linking Open Data sources.

Figure 2 - Translational Medicine Ontology Overview



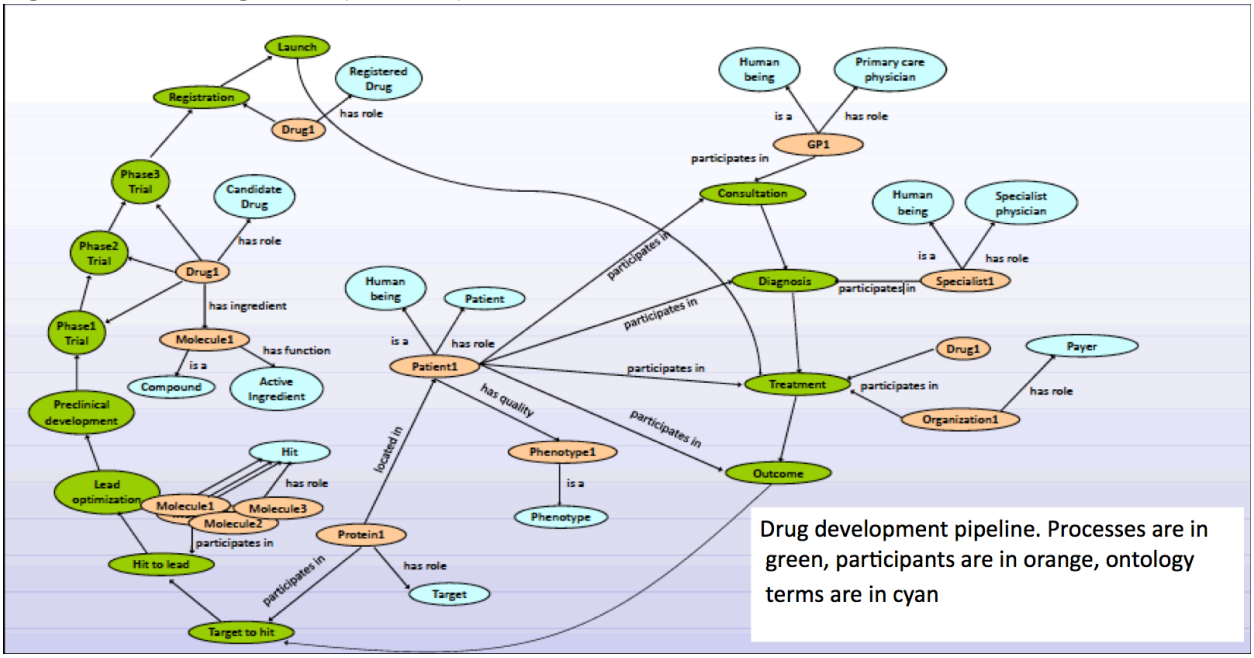
Overview of selected types, subtypes (overlap) and existential restrictions (arrows) in the Translational Medicine Ontology.

Figure 3 - Query #1: side effects



The data elements involved in query #1. The query can be formulated as "How many patients experienced side effects while taking Donepezil?"

Figure 4 - The Drug Development Pipeline



The drug development pipeline, outlining important processes, participants, and terms.

Tables

Table 1 - Users and their interests in translational medicine

The TMO defines 75 classes spanning material entities (e.g. molecule, protein, cell lines, pharmaceutical preparations), roles (e.g. subject, target, active ingredient), processes (e.g. diagnosis, study, intervention), and informational entities (e.g. dosage, mechanism of action, sign/symptom [39], family history). The TMO extends the basic types defined in the Basic Formal Ontology and uses relations from the Relation Ontology.[28]

[28]source: http://esw.w3.org/topic/HCLSIG/PharmaOntology/Roles

Category	User	Interest
Research	Biologist (*in vivo*, *in vitro*, cellular & molecular)	Target identification, assay development, target validation
	Bioinformatician	Biological knowledge management, cellular modeling
	Immunologist	Natural defense mechanisms
	Cheminformatician	Predictive chemistry
	Medicinal chemist	Drug efficacy
	Systems physiologist	Tolerance, adverse events
Clinic	Clinical trial specialist	Trial formulation, recruitment
	Clinical decision support	Data analysis, trend finding
	Primary care physician	General, conventional care
	Specialty medical provider	Specialized treatments
Business	Sales & marketing	Revenue generation
	Strategic/portfolio manager	Assessing market opportunities
	Project manager	Prioritizing resources & activities
	Health plan provider	Insurance coverage

Table 2 - Representative mappings between TMO and target terms

Abbreviations: ACGT- ACGT Master Ontology, NIFSTD – Neuroscience Information Framework Standardized ontology, CHEBI – Chemical Entities of Biological Interest, CTO – Clinical Trial Ontology, DOID – Human Disease Ontology, FMA – Foundation Model of Anatomy, FHHO – Family Health History Ontology, Galen – Galen Ontology, GO – Gene Ontology, GRO – Gene Regulation Ontology, LNC – Logical Observation Identifer Names and Codes, MSH- Medical Subject Headings, NCIt – NCI theraurus, NDFRT – National Drug File, OBI – Ontology for Biomedical Investigation, OCRe_- Ontology for Clinical Research, PATO – Phenotypic Quality Ontology, PRO – Protein Ontology, SNOMED-CT, SNOMED clinical terms, SO – Sequence Ontology, UMLS – Unified Modeling Language System.

