



The Translational Medicine Ontology

Facilitating Drug Development for Personalized Medicine

Elgar Pichler

W3C HCLSIG TMO Team Members

BioIT 2010, Boston, MA

April 21, 2010



Outline

- Questions & Problems
- Translational Medicine Ontology (TMO)
 - Ontology
 - Translational Medicine Knowledge Base (TMKB)
 - Example
- Comments

Introduction

Translational Medicine comprises many diverse areas including

- _ hypothesis management,
- _ discovery research,
- _ drug development and formulation,
- _ clinical research, and
- _ clinical practice.

Problem: Because of a **lack of a common terminology** for these areas, information flow between them is often hindered or very limited resulting in

- _ suboptimal patient care, and
- _ increased healthcare costs.

The **Translational Medicine Ontology (TMO)** attempts to **provide a common terminology for Translational Medicine.**

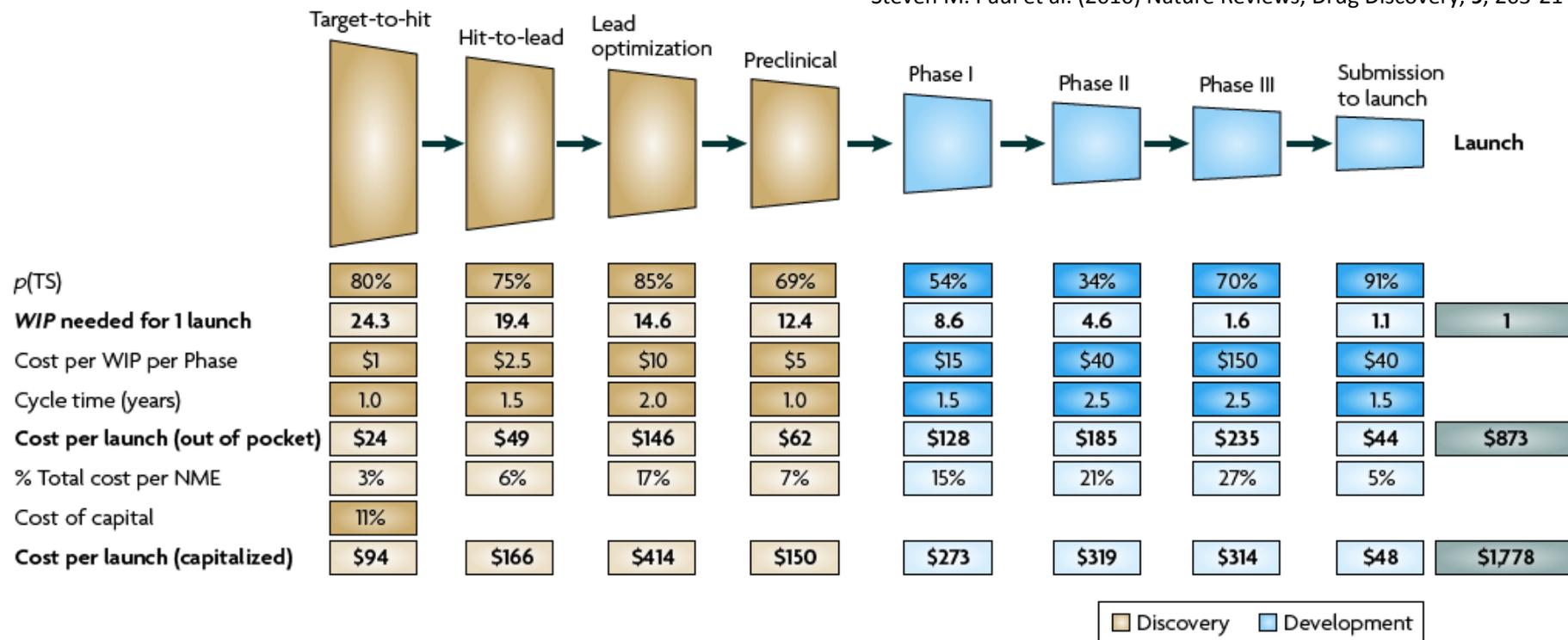
TMO should also serve as a global schema for data integration, and facilitate the formulation and answering of complex queries across heterogeneous sources. As such TMO work forms a basis for the development of a computational platform for managing information relevant to personalized medicine.

