Jena based Implementation of a ISO 11179 Metadata Registry

Gokce B. Laleci & A. Anil Sinaci
Agenda

- **Motivation**
  - A brief overview of SALUS Project
  - SALUS Semantic Interoperability approach
  - Role of CDE Repository in this approach

- **Background**
  - ISO/IEC 11179
  - Common Data Elements

- **Design & Implementation**
  - Why a semantic repository
  - Current Implementation Status

- **Future Work**

- A STREP funded under Objective ICT-2011.5.3b) Tools and environments enabling the re-use of electronic health records which aims to
  - Enable effective integration and utilization of electronic health record (EHR) data to improve post-market safety activities on a proactive basis
    - Build the necessary interoperability architecture for enabling ADE detection tools, signal validation and strengthening processes and real time screening of multiple, distributed, heterogeneous EHRs for early detection of adverse event signals
  - Enable semantic interoperability for reuse of EHRs in drug safety research
  - Build novel framework for open-ended temporal pattern discovery on top of the electronic health records
  - Ensure security and privacy
- Pilots in Lombardia Region (Italy) and Eastern Saxony (Germany)

**Partners**
- SRDC Ltd, Turkey (coordinator)
- EUROREC, France
- WHO- UMC, Sweden
- OFFIS, Germany
- AGFA Healthcare, Belgium
- ERS, Netherlands
- LISPA, Italy
- INSERM, France
- TUD, Germany
- ROCHE, Switzerland

29/11/2012 @ SALUS 2012
How SALUS extends current spontaneous reporting system to seamlessly exploit the already existing clinical data at EHRs

An ideal system for ADR surveillance would combine the strengths of case reports with those of EHRs
How SALUS enables exploratory/confirmatory signal detection and epidemiological research studies on top of heterogeneous EHRs

In order to realize nearly real time proactive post market safety studies, there needs to be a mechanism for screening the available heterogeneous and disparate EHRs for a specified time period for adverse event signal detection and also for conducting observational studies for validation of the suspected signals and for carrying out outcome research to see long term effects of drugs.
Selected Use Cases

- Enabling Semi-automatic Notification of Suspected ADEs and Reporting ADEs within a Hospital
  - Enabling Notification of Suspected ADEs
  - Enabling Semi-automatic ADE Reporting
- Supporting Clinical Evaluation of a Potential Signal through Accessing the EHRs
  - Characterizing the cases and contrasting them to a background population
  - Temporal pattern characterization
- Running Exploratory Analysis Studies over EHRs for Signal Detection
  - Temporal association screening on EHRs
  - Manual clinical review of relevant medical history
- Using EHRs as secondary use data sources for Post Marketing safety studies
  - Estimate incidence rates of CHF in diabetic patients with a recent acute coronary syndrome (ACS) event on different diabetic medications
Problem of *Interoperability*

- **Syntactic vs. Semantic**
  - The ability to exchange information
    - *access*
  - The ability to use the information once it has been exchanged
    - *understand*
SALUS Interoperability approach

- Enable Functional Interoperability among EHR Sources and Clinical Research Systems and Tools
  - To enable them to query and subscribe parts of medical summaries of selected patient population
    - To query EHRs for ADE identification, ADE reporting and signal follow-up studies and to subscribe clinical data for a selected cohort of patient for signal detection and outcome research over distributed EHRs
  - Transaction definitions!
  - Content agnostic

- Enable Semantic Interoperability among EHR Sources and Clinical Research Systems and Tools
  - To enable them to automatically interpret the queries and the resulting clinical data exchanged meaningfully and accurately in order to produce useful results
Semantic Interoperability Challenge

- Basically stems from usage of
  - Different Reference Information Models
  - Different Templates
  - Different Coding Systems

