From the Bench to the Bedside: The Role of Semantics in enabling the Vision of Translational Medicine

Vipul Kashyap
vkashyap1@partners.org
Senior Medical Informatician,
Clinical Informatics R&D, Partners Healthcare System
Wellesley, MA, USA

BIOSTEC 2008
Funchal, Madeira, Portugal
31st January, 2008
Outline

• Vision

• Healthcare and Life Sciences
  — What is Translational Medicine?
  — Ecosystem: Current and Future State
  — Clinical Information Systems: Current State and Goal Architectures

• Translation 1: Genomic Research and Clinical Practice
  — Value Proposition of Semantic Technologies

• Translation 2: Clinical Research and Clinical Practice

• Conclusions: Semantics – A Transformational Enabler
Vision

The Next Generation Healthcare Enterprise provides services across the healthcare and life sciences (HCLS) spectrum targeted towards delivering optimum wellness, therapy and care for the patient.

These holistic services cut across biomedical research, clinical research and practice and create a need for the accelerated adoption of genomic and clinical research into clinical practice.

This is the fundamental goal and promise of Translational Medicine!
What is Translational Medicine?

- Improve communication between basic and clinical science so that more therapeutic insights may be derived from new scientific ideas - and vice versa.

- Testing of theories emerging from preclinical experimentation are tested on disease-affected human subjects.

- Information obtained from preliminary human experimentation can be used to refine our understanding of the biological principles underpinning the heterogeneity of human disease and polymorphism(s).

- [http://www.translational-medicine.com/info/about](http://www.translational-medicine.com/info/about)
What is Translational Medicine?
Healthcare and Life Sciences: Current State

Characterized by silos with uncoordinated supply chains leading to inefficiencies in the system
Future State: A Knowledge Perspective

E.g., Application of Clinical/Genomic Decision Support Rules

Knowledge Application

Knowledge Asset Management

Knowledge Discovery

E.g., Analysis of clinical Care transactions for Rules, Patient Groups, Potential Biomarkers

E.g., Creation and Maintenance of Clinical Decision Support Rules
Future State: A Business Perspective

Biomedical Research

Adaptive trials

Build models

Launch

Clinical Practice

In-life tests, surveillance and lifecycle management

Continuous marketing application

6 - 24 months

Product life

Earlier launch

Product launch

Product launch
Future State: A Stakeholder Perspective

Translation 1: Genomic Research and Clinical Practice
Translation 2: Clinical Research and Clinical Practice
Current State of Clinical Systems: Lack of “Knowledge Orientation”

• Knowledge “hardwired” into applications

• Little or no standardization on terminologies, information models or ontologies

• Current knowledge engineering tools update knowledge directly in transactional environments

• No support for versioning, provenance, change propagation.

• System implementations have inadequate knowledge to meet current workflow and quality needs

• Resources required for converting “reference” knowledge into “executable” knowledge is vastly underestimated
  — Medical Knowledge doubles every 19 years (22 months for AIDS literature)
  — Physician needs 2 million facts to practice
  — Introduction of 300+ Molecular Diagnostic tests drastically increases the number of decision support variables!
Clinical Systems: Goal Architecture

PORTALS
R&D DIAGNOSTIC SvS LABs CLINICAL TRIALS CLINICAL CARE

APPLICATIONS
LIMS EHR TRANSLATIONAL MEDICINE
ASSAYS ANNOTATIONS DIAGNOSTIC TEST RESULTS ASSAY INTERPRETATIONS ORDERS AND OBSERVATIONS

APPLICATION COMPONENTS
Genomic Analysis Order Entry/Fulfillment Patient Administration

SERVICES
Ontology Engine Services Rule Engine Services Decision Support Services Medication Services Knowledge Management Services

SERVICE DISCOVERY, COMPOSITION AND CHOREOGRAPHY

DATA AND KNOWLEDGE INTEGRATION

DATA AND KNOWLEDGE REPOSITORIES
Clinical Data Laboratory Data Metadata Repository Database of Genotypic/Phenotypic Associations Ontologies Knowledge Bases and Rule Bases
Translation 1:
Genomic Research and Clinical Practice
Use Case Flow: Need for Shared Semantics