Questions & Problems

Example: The Drug Development Pipeline

"How to improve R&D productivity: the pharmaceutical industry's grand challenge",

Steven M. Paul et al. (2010) Nature Reviews, Drug Discovery, 9, 203-214



- The road is long, and costly.
- How do we contain costs and develop better drugs?

Questions & Problems

Iressa – How to Select Responsive Patients?

Scienceexpress

Report

Gefitinib-Sensitizing *EGFR* Mutations in Lung Cancer Activate Anti-Apoptotic Pathways

Raffaella Sordella, Daphne W. Bell, Daniel A. Haber, Jeffrey Settleman*

Center for Molecular Therapeutics, Massachusetts General Hospital Cancer Center and Harvard Medical School, Bldg 149, 13th Street, Charlestown, MA 02129, USA.

*To whom correspondence should be addressed. E-mail: settleman@helix.mgh.harvard.edu

Gefitinib (Iressa) is a tyrosine kinase inhibitor that targets the epidermal growth factor receptor (EGFR), and induces dramatic clinical responses in non-small cell lung cancers (NSCLCs) with activating mutations within the EGFR kinase domain. We report that these mutant EGFRs selectively activate Akt and STAT signaling pathways, which promote cell survival, but have no effect on Erk/MAPK signaling, which induces proliferation. NSCLC cells expressing mutant EGFRs underwent extensive apoptosis following siRNA-mediated knockdown of the mutant *EGFR* or treatment with pharmacological inhibitors of Akt and STAT signaling, and were relatively resistant to apoptosis induced by conventional chemotherapeutic drugs. Thus, mutant EGFRs selectively transduce survival signals on which NSCLCs become dependent; inhibition of those signals by Gefitinib may contribute to the drug's efficacy.

EGFRs, we generated stable lines of non-transformed mouse mammary epithelial cells (NMuMg) expressing wild-type or mutant EGFRs (10), and analyzed EGF-mediated autophosphorylation of multiple tyrosine residues linked to activation of distinct downstream effectors (Fig. 1A)(1). Cell lines were generated that expressed either wild-type EGFR or one of two recurrent mutations detected in tumors from Gefitinib-responsive patients: the missense mutation L858R and the 18bp in-frame deletion, delL747-P753insS (fig. S1). Significantly different tyrosine phosphorylation patterns were observed between wild-type and the two mutant EGFRs at several C-terminal sites (Fig. 1B). EGF-induced phosphorylation of Y1045 and Y1173 was virtually indistinguishable between wild-type and mutant EGFRs, whereas phosphorylation of Y992 and Y1068 was substantially increased in both mutants. Interestingly, Y845 was highly phosphorylated in the L858R missense mutant,

“mutant EGFRs selectively transduce survival signals on which NSCLCs become dependent; inhibition of those signals by Gefitinib may contribute to the drug's efficacy.”

- Would have been nice to know before the start of clinical trials.
- Biomarkers can help select the right patients for a treatment.

Questions & Problems

Aspirin – nothing new, right?

THE WALL STREET JOURNAL
WSJ.com

HEART BEAT | FEBRUARY 23, 2010

The Danger of Daily Aspirin

By ANNA WILDE MATHEWS

If you're taking a daily aspirin for your heart, you may want to reconsider.

For years, many middle-aged people have taken the drug in hopes of reducing the chance of a heart attack or stroke. Americans bought more than 44 million packages of low-dose aspirin marketed for heart protection in the year ended September, up about 12% from 2005, according to research firm IMS Health.

Now, medical experts say some people who are taking aspirin on a regular basis should think about stopping. Public-health officials are scaling back official recommendations for the painkiller to target a narrower group of patients who are at risk of a heart attack or stroke. The concern is that aspirin's side effects, which can include bleeding ulcers, might outweigh the potential benefits when taken by many healthy or older people.

