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?

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

“mutant EGFRs selectively transduce survival signals on which NSCLCs become dependent; inhibition of those signals by Gefitinib may contribute to the drug’s efficacy.”
Questions & Problems
Aspirin – nothing new, right?

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

- New findings every day.

- How does this affect the use of a drug? How does it affect me?
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

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


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
TMO Development
Mapping to Other Ontologies/Vocabularies

UMLS

e.g., eHR
**TMO Development**

**Mapping to Other Ontologies/Vocabularies**

**Mapping examples:**

<table>
<thead>
<tr>
<th>TMO class</th>
<th>Classes in other ontologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>pharmaceutical product (TMO_0002)</td>
<td>NCI:Finished_Pharmaaceutical_Product, UMLS:C1708062</td>
</tr>
<tr>
<td>institution (TMO_0025)</td>
<td>ACGT:Institution, BIRNLex:2085, LNC:LP76237-4, NCI:Institution, SNOMEDCT:385437003, UMLS:C1272753</td>
</tr>
<tr>
<td>clinical trial (TMO_0032)</td>
<td>HL7V3.0:CLNTRL, MSH:D016430, NCI:Clinical_Trial, SNOMEDCT:110465008</td>
</tr>
<tr>
<td>disease (TMO_0047)</td>
<td>ACGT:Disease, BIRNLex:11013, DOID:4, GRO:Disease, LNC:LP21006-9, MSH:D004194, NCI:Disease_or_Disorder, NDFRT:C2140, OBI:0000155, UMLS:C0012634</td>
</tr>
</tbody>
</table>
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
- ...
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

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

<table>
<thead>
<tr>
<th>drug_name</th>
<th>disease2_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s)-rolipram</td>
<td>Schizophrenia</td>
</tr>
<tr>
<td>(s)-rolipram</td>
<td>Autistic Disorder</td>
</tr>
<tr>
<td>(s)-rolipram</td>
<td>Bipolar Disorder</td>
</tr>
<tr>
<td>(s)-rolipram</td>
<td>Depression</td>
</tr>
<tr>
<td>irbesartan</td>
<td>Hypertension</td>
</tr>
<tr>
<td>lisinopril</td>
<td>Hypertension</td>
</tr>
<tr>
<td>lisinopril</td>
<td>Diabetes Mellitus, Insulin-Dependent</td>
</tr>
<tr>
<td>nifedipine</td>
<td>Hypertension</td>
</tr>
<tr>
<td>perindopril</td>
<td>Proteinuria</td>
</tr>
<tr>
<td>perindopril</td>
<td>Diabetes Mellitus, Non-Insulin-Dependent</td>
</tr>
<tr>
<td>perindopril</td>
<td>Cerebrovascular Accident</td>
</tr>
<tr>
<td>perindopril</td>
<td>Cardiovascular Diseases</td>
</tr>
<tr>
<td>perindopril</td>
<td>Dementia</td>
</tr>
<tr>
<td>perindopril</td>
<td>Hypertension</td>
</tr>
<tr>
<td>perindopril</td>
<td>Memory Disorders</td>
</tr>
<tr>
<td>pravastatin</td>
<td>Coronary Arteriosclerosis</td>
</tr>
</tbody>
</table>
**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 Alzheimer’s 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

• 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