Personalized Medicine Decision Support Services and Knowledge Repository

Knowledge Acquisition, Discovery and Management Services for Clinical Care

Test ordering and documentation guidance

Structured Test Result Interpretations

Therapeutic ordering and documentation guidance

Integrated Genotypic and Phenotypic Research Clinical Data Repository

Clinical Trials Referral

Tissue-bank

Structured Research Annotations

Bench R&D

Clinical Trials 1-4

Pharmacovigilance

Patient Encounter

R&D Discovery Services
Data Integration

Connecting Dx, Rx, Outcomes and Prognosis Data to Genotypic Data

Gene expression in HCM Test Results

<table>
<thead>
<tr>
<th>person</th>
<th>concept</th>
<th>date</th>
<th>raw value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z5937X</td>
<td>Syncope</td>
<td>3/4</td>
<td></td>
</tr>
<tr>
<td>Z5937X</td>
<td>ER visit</td>
<td>3/4</td>
<td></td>
</tr>
<tr>
<td>Z5937X</td>
<td>Palpitations</td>
<td>3/4</td>
<td></td>
</tr>
<tr>
<td>Z5937X</td>
<td>Gene-Chips</td>
<td>3/4</td>
<td></td>
</tr>
<tr>
<td>Z5937X</td>
<td>Echocardiop</td>
<td>4/6</td>
<td></td>
</tr>
<tr>
<td>Z5956X</td>
<td>Gene-Chips</td>
<td>5/2</td>
<td></td>
</tr>
<tr>
<td>Z5956X</td>
<td>Cardiomyop</td>
<td>5/2</td>
<td></td>
</tr>
<tr>
<td>Z5956X</td>
<td>Atrial Fib.</td>
<td>5/2</td>
<td></td>
</tr>
<tr>
<td>Z5956X</td>
<td>Echocardiop</td>
<td>5/2</td>
<td></td>
</tr>
<tr>
<td>Z5956X</td>
<td>EKG</td>
<td>3/9</td>
<td></td>
</tr>
<tr>
<td>Z5956X</td>
<td>Cardiac Arr.</td>
<td>3/9</td>
<td></td>
</tr>
<tr>
<td>Z5956X</td>
<td>ER Visit</td>
<td>3/9</td>
<td></td>
</tr>
<tr>
<td>Z5956X</td>
<td>Thalamus</td>
<td>3/9</td>
<td></td>
</tr>
</tbody>
</table>
Echo triggers guidance to screen for possible mutations:
- MYH7, MYBPC3, TNN2, TNNI3, TPM1, ACTC, MYL2, MYL3
Knowledge Maintenance

• Need for a Knowledge Repository to support Clinical and Genomic Decision Support

• Evolution of Biomedical Knowledge
  — Clinical Knowledge changes over time, e.g., value ranges for Clinical Normality
  — New knowledge is created, e.g., new molecular diagnostic test hits the market.

• Need for rapid knowledge change and maintenance
Data Integration: Ontology

OWL ontologies that blend knowledge from the Clinical and Genomic Domains
Data Integration: Architecture

Domain Ontologies for Translational Medicine Research

Instantiation

- Data assets remain as they are! They do not need to be modified
- The wrapper abstracts out details related to location, access and data structure
- Integration happens at the information level
- Highly configurable and incremental process
- Ability to specify declarative rules and mappings for further hypothesis generation
Bridging Clinical and Genomic Information

**Rule/Semantics-based Integration:**
- Match Nodes with same Ids
- Create new links: IF a patient’s structured test result indicates a disease
  THEN add a “suffers from link” to that disease
Bridging Clinical and Genomic Information

RDF Graphs provide a semantics-rich substrate for decision support. Can be exploited by SWRL Rules
Clinical and Genomic Decision Support

IF the patient’s LDL test result is greater than 120
AND the patient has a contraindication to Fibric Acid
THEN
    Prescribe Zetia Lipid Management Protocol

Contraindication to Fibric Acid: Clinical Definition (Old)
The patient is contraindicated for Fibric Acid if he has an allergy to Fibric Acid or has elevated Liver Panel

Contraindication to Fibric Acid: Clinical+Genomic Definition (New)
The patient is contraindicated for Fibric Acid if he has an allergy to Fibric Acid or has elevated Liver Panel or has a genetic mutation Missense: XYZ3:Ser@$#Pro

Please note: Hypothetical – assume a genetic variant is a biomarker for patients contraindicated to Fibric Acid.
Class Patient: Person
method get_name(): string;
method has_genetic_test_result(): StructuredTestResult;
method has_liver_panel_result(): LiverPanelResult;
method has_ldl_result(): real;
method has_contraindication(): set of string;
method has_mutation(): string;
method has_therapy(): set of string;
method set_therapy(string): void;
method has_allergy(): set of string;