"Not everybody needs to take aspirin," says Sidney Smith, a professor at the University of North Carolina who is chairing a new National Institutes of Health effort to compile treatment recommendations on cardiovascular-disease prevention. Physicians are beginning to tailor aspirin recommendations to "groups where the benefits are especially well established," he says.

- New findings every day.
- How does this affect the use of a drug? How does it affect me?

New recommendations for cardiovascular disease prevention with Aspirin:

- slightly lower daily dose than baby aspirin
- yes for person with risk factors but no history of bleeding and ulcers; for men >45y, women >55y
- no for men <45y, women <55y, or >80y

TMO

Mission

Focuses on the development of a **high level patient-centric ontology for the pharmaceutical industry**. The ontology should enable silos in **discovery research, hypothesis management, experimental studies, compounds, formulation, drug development, market size, competitive data, population data**, etc. to be brought together. This would enable scientists to answer new questions, and to answer existing scientific questions more quickly. This will help pharmaceutical companies to model patient-centric information, which is essential for the tailoring of drugs, and for early detection of compounds that may have sub-optimal safety profiles. The ontology should **link to existing publicly available domain ontologies**.

TMO Development

Concept Identification via Use Cases

Process (bottom-up approach):

- describe users & roles
- work out use cases
- identify used concepts
- map concepts to other ontologies/vocabularies
- align with Basic Formal Ontology (BFO)
- identification of candidate domain ontologies
- refine and start over again

TMO Development

Users & Roles

Category	User / Role	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

TMO Development

Concept Identification via Use Cases

Example

(see <http://esw.w3.org/topic/HCLSIG/PharmaOntology/UseCases>):

1. Patient [OBI:0000093, patient role] (and family members [NCI:Patient_Family_Member_or_Friend]) report symptoms [IDO:0000048, Symptom] to physician/clinician [NCIt:Physician]. Physician/clinician enters reported symptoms into eHR.
 2. Physician [NCIt: Physician] makes a list of differential diagnoses, with a working diagnosis [OBI:0000075] of Alzheimer Disease [DOID:10652]. (Data Source: Physician's head).
 3. Physician [NCIt:Physician] arranges for patient [OBI:0000093, patient role] to have a basic biochemical/haematological, and SNP [SO:0000694, SNP] profile undertaken. Biochemistry, Haematology, and SNP requests are input by respective departments directly into patient's eHR [HL7:EHR, UMLS:C1555708, HID:20081] from laboratory (Data Source: eRecord). Preliminary SNP and genetic data will be submitted directly to the NIH Pharmacogenetics Research Network (PGRN).
- [...]

TMO Development

Mapping to Other Ontologies/Vocabularies

NCBO

e.g., patient role

The screenshot shows the NCBO BioPortal interface for the 'Ontology for Biomedical Investigations' (OBI). The page title is 'Ontology for Biomedical Investigations - patient role - Mozilla Firefox'. The URL is http://bioportal.bioontology.org/visualize/40832/?conceptid=obo%3AObi_0000093. The page features a navigation bar with 'BioPortal', 'Browse', 'Search', 'Projects', 'Annotate', 'All Mappings', and 'All Resources Alpha'. Below the navigation bar, there are tabs for 'NCI Thesaurus' and 'Ontology for Biomedical Investigations'. The main content area displays the 'Ontology for Biomedical Investigations Version 2009-11-06 Philly (aka version 1.0) Release Candidate' with a search bar containing 'patient role' and links for 'Link Here' and 'Subscribe'. A 'View Ontology Summary' section is visible, showing a 'Jump To:' field and a 'Legend' section with a list of roles including 'patient role'. A table on the right provides details for the 'patient role' concept.