- Examples
  - Different models to represent the query through (Patients having experienced Myocardial Infarction and who has recently used Nifedipine)
    - Precondition element in an HL7 study design message (criterion coded with ASPIRE Data Set)
    - Precondition element in an HL7 study design message (criterion coded with codes from different terminology systems like SNOMED CT, MedDRA)
    - Through Population Criteria representation in Health Quality Measures Format (HQMF)
    - Through proprietary information model to define eligibility criteria
  - Different models to represent Result Set
    - OMOP CDM (an information model and selected set of terminologies)
    - CDISC ODM annotated with CDASH
    - HL7 CDA templates
    - EN 13606 EHR Extract and archetypes
Proposed Semantic Mediation Approach

SALUS Semantic Interoperability Platform aims to build a semantic architecture, where these will be mediated to one another through a common ontology supported through semantic resource sets (domain ontologies, terminology systems).

How to build the Common Ontology?

SALUS Semantic Interoperability Platform aims to build a semantic architecture, where these will be mediated to one another through a common ontology supported through semantic resource sets (domain ontologies, terminology systems).
Building Common Ontology as a CDE Set

- We aim to create the Common Ontology as a set of meaningful fragments:
  - **Objective:** Identify a core set of common data elements (CDE) as meaningful EHR fragments that needs to be exchanged within the scope of post market safety studies
  - Step 1: Build and maintain a CDE repository in conformance to ISO/IEC 11179 standard for metadata registries
    - Provide necessary tools to create, select, adapt and manage the CDEs
  - Step 2: Load available common models to CDE Repository such as BRIDG (in future CDISC Share elements?)
    - Note that we may not be directly using BRIDG in SALUS as it is for regulated clinical research studies
  - Step 3: Identify possible source and target content models to represent the required data
    - Input: Data requirements of the selected SALUS Use cases
    - Define these as HL7 CCD templates, 13606 archetypes, OMOP CDM, CDISC ODM
  - Step 4: Load these Content Models to CDE Repository:
    - Analyse the content models, extract candidate CDEs, try to map them with existing CDEs, when necessary curate new CDEs.
    - Annotate the Content models with these CDEs
  - Step 5: Export the selected subset of CDEs as common ontology
  - Step 6: Using intelligent heuristics to semi-automatically define semantic mediation rules between content models (annotated with CDEs) and Common Ontology

- This evolving ontology will act as a common semantic dictionary of the clinical terms to be exchanged between EHR Systems and clinical research systems
- In this way we aim to achieve **semantic interoperability of meaningful fragments**
Summary: Relation between Task 4.1-4.2-4.3-4.4

2. Produces content models as HL7 CDA templates, EN 13606 templates, Archetypes, etc. (SIAMS?)
   -- This is based on the selected Pilot Scenarios--

Task 4.1
Content Model Creation

Content Models

Task 4.2
CDE Repository

3. Content Models are imported to CDE Repository as “Content Model Ontologies”
4. “Content Model Ontologies” are annotated with Terminology Codes
5. Matching CDEs are found, “Content Model Ontologies” annotated with these CDEs
   If necessary new CDEs are created

Task 4.3-4.4
Semantic Mediation Framework

6. SALUS Common Ontology is created from CDEs
7. Annotated Content Model Ontologies are loaded
8. Using Annotated Content Model Ontologies, “Mapping Definitions” between these and SALUS Common Ontology is created → A Step to create Harmonized Ontology
9. Using “Mapping Definitions”, Clinical Content represented in Content Models can be semantically mediated to/from SALUS Common Ontology

1. Imports existing Domain Models as a basis, identifies and maintains CDEs as an ontology conforming to ISO-IEC 11179 Repository Metamodel
   CDEs are annotated with Terminology Codes

CDISC
SHARE
BRIDG
DAM
OMOP
CDM
CDISC
SHARE
Meta-data for Semantic Interoperability

- Precise knowledge about how data is structured
- More efficient and productive with a central, well-administered place to seek for meta-data
  - Central, easily consumable
- Classifications with well-known terminology systems
- Build (or map) data models based on a common meta-model
Identified Requirements for SALUS CDE Repository