Label	TMO	Target
Protein	0035	ACGT:Protein, BIRNLex:23, CHEBI:36080, FMA:Protein, GO:0003675, GRO:Protein, Galen:Protein, NCIt:Protein, PRO:000000001, SNOMEDCT:88878007, SO:0000358, UMLS:C0033684
Gene	0037	FMA:Structural_gene, GRO:Gene, Galen:Gene, LNC:LP32747-5, MSH:D005796, NCIt:Gene, NCIt:Gene_Object, NDFRT:C242394, PRO:Gene, SNOMEDCT:67271001, SO:0000704, UMLS:C0017337
Diagnosis	0031	ACGT:Diagnosis, FHHO:Diagnosis, Galen:Diagnosis, LNC:LP72437-4, MSH:D003933, NCIt:Diagnosis, OBI:0000075, OCRe_clinical:Diagnosis, SNOMEDCT:439401001, UMLS:C0011900
Disease	0047	ACGT:Disease, BIRNLex:11013, DOID:4, GRO:Disease, LNC:LP21006-9, MSH:D004194, NCIt:Disease_or_Disorder, NDFRT:C2140, OBI:0000155, UMLS:C0012634

Table 3 - Data sources used in this study

All datasets except for PharmGKB, diagnostic criteria and patient records are available through the Linking Open Drug Data (LODD)[29] project [13]. Alzheimer's diagnostic criteria were obtained from Dubois et al. (Dubois et al. 2007). Seven patient health records were manually created to capture demographic information, contact information, family history, life style data, allergies, immunizations, information on conditions, procedures, prescriptions, and encounters with members of the medical community. The patient record was defined by an XML schema, based in part on the Indivo schema[30], and converted into RDF using an XSL stylesheet.

LODD	Prefix	Dataset	Description
x	linkedct	Clinicaltrials.gov	Registry of clinical trials
	dubois	AD diagnostic	AD diagnostic criteria
x	dailymed	DailyMed	Marketed & FDA approved drugs
x	diseasome	Diseasome	The genetic basis of disease
x	drugbank	DrugBank	Detailed drug data & drug target
x	medicare	Medicare	Medicare D approved drugs
	pchr	Patient	synthetic patient data
	pharmgkb	PharmGKB	Drug response to genetic variation
x	sider	SIDER	Side effects of marketed drugs

LODD – 'x' indicates a linking open drug data dataset

Table 4 - Questions and answers using TMO-integrated data sources

[29] http://esw.w3.org/HCLSIG/LODD/Data
[30] http://wiki.indivohealth.org/index.php/Main_Page

Question	Answer
Clinic	
What are the diagnostic criteria for AD?	There are 12 diagnostic inclusion criteria and 9 exclusion criteria.
Does Medicare D cover Donepezil?	Medicare D covers 2 brand name formulations of Donepezil: Aricept and Aricept ODT.
Have any AD patients been treated for other neurological conditions	Patient 2 was found to suffer from AD and depression.
Clinical Trial	
Since my patient is suffering from drug-induced side effects for AD treatment, identify an AD clinical trial with a different mechanism of action (MOA)	Of the 438 drugs linked to AD trials, only 58 are in active trials and only 2 (Doxorubicin and IL-2) have a documented MOA. 78 AD-associated drugs have an established MOA.
Find AD patients without the APOE4 allele as these would be good candidates for the clinical trial involving Bapineuzumab?	Of the four patients with AD, only one does not carry the APOE4 allele, and may be a good candidate for the clinical trial.
What active trials are ongoing that would be a good fit for Patient 2?	58 Alzheimer trials: 2 mild cognitive impairment, 1 hypercholesterolaemia, 66 myocardial infarction, 46 anxiety, and 126 depression.
Research	
What genes are associated with or implicated in AD?	Diseasome and PharmGKB indicate at least 97 genes have some association with AD.
Which SNPs may be potential AD biomarkers?	PharmGKB reveals 63 SNPs
Which market drugs might potentially be re-purposed for AD because they modulate AD implicated genes?	57 compounds or classes of compounds that are used to treat 45 diseases, including AD, hyper/hypotension, diabetes and obesity.