Methodology for Standards-Based Biomedical and Healthcare Data Instance Generation

John T.E. Timm
IBM Almaden Research Center
San Jose, CA

Ariel Farkash
IBM Haifa Research Lab
Haifa, Israel

Sondra Renly
IBM Almaden Research Center
San Jose, CA
Two different types of user roles:

- **Clinical Domain Expert / Researcher** – cares about local ontology, mappings to common ontology and standard terminologies
- **Healthcare IT Expert** – cares about healthcare interoperability standards, software modeling technologies (e.g. UML and OCL)

Clinical domain experts / researchers are integrating data using **semantic web** technologies (e.g. RDF, OWL)

Would like to **contribute data to / use data from** larger communities and information networks such as the Nationwide Healthcare Information Network (NHIN)

Healthcare data on such networks is **exchanged** using **XML-based interoperability standards** such as the HL7 Clinical Document Architecture (CDA)
Goals

- Provide a methodology that enables **semantic interoperability** and data sharing across these two distinct user communities
  - Provide mechanism for healthcare data standard instance generation
  - Leverage strengths of several different standards and technologies
  - Capture similarities while preserving disparities using template mechanism
Data Integration using Semantic Web Technologies

HYPERGENES
European Network for Genetic-Epidemiological Studies

- FP7 funded Integrated Project (IP), 19 European partners
- Genetic-epidemiological study of essential hypertension
- Key Components
  - Heterogeneous data sources (cohorts)
  - Data extraction from proprietary format
  - RDF-based data container
  - Ontology-based harmonization
  - Normalization and validation of data
Data Extraction to RDF (RDFizers)

- Data may come in various formats and schemas
- Some of the more common: MySQL, Excel / CSV, Plain Text
- Need to converge to a generic format: RDF data container using cohort ontology
Harmonization and Normalization

- Each cohort has its own local ontology and may refer to different terminologies
- Variables may be focused or very vague and may differ in units, methods of measurement and timing
- Need mapping to a core ontology with standard terminology and normalized units
IBM Research

Semantic Data Integration

Clinical Domain Expert/Researcher

Cohort Ontologies

Standard Terminologies

Mapping

Core Ontology (OWL)

conforms to

RDF Data Graph

Clinical Data Producers/Consumers
HL7 Clinical Document Architecture (CDA)
Implementation Guides

- A text document that specifies additional constraints on a CDA document for a particular domain or use case
- Examples:
  - Continuity of Care Document (CCD)
  - Public Health Case Report (PHCR)
  - Essential Hypertension Summary Report (EHCDA)
- Constraints/conformance rules are grouped into templates

Templates

- Every template has a unique identifier, thus it may be reused
- Templates constrain specific classes in the CDA model
- Examples:
  - Discharge Summary is a document template
  - Allergy is a clinical statement template (e.g., hives as reaction to penicillin)
Modeling Healthcare Interoperability Standards

Open Health Tools

- Model-Driven Health Tools (MDHT)
  - Joint open-source project between IBM and VHA
  - Build Eclipse-based healthcare tools based on software industry standard modeling languages (e.g. UML/OCL)

- CDA Tools subproject
  - Provide a methodology and tooling for the design and implementation of CDA and templates using widely adopted modeling languages and paradigms
Template model represents domain-specific specialization of base CDA model

- Templates modeled as classes that subclass classes in CDA model (or other templates)
- Constraints on template attributes modeled using property redefinition
- Constraints between templates modeled using directed associations
- CDA UML Profile created to capture additional template/constraint related metadata

A model-to-model transformation is used on the template model to produce OCL constraints used for validation at runtime
PHCR Tuberculosis Template Model (Excerpt)

- Property redefinitions
- Directed associations
- Inheritance
MDHT CDA Tools

Implementaiton Guide (PDF)

Domain Template Model (UML+OCL)

Healthcare IT Expert

Transformation

Implementation Model (UML+OCL)

Transformation

Code Generation Model (EMF)

Runtime API (Java)

Produces

CDA Instance (XML)
How do we bring these together?

