A Semantic Web-Based Framework for Quality Assurance of Electronic Medical Records Data for Secondary Use

Guoqian Jiang, Harold Solbrig, Christopher Chute
Mayo Clinic

W3C RDF Validation Workshop
September 10, 2013
Secondary use of EHR data

- The secondary use of electronic health records (EHR) sourced data is a broad domain.
- It includes patient safety and clinical quality metrics and development programs.
- The entire categories of clinical and translational research are fundamentally dependent on effective secondary use of clinical information, including:
  - clinical trials,
  - observational cohorts,
  - outcomes research,
  - comparative effectiveness,
  - and best evidence discovery.
The dimensions of data quality and methods of data quality assessment

Source from: Weiskopf NG, Weng C. J Am Med Inform Assoc. 2013
Use of Data Standards

• A central part of good data practice that improves data quality

• Meaningful Use
  • A national strategy in the US
  • A roadmap for the advancement of health information exchange
  • Using certified EHR technology to improve quality, safety, efficiency and reduce health disparities
Clinical Information Modeling Initiative (CIMI)

• an international collaboration effort,

• has been actively working on developing a common format for detailed specifications for the representation of health information content

• http://informatics.mayo.edu/CIMI/index.php/Main_Page
Standardized clinical information models for secondary use

• There is an emerging interest in the use of clinical information models in the context of secondary use of Electronic Medical Records (EMR) data

• The Office of the National Coordinator (ONC)’s Strategic Health IT Advanced Research Projects Area 4 (SHARPn) project has adopted Intermountain Healthcare clinical element models (CEMs) for normalizing patient data extracted from EMR systems

• [http://sharpn.org](http://sharpn.org)
# SHARPn Project Organization (Secondary Use of EHR Data)

<table>
<thead>
<tr>
<th>Themes</th>
<th>Projects</th>
<th>Players</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Normalization</td>
<td>Clinical Data Normalization</td>
<td>IBM, Mayo, Utah, Agilex</td>
</tr>
<tr>
<td>Phenotype Recognition</td>
<td>Natural Language Processing (NLP)</td>
<td>Harvard, Group Health, IBM, Utah, Mayo, MIT, SUNY, i2b2, UCSD, Colorado</td>
</tr>
<tr>
<td>Data Quality and Evaluation Frameworks</td>
<td>Phenotyping</td>
<td>CDISC, Centerphase, Mayo, Utah</td>
</tr>
<tr>
<td>UMIA and Scaling Capacity</td>
<td>Data Quality</td>
<td>IBM, Mayo, Mirth Corp</td>
</tr>
<tr>
<td>Evaluation Framework</td>
<td></td>
<td>Agilex, MN HIE, Mayo, Utah, Mirth Corp</td>
</tr>
</tbody>
</table>
Clinical Element Models

• Developed by the Intermountain Healthcare
• A model for describing and representing detailed clinical information through defining a standard data structure
• Presently, over 4,000 CEMs, such as blood pressure measurement or specific laboratory tests, are defined.
• Authored in the Constrain Definition Language (CDL)
• Distributed in XML schema (XSD)
System Architecture

Semantic Web Layer

Current Normalization Pipeline

SPARQL Inference Notation (SPIN) Framework

- CEMs in OWL
- Instance Data in RDF
- CEMs in XSD
- Instance Data in XML
- CEMs in CDL
- Patient Data in EMRs
SecondaryUseNotedDrug (SHARP)
SecondaryUseAllergyIntolerance (SHARP)
SecondaryUseAllergyIntoleranceConcern (SHARP)
SecondaryUseAssertion (SHARP)
SecondaryUseAssessmentCoded (SHARP)
SecondaryUseAssessmentOrdinary (SHARP)
SecondaryUseAssessmentQuantitative (SHARP)
SecondaryUseDiseaseDisorder (SHARP)
SecondaryUseNotedDrug (SHARP)
SecondaryUseNotedDrugList (SHARP)
SecondaryUsePatient (SHARP)
SecondaryUseProblemConcern (SHARP)
SecondaryUseProcedure (SHARP)
SecondaryUseStandardLabObsCoded (SHARP)
SecondaryUseStandardLabObsIntervalQuantitative (SHARP)
SecondaryUseStandardLabObsNarrative (SHARP)
SecondaryUseStandardLabObsOrdinary (SHARP)
SecondaryUseStandardLabObsQuantitative (SHARP)
SecondaryUseStandardLabObsTiter (SHARP)
SecondaryUseStandardLabPanel (SHARP)
SecondaryUseVitalSignPanel (SHARP)
CEMs in XSD -> OWL
Instance Data in XML -> RDF

```xml
<?xml version="1.0" encoding="UTF-8"?>
<secondaryUseNotedDrug class="statement" xsi:noNamespaceSchemaLocation="SecondaryUseNotedDrug.xsd">
  <key type="CD" class="data">
    <isNull type="xs:boolean">false</isNull>
    + <code type="CODE" class="data"></code>
    + <codeSystem type="CS" class="data"></codeSystem>
    + <codingRationale type="CODE" class="data"></codingRationale>
    <originalText>NOTEDDRUG</originalText>
  </key>
  <CD type="CD" class="data">
    <isNull type="xs:boolean">false</isNull>
    <code type="CODE" class="data">
      "OXYCONTIN"
    </code>
  </CD>
</secondaryUseNotedDrug>
```

```xml
a cem:SecondaryUseNotedDrug ;
  cem:CD [ a cem:CD ;
    cem:class "data" ;
    cem:code [ a cem:CODE ;
      cem:class "data" ;
      cem:type "CODE" ;
      cem:value [ rdf:value "OXYCONTIN" ;
        cem:atomicType "xs:string"
      ] ] ;
    cem:codeSystem [ a cem:CS ;
      cem:class "data" ;
      cem:type "CS" ;
    ] ] ;
```
Check constraints and perform data validation