- Browse/View CDEs
- Search CDEs
- Annotate the Content Model with the CDEs in SALUS CDE Set
- Import content Model to the CDE Repository
- Import an XSD Schema to the CDE Repository
- Import a UML Model to the CDE Repository
- Import a local Domain Ontology to the CDE Repository

CDE Repository User

Domain Expert

Semantic Mediation Framework Knowledge Base
Example Use Cases (from SALUS)

- Medical Summary is provided in HL7 CCD based template
- Safety Analysis Tools requests the data in OMOP CDM
  - Or, can query the data through CDASH datasets
- Assume that we have initially loaded BRIDG to CDE Repository
Example Use Cases (from SALUS)

- CDE Repository now contains some CDEs extracted from BRIDG UML Model:

- Example: `PerformedMedicalConditionResult` in Study Conduct sub domain
  - `PerformedMedicalConditionResult.OccurrenceDateRange.IVL_TS`
  - `PerformedMedicalConditionResult.value.CD`
  - `PerformedMedicalConditionResult.medicalHistoryIndicator.BL`
  - `PerformedMedicalConditionResult.severityCode.CD`
Example Use Cases (from SALUS)

- CCD Section template for representing “Past Medical History is loaded.

- Identified Candidate CDEs:
  - PastMedicalHistorySection.entry.Act
  - Observation.code.CD
  - Observation.value.CD
  - Observation.effectiveDate.IVL_TS
  - Observation.problemStatus.Observation
  - Observation.severity.Observation
  - Observation.healthStatus.Observation
Map Candidate CDEs with available CDEs

- **CDEs available**

  - ObservedProperty
  - Observation.value.CD
  - Observation.effectiveDate.IVL_TS
  - Observation.problemStatus.Observation
  - Observation.severity.Observation
  - Observation.healthStatus.Observation

- **Identified Candidate CDEs:**

  - Observation

We are planning both:

**a manual mapping process through the graphical interfaces to be provided,**

**a semiautomatic mapping process supported with semiautomatic annotation with terminology codes and reasoning:**

This is why we foresee a semantic repository

No mapping CDE, a new CDE can be curated and added to CDE Rep.
Example Use Cases (from SALUS)

- CCD Section template annotated with available CDEs.

```xml
<section code="11348-0">
  <entry act="statusCode">
    <observation (moodCode=EVN)>
      <code codeSystemName="codeSystemName" display="displayName">
        <value code="55607006">
          <code codeSystemName="codeSystemName" display="displayName">
            <effectiveDate low="low" high="high">
              <entryRelationship>
                <observation>
                  <code codeSystemName="codeSystemName" display="displayName">
                    <value code="33999-4">
                      <code codeSystemName="codeSystemName" display="displayName">
                        <PerformedMedicalCondition Result.value.CD>
                          <PerformedMedicalCondition Result.OccurrenceDateRange.IVL_TS>
                            <PerformedMedicalCondition Result.medicalHistoryIndicator.BL>
                              <29/11/2012>
                        @ SALUS 2012
```
Adding OMOP CDM Model and annotating it with available CDEs

- Identified Candidate CDEs:
  - ConditionOccurrence.condition_concept.
  - ConditionOccurrence.startDate.Date
  - ConditionOccurrence.endDate.Date

- CDEs available
  - PerformedMedicalConditionResult.value.CD
  - PerformedMedicalConditionResult.OccurrenceDateRange.IVL_TS

29/11/2012 @ SALUS 2012
Adding CDASH Datasets and annotating them with available CDEs

- Identified Candidate CDEs:
  - MHTERM
  - MHTERM_CD
  - MHSTDAT
  - MHENDAT
  - MHONGO

- CDEs available:
  - `PerformedMedicalConditionResult.value.CD`
  - `PerformedMedicalConditionResult.OccuranceDateRange.IVL_TS`
  - `PerformedMedicalConditionResult.medicalHistoryIndicator.BL`
Possible uses of these annotations

