**Assignment:** 2

**Course:** BMI 506

**Date:** March 2, 2015

**Goal:** Define knowledge sources, knowledge models, data sources, data models, and representation approaches for your project.

**Project Title:** Patients Like Mine

**Team Members:** Matt Halbert (591), Eric Holden (506), Tara Salehpour (506,591), Michelle Winerep (591)

1. Project Title: Patients Like Mine
2. Team Members: Eric Holden, Tara Salehpour
3. Brief summary of aims/goals of project and rationale for need for it

The goal of this project is to implement a tool within the framework of the PLM system that allows a domain expert to author a template for use in the generation of a database query. The templates are composed of an input template, which defines the variables, their ranges, and their relative timeframes in order to determine which patients are considered similar to the patient of interest, and an output template, which defines the variables and their relative timeframe to be extracted from the cohort of similar patients. The authoring tool will allow the domain expert to select the variables to be included in both the input and output templates from an internal knowledge base. The template will then be used in conjunction with a patient of interest's information in order to generate a database query which will then be run on the past patients database(s) to extract information from patients that are determined to be suitably similar. In practice, the query generation step will be initiated by a clinician, not a domain expert.

The PLM system as a whole is a novel concept. Although there have been instances of patient databases and extracting patient information for use in research studies, there are few that extract this information in real-time for use in a clinical setting. Certainly, no such system exists for Mayo Rochester, where PLM is being proposed. Comparing patients is typically done between large groups of patients in comparative effectiveness studies, rather than at an individual level as proposed by PLM.

1. Data Items

The major data items needed for this project are the variables used in defining the input and output templates. This information will be guided by input from Dr. Peter Li as to which variables are most relevant. This information will be stored in a tiered architecture to allow the domain expert to easily filter down to the particular item of interest. For example, selecting Heparin might follow this selection path: prescriptions ? cardiac medications ? anticoagulants ? Heparin. Some examples of categories are listed below. This listing may be subject to change based on feedback from Dr. Li.

* Demographics
* Gender
* Age
* Family history
* Medical history
* Encounters (providers, surgeons, ER)
* Date
* Provider
* Appointment type or purpose
* Outcome
* Medication: New, stop, continue
* Signs and symptoms
* Referral: type, provider or facility
* Treatment: type
* Recovery
* Diagnosis
* Status: New, change, end (recovered), continued/ chronic
* Code
* Date
* Related information: Medication, treatments, signs and symptoms
* Lab test
* Value
* Unit
* Normal range of test
* Date test administered
* Variables to be extracted from the cohort of similar patients will be selected from the same group of variables as the similarity criteria. The major difference, however, is that the domain expert will not need to define a range of similarity for these variables, just a relative timeframe.

Data sources needed

* Database
* Server, web interface

Where will you get them?

* Dr. Peter Li
* Database, EHR
* Synthetic database (Antonella Di Pace)

1. What are the forms of knowledge you will use or create

Create

Authoring Tool for Query

* Data Items
* Range
* Trends
* Limits
* Related
* Descriptive Logic Ontology: Recognition, Classification

Utilize

* Ontologies: Potential - Logical Observation Identifiers Names and Codes (LOINC), RxNorm, SNOMED, UMLS
* ICD-9, CPT
* Literature
* Domain expert knowledge, Dr. Peter Li,
* Instructor Knowledge

Where will you get the knowledge content (knowledge base?)

* Create
* Ontologies

1. How do you plan to represent/encode the data

* Semantic logic, logical data model
* Clinical Data Model: Fast Healthcare Interoperability Resources (FHIR), HL7
* Schema, UML, XML, JASON
* Web Ontology Language (OWL)

1. How do you plan to represent/encode the knowledge

* Semantic logic, logical data model
* Fuzzy logic?
* Unified Modeling Language (UML), Enterprise
* Ontologies: Potential - Logical Observation Identifiers Names and Codes (LOINC), RxNorm, SNOMED, UMLS
* Web Ontology Language (OWL), Protégé

1. Justification for your choices for the above items 4-7.

The most crucial data items are: age, gender, surgical procedures, surgical date, disease, lab test results, normal lab test ranges, lab test dates, signs and symptoms (dates), medications, health status (encounter status), encounter dates, treatments, treatment dates, family history and exclusion criteria. These are all parameters that will be used to generate a similar cohort to a patient of interest. Specific items to be determined through consult with Dr. Peter Li

Our team has decided to use the Fast Healthcare Interoperability Resources (FHIR) Clinical Information Model. The data has already been structured in FHIR, so we feel it is logical to continue to use this model. According to our stakeholder and domain expert, Dr. Peter Li, the values and types in FHIR work well with the data we will be using.

FHIR can satisfy the needs covered by all of the previous primary HL7 standards (V2, V3 and CDA). In many cases, it also provides additional benefits in terms of ease of interoperability. FHIR has a superior ability to interact with EMRs, which is a feature that we will very likely be implementing In order to query and retrieve patient data.

We also need to build a reasoning layer to classify patient data from our database. We plan to utilize Descriptive Logic Ontology using Protégé. This type of mapping will allow us to employ ontologies such as: SNOMED, UMLS, LOINC, and RxNorm in our knowledge model.

**Contributions:**

* Both, Eric and Tara contributed equally.

**Questions** for Dr. Li:

1. Which patient attributes are relevant (specific lab tests, time frames, trends, limits, exclusion criteria, diagnosis/ disease)?
2. What type of database are we using?
3. Will we need to extract information directly from the EHR?
4. How do we classify data items in the database/ EHR? (diagnosis, mediation, etc)

* Data codes used in EHR/database for disease, medication etc. (use OWL to map these codes to ontologies?)

1. Do you have PLM framework/ schema files we can use?
2. Are you using ontologies?
3. Are you using UML diagrams?