Class StructuredTestResult
method get_patient(): Patient;
method indicates_disease(): Disease;
method identifies_mutation(): set of string;
method evidence_of_mutation(string): real;

Class LiverPanelResult
method get_patient(): Patient;
method get_ALP(): real;
method get_ALT(): real;
method get_AST(): real;
method get_Total_Bilirubin(): real;
method get_Creatinine(): real;
Flexible Clinical Decision Support

IF the_patient.has_ldl_result() > 120
AND ((the_patient.has_liver_panel_result().get_ALP() ≥ <NormalRange>
    AND the_patient.has_liver_panel_result().get_ALT() ≥ <NormalRange>
    AND the_patient.has_liver_panel_result().get_AST() ≥ <NormalRange>
    AND the_patient.has_liver_panel_result().get_Total_Bilirubin() ≥ <NormalRange>
    AND the_patient.has_liver_panel_result().get_Creatinine() ≥ <NormalRange>)
OR "Fibric Acid Allergy" ∈ the_patient.has_allergy()
OR "Missense: XYZ3:Ser@$#Pro" ∈ the_patient.has_mutation())
THEN
    the_patient.set_therapy(“Zetia Lipid Management Protocol”)

IF the_patient.has_ldl_result() > 120
AND the_patient.get_state() = “Fibric Acid Contraindication”
THEN
    the_patient.set_therapy(“Zetia Lipid Management Protocol”)
• Introduction of new component (Ontology Engine) could impact performance negatively
• Could be mitigated by co-location of Rules and Ontology Engines
• Use of load balancing servers (multiple copies of ontology engine) to improve performance
• Smart Data Caching is a critical need.
Knowledge Base Simplification and Maintenance

- Identification of “Rule Design Patterns”

- Definitions are specifications of conditions that describe:
  - Patient States, e.g., “Patient with Fibric Acid Allergy”
  - Physiological States, e.g. Normal range of LDL, Bilirubin, etc.

- Actions/Recommendations are specifications that propose therapies, medications, referrals, etc.
  - E.g., Ordering of fibrate therapies for patients with contraindication to fibric acid

- Decoupling of definitions from decisions enables greater knowledge re-use
  - Context: Abmbulatory Care, In-patient
  - Disease States
  - Indications and contraindications for interventions
  - Formularies, Resource Availability

- Enables better change management!
Knowledge Base Maintenance

- Rule simplification promotes ease of authoring
- Encapsulation of rule conditions in definitions makes rule base maintenance easier, Consider:
  IF $A$ and $B$ and $C$ and $D_1$ THEN Action$_1$
  IF $A$ and $B$ and $C$ and $D_2$ THEN Action$_2$
  ...
  IF $A$ and $B$ and $C$ and $D_K$ THEN Action$_K$

- *What if $A$ changes to $A'$?*

- Alternative:
  — Let \( \text{Definition} \equiv A \ \text{and} \ B \ \text{and} \ C \)
  — IF Definition and $D_1$ THEN Action$_1$
  — IF Definition and $D_2$ THEN Action$_2$
  — ...
  — IF Definition and $D_K$ THEN Action$_K$
Patient with Biomarker has mutation: "Missense: XYZ3:Ser@$#Pro"
Knowledge Change and Provenance
Value Proposition

• Simplification of Application Software Complexity
  — Significant chunks of functionality can be implemented using semantic web technologies such as OWL and RDF engines
  — Has impact on Software maintenance

• Simplification of Knowledge Base Complexity
  — Enables easier authoring and maintenance of the knowledge base

• Ability to incrementally integrate new data in a flexible and cost-effective manner

• Ability to handle rapid change in complex knowledge

• Common “Knowledge Standard” for Clinical Decision Support and Quality Reporting
Evaluation: Initial Experiences

Is the time and efficiency gained in authoring offsets the cost required to enhance clinical decision support performance to acceptable levels?
Translation 2: Clinical Research and Clinical Practice

http://esw.w3.org/topic/HCLS/ClinicalObservationsInteroperability
Research Coordinator selects protocol for patient screening:

Research Coordinator views list of patients and selects which ones to approach in person for evaluation and recruitment.