Details	Visualization	Notes (0)	Mappings (0)	Resource Index
ID:				obo:Obi_0000093
Full Id:				http://purl.obolibrary.org/obo/Obi_0000093
Has Curation Status:				obo:IAO_0000120
Label:				patient role
Example Of Usage:				a hospitalized person; a person with controlled diabetes; the patient's role http://www.fertilityjourney.com/testingAndDiagnosis/theRightDoctor/thePatientsRole/index.asp?C=55245395146924652778
Definition Editor:				GROUP:Role Branch
Definition:				Patient is a role which inheres in a person and is realized by the process of being under the care of a physician or health care provider
Editor Preferred Term:				patient role
Definition Source:				OBI, CDISC
Disjoint With:				analyte role study group role supernatant role cloning insert role buffer role restricting MHC role

TMO Development

Mapping to Other Ontologies/Vocabularies

UMLS

e.g., eHR

The screenshot shows the 'Rich Release Format Browser 2009AB C1555708' application. The interface includes a menu bar (File, Edit, View, Options, Help), a toolbar, and a search section. The search section has a 'Refine Search by' dropdown set to 'None' and a 'Modify' button. Below this are tabs for 'Tree Browser', 'UI Search', and 'Word Search'. The 'Word Search' tab is active, showing a search input field with 'electronic health record' and a 'Search' button. Below the search input is a list of search results, with 'C1555708 electronic health record - ActClass' selected. The right pane shows the details for the selected concept, including 'Raw View' and 'Report View' tabs. The 'Report View' is active, displaying the following information:

- Concept:** [C1555708] electronic health record - ActClass
- DA Date Added 20051121
- MR Major Revision Date 20090914
- ST Status R
- Semantic Type**
 - Idea or Concept
- Definition**

HL7V3.0/PT|<p>A context that comprises all compositions The EHR is an extract that includes the entire chart.</p><p> NOTE: In an exchange scenario, an EHR is a specialization of an extract.</p>
- Atoms (2):** [AUI/RSAB/TTY]
 - ⊕ electronic health record [A8322888/HL7V3.0/PT] CODE:...
- electronic health record - ActClassContainer [A15664686,...
- Contexts (2)**
 - ⊕ HL7V3.0/PT/EHR 1 electronic health record
 - ⊕ HL7V3.0/PT/EHR 2 electronic health record
- Concept Relations (1)**
 - [R0||MTH] [C2362543](#) Electronic Health Records

TMO Development

Mapping to Other Ontologies/Vocabularies

Mapping examples:

TMO class	Classes in other ontologies
pharmaceutical product (TMO_0002)	NCIt:Finished_Pharmaceutical_Product, UMLS:C1708062
institution (TMO_0025)	ACGT:Institution, BIRNLex:2085, LNC:LP76237-4, NCIt:Institution, SNOMEDCT:385437003, UMLS:C1272753
protein (TMO_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
clinical trial (TMO_0032)	HL7V3.0:CLNTRL, MSH:D016430, NCIt:Clinical_Trial, SNOMEDCT:110465008
disease (TMO_0047)	ACGT:Disease, BIRNLex:11013, DOID:4, GRO:Disease, LNC:LP21006-9, MSH:D004194, NCIt:Disease_or_Disorder, NDFRT:C2140, OBI:0000155, UMLS:C0012634

TMO Development

Use of Other Ontologies/Vocabularies

Ontologies used in TMO:

- Basic Formal Ontology (BFO): basic structure
- Relation Ontology (RO): relations
- Information Artifact Ontology (IAO): class annotations

with mappings to:

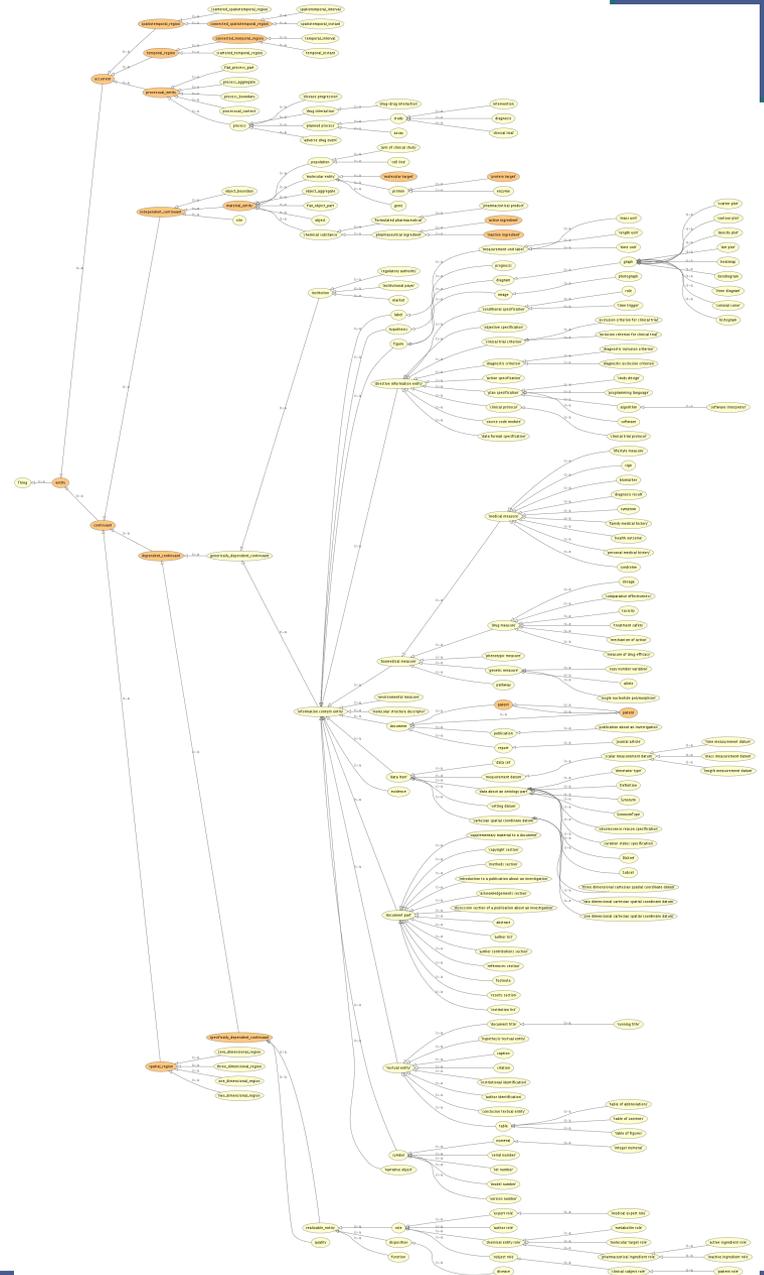
- Experimental Factor Ontology (EFO): cell line
- Ontology for Biomedical Investigations (OBI): planned process, molecular entity, metabolite
- Protein Ontology (PRO): protein
- Sequence Ontology (SO): SNP, gene, copy number variation, genotype
- ...

cell line

TMO Development Ontology Structure

<100 main TMO classes:

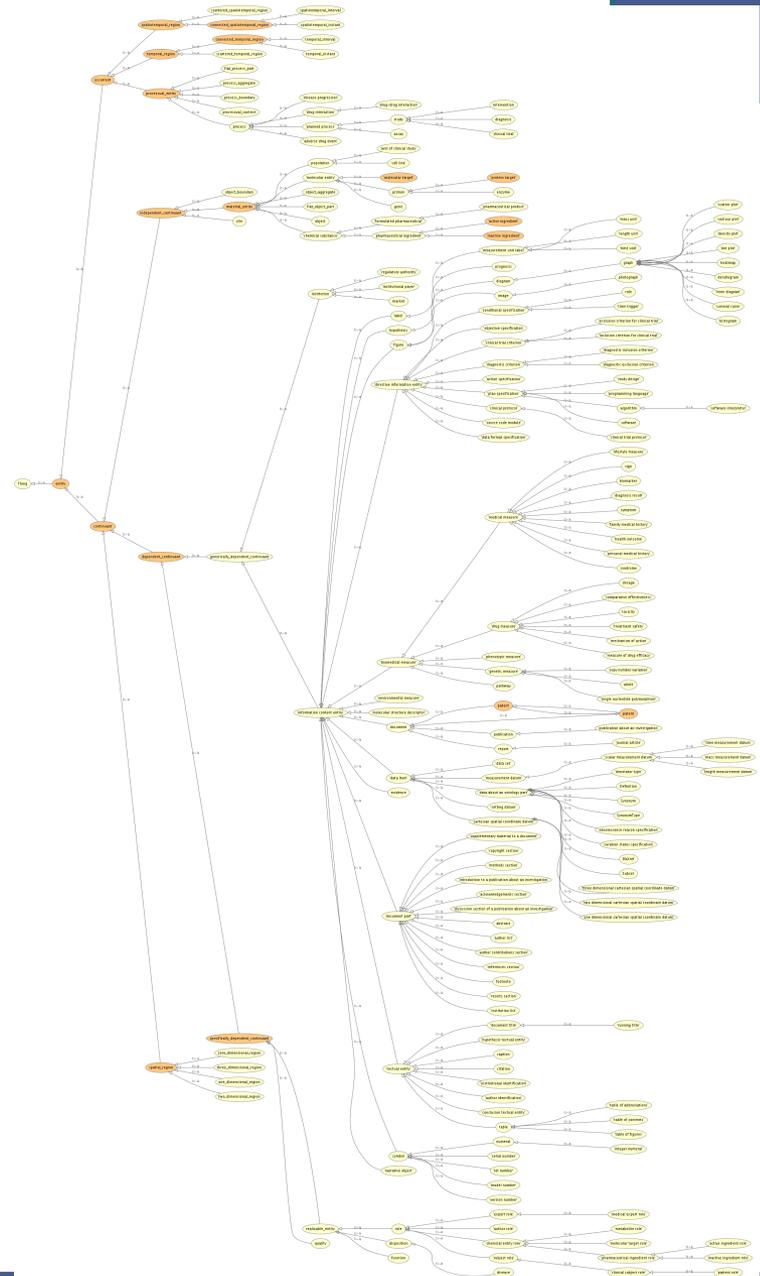
- **material entities** (e.g. molecule, protein, cell lines, pharmaceutical preparations),
- **roles** (e.g. subject, target, active ingredient),
- **processes** (e.g. diagnosis, study, intervention), and
- **informational entities** (e.g. dosage, mechanism of action, sign/symptom, family history).



TMO Development Ontology Structure

TMO also contains:

- **223 class equivalence mappings** (using owl:equivalentClass)
- from 60 TMO classes to 201 target classes in 40 other ontologies & source vocabularies (e.g. molecule, protein, cell lines, pharmaceutical preparations)



TMKB

Data Aggregation

Process:

- rdf-ize data
- load data into Virtuoso triple store
- generate mappings between data sources and TMO via
 - same IDs
 - string & semantic matching (LinQuer, SILK)

TMKB

Data Sources

Name	Topic	Short Description	Size	LODD
DailyMed	Drugs	dailymed.nlm.nih.gov provides information about approved prescription drugs, includes FDA approved labels (package inserts).	164,276 triples; 4,039 drugs	x
DBpedia	Drugs / Diseases / Proteins	RDF data about 2.49 million things that has been extracted from Wikipedia.	218M triples; 2,300 drugs; 2,200 proteins	x
Diagnostic Data	Disease / Diagnosis	AD specific diagnostic data extracted from a paper by DuBois et al (2007).		
Diseasome	Diseases / Genes	Diseasome describes characteristics of disorders and disease genes linked by known disorder–gene associations.	91,182 triples; 2,600 genes	x
DrugBank	Drugs	Drugbank.ca provides drug (i.e., chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e., sequence, structure, and pathway) information.	766,920 triples; 4,800 drugs; 2,500 protein sequences	x
LinkedCT	Clinical Trials	Linked data source of trials from ClinicalTrials.gov	7M triples; 62000 trials	x
Medicare	Medicare Formulary	List of drugs that recipients of Medicare D are eligible to receive.		x
Patient Records	Patient Data	Hand-generated test patient data, assuming data was collected within a PCHR (personally controlled health record).		
PharmGKB	Genetic Information / Drug Response	Contains information that relates genetic variation to variation in drug response.		
SIDER	Diseases / Side Effects	SIDER contains information on marketed drugs and their adverse effects.	192,515 triples; 1,737 genes	x
STITCH	Chemicals / Proteins	STITCH contains information on chemicals, proteins, and their interactions.	7,500,000 chemicals; 500,000 proteins; 370 organisms	x