Define SPIN constraints
Check whether the units are identical

cem:SecondaryUseNotedDrug

```sparql
# the units of takenDoseLowerLimit and takenDoseUpperLimit are not identical
ASK WHERE {
  ?s cem:takenDoseLowerLimit ?b .
  ?c cem:unit ?d .
  ?d cem:value ?e .
  ?e rdf:value ?lowerUnitValue .
  ?e1 rdf:value ?upperUnitValue .
  FILTER (?lowerUnitValue != ?upperUnitValue) .
}
```
Check whether lowerLimit is greater than upperLimit
cem:SecondaryUseNotedDrug

```sparql
# the value of takenDoseLowerLimit is greater than the value of takenDoseUpperLimit
ASK WHERE {
    ?s cem:takenDoseLowerLimit ?b .
    ?c cem:unit ?d .
    ?d cem:value ?e .
    ?e rdf:value ?lowerUnitValue .
    ?e1 rdf:value ?upperUnitValue .
    FILTER ((?lowerUnitValue = ?upperUnitValue) && (?lowerLimitValue > ?upperLimitValue)) .
}
```
Constraint Violation Warning

When
(takenDoseLowerLimit.PQ.unit.value "TAB" =
takenDoseUpperLimit.PQ.unit.value "TAB") and
(takenDoseLowerLimit.PQ.value "3" >
takeDoseUpperLimit.PQ.value "2")

Warnings for <http://informatics.mayo.edu/sharp/images/d...>

the value of takenDoseLowerLimit is
greater than the value of
takenDoseUpperLimit
(SPIN constraint at
cem:SecondaryUseNotedDrug)
Standardized Semantic Web Services

- We develop a CIMI-compliant semantic web representation and services for clinical information models
  - Convert the CIMI reference model in the UML to OWL
  - Generate domain templates from the BRIDG model (which is a domain analysis model in the domain of clinical research)
The Parts of the CIMI Reference Model

• Core Model
• Party Model – with HL7 V3 RIM mappings
• Data Value Types
• Primitive Types
CIMI Reference Model – Data Values
CIMI Reference Model in OWL

http://informatics.mayo.edu/ontologies/cimi/cimi-rm-1.0.12b.owl
Domain Templates in CIMI-compliant Representation Using Semantic Web
Clinical Study Meta-Data Standards Development Applications

BRIDG Browser
Template Generator
Template Mind Map View

BRIDG Template Generation and Visualization User Interface (SmartGWT Widgets)

SPARQL Query Services

Integrated BRIDG Model Artifacts in a RDF Store

Normalization Pipeline

BRIDG Model in OWL
ISO 21090 Datatypes In OWL
bridg:AdverseEventSeriousnessTemplate
  a owl:Class, cimi:ENTRY ;
  rdfs:label "AdverseEventSeriousnessTemplate" ;
  <http://opencimi.org/rm#CLUSTER.item>
    a cimi:CLUSTER ;
    rdfs:comment "(describes) describedAdverseEvent (1-1): AdverseEvent"^^xsd:string ;
    rdfs:label "AdverseEvent"^^xsd:string ;
    cimi:constrains bridg:AdverseEvent
    ] ;
  <http://opencimi.org/rm#CLUSTER.item>
    a cimi:CLUSTER , spl:attribute ;
    rdfs:comment "AdverseEventSeriousness"^^xsd:string ;
    rdfs:label "AdverseEventSeriousness"^^xsd:string ;
    <http://opencimi.org/rm#CLUSTER.item>
      a cimi:ELEMENT , spl:Attribute ;
      rdfs:comment "date TS.DATETIME (0-1)"^^xsd:string ;
      rdfs:label "date"^^xsd:string ;
      spl:maxCount 1 ;
      spl:minCount 0 ;
      spl:predicates <http://www.bridgemodel.org/owlt#AdverseEventSeriousness.date> ;
      spl:valueType <http://www.hl7.org/iso-dt-2.0#TS.DATETIME>
    ] ;
  <http://opencimi.org/rm#CLUSTER.item>
    a cimi:CLUSTER , spl:attribute ;
    rdfs:comment "describes AdverseEvent(1-1)"^^xsd:string ;
    rdfs:label "describes"^^xsd:string ;
    spl:maxCount 1 ;
    spl:minCount 1 ;
    spl:predicates <http://www.bridgemodel.org/owlt#AdverseEventSeriousness.describesAdverseEvent> ;
    spl:valueType bridg:AdverseEvent
  ] ;
  <http://opencimi.org/rm#CLUSTER.item>
    a cimi:CLUSTER , spl:attribute ;
    rdfs:comment "code CD(0-1)"^^xsd:string ;
    rdfs:label "code"^^xsd:string ;
    spl:maxCount 1 ;
    spl:minCount 0 ;
    spl:predicates <http://www.bridgemodel.org/owlt#AdverseEventSeriousness.code> ;
    spl:valueType dt:CD
  ] ;
  cimi:constrains bridg:AdverseEventSeriousness
] .
Discussion

• RDF validation mechanism against CIMI models
  • Model level (CIMI models)
  • Instance level (Patient data)

• Challenging issues
  • Data types
  • Value set binding

• XML Semantics Reuse technology
  • XSD -> OWL
  • XML -> RDF
BRIDGing Technology?

XML Rules
XSD/XML

Object Constraint Language (OCL)
UML

SWRL/SPARQL Rules
RDF/OWL

XML Rules
XSD/XML

Object Constraint Language (OCL)
UML

SWRL/SPARQL Rules
RDF/OWL
Questions & Discussion