Clinical Research Protocol
Eligibility Criteria:
- Inclusion
- Exclusion

EMR DATA
- Meds
- Procedures
- Diagnoses
- Demographics

<table>
<thead>
<tr>
<th>Patient MR #</th>
<th>Potentially Eligible for Protocol</th>
<th># Criteria Met / Total Criteria in Protocol</th>
<th>Criteria #1 (Pass/Fail/Researcher Needs to Evaluate)</th>
<th>No Criteria #2 (Pass/Fail/Researcher Needs to Evaluate)</th>
<th>Criteria #3 (Pass/Fail/Researcher Needs to Evaluate)</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>00111111</td>
<td>Yes</td>
<td>6/8 criteria met</td>
<td>Pass</td>
<td>Pass</td>
<td>Pass</td>
<td>...</td>
</tr>
<tr>
<td>00222222</td>
<td>No</td>
<td>3/8 criteria met</td>
<td>Pass</td>
<td>Fail</td>
<td>Pass</td>
<td>...</td>
</tr>
<tr>
<td>00333333</td>
<td>Yes</td>
<td>5/8 criteria met</td>
<td>Pass</td>
<td>Pass</td>
<td>Fail</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Secondary Scenario, (Patient-Centric) Eligibility Screening

**Available protocols mapped to EMR:**

- **Research Protocol #1**
  - Eligibility Criteria:
    - Inclusion
    - Exclusion

- **Research Protocol #2**
  - Eligibility Criteria:
    - Inclusion
    - Exclusion

- **Research Protocol #3**
  - Eligibility Criteria:
    - Inclusion
    - Exclusion

- ... **Research Protocol #n**
  - Eligibility Criteria:
    - Inclusion
    - Exclusion

**EMR DATA**

- Meds
- Procedures
- Diagnoses
- Demographics

**Protocol #** | **Potentially Eligible for Protocol** | **# Criteria Met / Total Criteria in Protocol** | **Criteria #1 (Pass/Fail/ Researcher Needs to Evaluate)** | **No Criteria #2 (Pass/Fail/ Researcher Needs to Evaluate)** | **Criteria #3 (Pass/Fail/ Researcher Needs to Evaluate)** |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>6/8 criteria met</td>
<td>Pass</td>
<td>Pass</td>
<td>Pass</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>3/8 criteria met</td>
<td>Pass</td>
<td>Fail</td>
<td>Pass</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>5/8 criteria met</td>
<td>Pass</td>
<td>Pass</td>
<td>Fail</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Architecture

CTMS

Protocol Specification Interface

Eligibility Checking Module

Mapping Module

RDF Transformation Engine

EMR System
Goals: Demonstration of Re-use

- Re-use of data from the EMR for Clinical Research
- Re-use of existing vocabularies, e.g., NCI Thesaurus, Snomed, MedDRA
- Re-use of pre-existing information models e.g., HL7/RIM/DCM, CDA, SDTM
- Identify and Re-use software components that can be used to enable a wide range of use cases
  — Patient Recruitment
  — Adverse Drug Event Detection
  — Tracking a Patient through a Clinical Trial
**Model Re-use:**
**Precise, Generalized and Extensible Specification**

A Systolic Blood Pressure measurement is a pressure measurement that has a value from 0 to 220 and units are mmHg

```
SystolicBloodPressureMeasurement
equivalentClass
ClinicalElement
that key value "SnomedCodeForSystolicBP"
and magnitude only float[>= 0.0, <= 220.0]
and units value mmHg
```

A Sitting Systolic Blood Pressure Measurement is a systolic blood pressure measurement taken when a patient is sitting

```
SittingSystolicBloodPressureMeasurement
equivalentClass
SystolicBloodPressureMeasurement
that bodyPosition value Sitting
```
Shareable Open Source Models of Clinical Data

Healthcare Provider 1

Healthcare Provider 2

: : 

Healthcare Provider N

Clinical Trial 1

Clinical Trial 2

: : 

Clinical Trial M

Clinical Observations

Open Source Clinical Models
- DCM
- SDTM
- BRIDG
- Snomed
- MedDRA
- NCIT

Clinical Observations
Re-use of Industry Standards

**DCM:**
SystolicBloodPressureMeasurement
subClassOf
key value “SnomedCodeForSystolicBP”
and magnitude only float[>= 0.0, <= 220.0]
and units value HL7:PQ:mmHg