TMO Use

Sample Queries ... and Answers!

Physician

_ Q: What are the diagnostic criteria for AD?

A: 12 Diagnostic inclusion criteria and 9 exclusion criteria were obtained from the criteria outlined in Dubois et al.

_ Q: Is Donepezil covered by Medicare D?

A: Yes, Medicare D covers two brand name formulations of Donepezil (Aricept and Aricept ODT).

Clinical:

_ Q: What active trials are ongoing that would be a good fit for Patient 2?

A: 58 Alzheimer trials, 2 mild cognitive impairment trials, 1 hypercholesterolaemia trial, 66 myocardial infarction trials, 46 anxiety trials, and 126 depression trials.

Discovery Research:

_ Q: What genes are associated with or implicated in AD?

A: At least 97 genes have some association with AD.

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

A: 57 compounds or classes of compounds that are used to treat 45 diseases.

TMO Use

Query & Answer – More Detail

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

A: 57 compounds or classes of compounds that are used to treat 45 diseases, including AD, hyper/hypotension, diabetes and obesity.

drug_name	disease2_name
(s)-rolipram	Schizophrenia
(s)-rolipram	Autistic Disorder
(s)-rolipram	Bipolar Disorder
(s)-rolipram	Depression
⋮	⋮
irbesartan	Hypertension
lisinopril	Hypertension
lisinopril	Diabetes Mellitus, Insulin-Dependent
nifedipine	Hypertension
perindopril	Proteinuria
perindopril	Diabetes Mellitus, Non-Insulin-Dependent
perindopril	Cerebrovascular Accident
perindopril	Cardiovascular Diseases
perindopril	Dementia
perindopril	Hypertension
perindopril	Memory Disorders
pravastatin	Coronary Arteriosclerosis

Summary

Collaboration between researchers and clinicians will be ever more important in the future to achieve the goals of personalized medicine.

TMO aims to support translational medicine by providing a terminology that facilitates the integration and analysis of disjoint data sets from basic biomedical, pharmaceutical and clinical research, and health care.

Summary

The TMO team has developed and made available:

- TMO, a candidate ontology for Translational Medicine.
- TMKB, a prototype Translational Medicine Knowledge Base containing several pharma/drug/health care relevant data sets.

An Alzheimers Disease use case demonstrates the use and usefulness of TMO in the selection of treatment and clinical trial options for a hypothetical AD patient.

The TMO project is a great example of a collaboration between industry, academia, and W3C HCLS in the pre-competitive space. Comments on and contributions to this work by the community are welcome and encouraged.

Acknowledgements

- TMO
 - Bosse Andersson, Colin Batchelor, Christopher Domarew, Michel Dumontier, Anja Jentsch, Joanne Luciano, Eric Prud'hommeaux, Susie Stephens, Patricia Whetzel, Olivier Bodenreider, Tim Clark, Lee Harland, Vipul Kashyap, Peter Kos, Julia Kozlovsky, James McGurk, Chimezie Ogbuji, Matthias Samwald, Lynn Schriml, Peter J. Tonellato, Jun Zhao
- W3C / W3C Semantic Web for Health Care and Life Sciences Interest Group / LODD

TMO

Pointers

- project home links:
 - <http://esw.w3.org/topic/HCLSIG/PharmaOntology>
- example queries:
 - <http://esw.w3.org/topic/HCLSIG/PharmaOntology/Queries>