Large Scale Healthcare Data Integration using the Semantic Web

John T.E. Timm (presenter)
Ariel Farkash
Sondra R. Renly

Oral Presentation at MIE 2011
August 30th, 2011
Oslo, Norway
Objectives

- Create a platform that capitalizes on the strengths of multiple technologies / representations
  - Semantic web formats for integration
  - Standard healthcare information models for semantics and XML exchange capabilities
  - UML + OCL for tooling / information model design

- System must support multiple roles / expertise
  - Clinical domain expert
  - Health IT standards expert

- Leverage existing healthcare information models available in open source
Previous Work – Hypergenes

- **Cohort ontology**
  - OWL ontology that represents clinical data at the data source

- **Core ontology**
  - OWL ontology that represents clinical data for the essential hypertension domain

- **Mapping between cohort and core ontologies**
  - Convert RDF conforming to cohort ontology into RDF conforming to core ontology
  - Normalization / harmonization – unit conversion, terminology

- **Instance generation**
  - Convert RDF conforming to core ontology into XML instances that conform to CDA standard (using templates)
Previous Work – Hypergenres (cont’d)

- We successfully used semantic web technologies to integrate multiple cohorts
- We now have a repository of CDA documents
- CDA XML instances are especially challenging to query
- Used ad hoc XQuery => RDF solution to export data to relational data mart
- How can we query the data while it is still in semantic web format but also leverage the richness of templated CDA information models and make CDA XML instances available on-demand?
Clinical Domain Expert/Researcher

- Cohort Ontologies
- Standard Terminologies

mapping

Core Ontology (OWL)

conforms to

Data Container (RDF)

Clinical Data Provider

Cohort Data

Clinical Data Consumers

Core Ontology (OWL)

generate

Triple Store (RDF)

Instance generation

Standard CDA instances (XML)

IHE XDS / QED interface

XQUERY

Clinical Data Consumers
Model-Driven Health Tools (MDHT)

- Open Source Project built on the Eclipse platform
- Design-time tooling for standards developers to create Implementation Guides (CDA templates) for the HL7 Clinical Document Architecture (CDA) using UML and OCL formalisms
- Uses model-driven approach to generate documentation and reference implementation (Java APIs) for applications, testing and conformance
- Library of existing CDA template models (CCD, IHE PCC, HITSP)
Clinical Domain Expert/Researcher

- Cohort Ontologies
- Standard Terminologies

Core Ontology (OWL)

Data Container (RDF)

Clinical Data Provider

- Cohort Data

Clinical Data Consumers

Clinical Domain Expert

- Healthcare IT Expert
- Template Model (UML/OCL)

CDA Model (UML)

HL7v3 Datatypes (UML)

HL7v3 Vocabulary (UML)

Healthcare IT Expert

Triple Store (RDF)

Instance generation

Standard CDA instances (XML)

IHE XDS / QED interface

Clinical Data Consumers

Mapping

Conforms to

Generate

SPARQL

XQUERY
Methods

- Determine mappings between UML + OCL formalisms and OWL based on existing best practices and standards such as the OMG Ontology Definition Metamodel (ODM)
  - UML class -> OWL class
  - UML property -> OWL datatype property
  - UML relationship -> OWL object property
  - UML enumeration -> OWL one of

- Determine how to represent CDA template models into OWL using restrictions

- Create model transformations to convert standard information models (CDA) to OWL
Clinical Domain
Expert/Researcher

- Cohort Ontologies
- Standard Terminologies

Healthcare IT
Expert

- Template Model (UML/OCL)
- Uses

Core Ontology (OWL)

CDA Model (UML)

- HL7v3 Datatypes (UML)
- HL7v3 Vocabulary (UML)

Data Container (RDF)

- Conforms to
- Uses

CDA instances (RDF)

Standard CDA instances (XML)

Clinical Data
Provider

Cohort Data

Clinical Data Consumers

Clinical Domain Expert/Researcher

- Cohort Ontologies
- Standard Terminologies

Healthcare IT Expert

- Template Model (UML/OCL)
- Uses

Core Ontology (OWL)

CDA Model (UML)

- HL7v3 Datatypes (UML)
- HL7v3 Vocabulary (UML)

Data Container (RDF)

- Conforms to
- Uses

CDA instances (RDF)

Standard CDA instances (XML)

Clinical Data Consumers
Conclusion and Future Work

- We were able to query semantic data using SPARQL and information models based on CDA
- We eliminated the need for querying directly against XML instances to export to data mart
- Still provided the ability to produce CDA XML instances on-demand
- How do we represent complex template constraints? SWRL?
- Develop query builder tool using SPARQL query templates
Questions