**SDTM:**
SystolicBloodPressureMeasurement
equivalentClass
VSTEST
that VSTESTCD value “NCITCodeForSYSBP”
and VSORRESU value mmHg

**Alignment/Mappings:**

DCM:SystolicBloodPressureMeasurement equivalentClass
SDTM:VSTEST that VSTESTCD value “NCITCodeForSYSBP”

DCM:key equivalentProperty SDTM:VSTESTCD
DCM:units equivalentProperty VSORRESU
DCM:magnitude equivalentProperty VSSORRES

“SnomedCodeForSystolicBP” sameAs “NCITCodeForSYSBP”
HL7:PQ:mmHg sameAs VSORRESU:mmHg
Enabling Clinical Observations Interoperability

“Mr. X”

name

Patient (id = URI1)

systolicBP Measurement1

SystolicBP

magnitude

units

key

SnomedCodeForSystolicBP

mmHg

EMR Data

120

“T1”

recording_time

“Mr. X”

name

Patient (id = URI1)

systolicBP Measurement2

SystolicBP

VSORRES

VSORRESU

VSTESTCD

mmHg

NCITCodeForSYSBP

Clinical Trials Data

130
Enabling Clinical Observations Interoperability

Patient (id = URI1)

“Mr. X”

name

SystolicBP

“mmHg”

“NCITCodeForSYSBP”

“T1”

recording_time

SystolicBP Measurement1

“T2”

recording_time

SystolicBP Measurement2

magnitude

130

120

magnitude
Precise Protocol Specification

Prophylactic Irradiation to the Contralateral Breast for BRCA Mutation Carriers Undergoing Treatment for Breast Cancer


Ages Eligible for Study: 30 Years - 90 Years, Genders Eligible for Study: Female

Inclusion Criteria:
- Female patient diagnosed with stage I-III breast cancer (AJCC 6), undergoing breast irradiation as part of her adjuvant therapy.
- The patient must be a carrier of a deleterious mutation in BRCA 1/2.
- ...

Exclusion Criteria:
- Metastatic breast cancer.
- Previous irradiation of the breast or chest wall.
- Pregnancy.
- Patients with active connective tissue diseases are excluded due to the potential risk of significant radiotherapy toxicity.
- ...
Precise Protocol Specification

Patient

that hasAge only float[>=30, <=60]
and hasGender value “Female”
and hasDiagnosis some StageI-IIIBreastCancer
and hasTherapy some BreastIrradiation
and hasMutation some (mutationType value deleterious and mutationGene value BRCA1/2)
and not (hasDiagnosis some (BreastCancer that cancerType value Metastatic))
and not (hasTherapy some Irradiation that hasLocation some (Chest or BreastWall))
and not (hasCondition some Pregnancy)
and not (hasDisease some (Disease that hasLocation some ConnectiveTissue
    and type value Active))

...
Identify Patients in the absence of explicit data about a clinical condition

<table>
<thead>
<tr>
<th>Patient Record 1</th>
<th>Patient Record 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name = Mr. X</td>
<td>Name = Mr. Y</td>
</tr>
<tr>
<td>Disease = Hematology Disorder</td>
<td>Hypereosinophilia = 2.0</td>
</tr>
<tr>
<td></td>
<td>Lymphosytosis = 6</td>
</tr>
<tr>
<td></td>
<td>Blood Lymphocytes = ATypical</td>
</tr>
</tbody>
</table>

Which of these two patients have Hematology Disorder?

Assert the following Mapping:
PatientWithHematologyDisorder
equivalentClass
Patient
that hypereosinophil only float[>1.5]
and lymphosytosis only float [> 5]
and bloodLymphocytes value ATypical

The OWL Ontology Engine will infer the rest.
Conclusion: Semantics a “Transformational” Enabler

• Data and Knowledge Integration across Biomedical Research, Clinical Research and Clinical Practice is a critical requirement for the Next Generation Healthcare Enterprise
  — The ability to capture and re-use information semantics is a critical enabler!
  — The ability to perform “as needed”, incremental and cost effective integration
  — The ability to support flexible and rapid change management
  — Increased productivity for authoring different forms of knowledge

• Rapid adoption of innovations from Genomic Research into Clinical Practice is crucial for improving patient outcomes and quality of care and is likely to transform the state and impact of healthcare
  — The rate of knowledge change will increase dramatically as Genomic Knowledge explodes!
  — Automated Semantics-based Knowledge Update and Propagation will be key in keeping the knowledge updated and current