Implementation Guide (PDF) — Healthcare IT Expert

Domain Template Model (UML+OCL)

Transformation

Implementation Model (UML+OCL)

Transformation

Code Generation Model (EMF)

Code Generation

Runtime API (Java)

Produces

Clinical Domain Expert/Researcher

Cohort Ontologies

Mapping

Conforms to

Standard Terminologies

Core Ontology (OWL)

RDF Data Graph

CDA Instance (XML)

Clinical Data Producers/Consumers
IBM Research

Integrated Workflow

- Implementation Guide (PDF)
- Healthcare IT Expert
- Domain Template Model (UML+OCL)
- Annotated Domain Template Model (UML+OCL)
- Implementation Model (UML+OCL)
- Code Generation Model (EMF)
- Transformation
- {variable, path} mapping
- Code Generation
- Clinical Domain Expert/Researcher
- Clinical Data Producers/Consumers
- Cohort Ontologies
- Standard Terminologies
- Core Ontology Model (UML+ODM)
- Instance Generation Engine (Java)
- Runtime API (Java)
- RDF Data Graph
- CDA Instance (XML)
- Import
- Conforms to
- Uses
- Produces

- Annotation
- Transformation
- Annotation
- Mapping

- Cohort Ontologies
- Standard Terminologies

- Healthcare IT Expert
- Clinical Domain Expert/Researcher
- Clinical Data Producers/Consumers

- Import
- Conforms to
- Uses
- Produces
Core Ontology Model

- Options for creating core ontology model
  - Create core ontology using ontology editor
    - Using an ontology editor (e.g. Protégé)
    - Programmatically using library (e.g. Jena API)
    - Import ontology into UML using ODM mapping rules (work-in-progress)
  - Create core ontology using UML editor
    - Use ODM UML Profile to specify ontology related metadata
    - Generate OWL representation
  - Apply ODM UML Profile to template model directly
    - Generate OWL representation
Annotated Template Model

- Created UML profile for adding annotations to template model
- Profile captures mappings between concepts (variables) from the core ontology and paths in the template model
- Paths are relative to template class on which they are applied
- Traverse template model to produce set of absolute paths
  - Directed associations between templates are expanded (Section Template -> Observation Template) => (Section -> Entry -> Observation)
  - Paths are decorated with specialized type (prefix:TypeName(comp))
  - Create {variable name => absolute path} map @treatmentGiven
Runtime API is produced from EMF model
  - Customized code generation templates

Input:
  - Core ontology
  - RDF data graph that conforms to Core Ontology
  - Variable to path mappings from annotated template model

Process:
  - Extract variable names and data values from datatype properties in RDF data graph using Jena API
  - Look up correct path using variable name
    - \{variable name => data value\} -> \{absolute path => data value\}
  - Path is used to create EMF object graph

Output:
  - Standard CDA XML instance that conforms to template model
Integrated Workflow

Implementation Guide (PDF)

Domain Template Model (UML+OCL)

Annotated Domain Template Model (UML+OCL)

Implementation Model (UML+OCL)

Code Generation Model (EMF)

Core Ontology Model (UML+ODM*)

Instance Generation Engine (Java)

Runtime API (Java)

Cohort Ontologies

Standard Terminologies

Clinical Domain Expert/Researcher

Healthcare IT Expert

Transformation

{variable, path} mapping

uses

produces

Clinical Data Producers/Consumers

import

conforms to

Annotation

Mapping

Code Generation

CDA Instance (XML)

RDF Data Graph

Core Ontology (OWL)

Core Ontology Model (UML+ODM*)

Transformation

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation

{variable, path} mapping

Annotation
Summary and Future Work

- **Summary**
  - Method for integrating data from different sources using semantic technologies (e.g. OWL, RDF)
  - Method for creating standard instances used for exchange, analysis, etc. using standard modeling technologies (e.g. UML)
  - Working to bring together these approaches through annotations and instance generation engine

- **Work left to do**
  - Leverage Ontology Definition Metamodel (ODM)
  - Annotation Model
    - Template model should be reusable
    - Create separate annotation model that wraps template model
  - Reverse direction – Standard Instance -> RDF/OWL
  - Test usability of approach/tooling with both clinical domain experts/analysts and healthcare IT experts in pilot engagements
The End

Thank You 😊