- We aim to create Semantic Mediation Rules to map one content model to another through Common Ontology through the CDE annotations.
Possible uses of these annotations

- We aim to query the clinical content that is gathered and translated to SALUS common Ontology instances semantically through selected CDE annotations inline with the new IHE DX Profile proposition

**Diagram Steps:**

1. EHR extracts pre-population data as CCD
2. Pre-population data in CCD is mapped to BRIDG OWL representation
3. An EDC processes a study design model to extract the CDASH terms annotating the data fields in the CRF
4. Data Fields in CRF are queried from the Pre-population data using the SPARQL mappings of CDASH variables
5. The CRF is pre-populated and presented to the EHR system to be further checked and completed by the health care professional
6. Filled CRF is sent
7. Patient Medical history can be queried through CDASH/SDTM variables and terminology systems codes to collect other underlying conditions and active medications of the patient

**Flowchart:**

- **EHR System**
- **Patient Summary represented in CCD Model**
- **CDASH annotated Study Design in ODM**
- **EDC System**
- **SALUS Core Ontology**
- Statistical Analysis Tool (Regulatory Body)
Other possible uses

- Proliferation of common models (sometimes for the same purpose)
  - Transitions of Care (ToC) Clinical Element Data Dictionary (CEDD)
  - GE/Intermountain Healthcare Clinical Element Models (CEM)
  - Federal Health Information Model (FHIM)
  - Object Model of Green CDA representation of HITSP C32 (Used by HQuery)
  - OMOP CDM
  - BRIDG DAM
  - Mini Sentinel Common Data Model
  - .... Many more
    - Including the ones produced by different EU projects…
  - A metadata repository would also be useful to see the correspondences between these data models (through mapping them to common CDE set, by extending the CDE set when necessary)
    - Such mapping can be used by domain experts
    - Also can be exploited to semi-automatically defining mapping between them (to be used by programs)
      - As we are planning to do in SALUS
Design and Implementation of CDE Repository
What is Meta-data?

• data about data… (*deprecated*)
  • at design time the application contains no data
    • Descriptive metadata
  • **Structural Metadata** – data about the containers of data

• meta-data is data
  • can be stored and managed
  • meta-data registries
Importance of Meta-data

- Personal Identification:
  - TC Citizenship No: 33571191821
  - First Name: Ali Anil
  - Last Name: Sinaci
  - Date of Birth: 1985-02-02
  - Gender: Male

- Data:
  - Patient: Ali Anil Sinaci
  - B. Date: 02/02/1985
  - E-mail: anil@srđc.com.tr
What is ISO/IEC 11179?

- Family of standards addressing the:
  - Semantics of Data
  - Representation of Data
  - Registration of Data

- ISO/IEC 11179 is:
  - Description of metadata in terms of Data Elements
  - Procedures to manage registry of Data Elements

- Consists of 6 parts defining:
  - Framework for Specification
  - Classification
  - Registry Metamodel
  - Formulations of Data Definitions
  - Naming and Identification Principles
  - Registration

of Data Elements.
Benefits of ISO/IEC 11179

- Similar CDE’s linked to same Concept’s;
  - reduced search time
- All representations of a CDE can be shown together;
  - increased flexibility
- CDE’s having same value domain can be shown together;
  - easy administration of registry
- Concept of Object Class, Property and Value Domain;
  - subject-predicate-object
  - allows Linked Data representation
- Classification through External Vocabularies;
  - allows Linked Data integration
Common Data Element

The concept
What?

The representation
How?

Data Element

Person Birth Date Value

Data Element Concept

Person Birth Date

Value Domain

Birth Date Value

Data type: Calendar

Object Class

Person

Property

Birth Date

@ SALUS 2012

29/11/2012
Common Data Element

Linked Data

Data Element (DE): Person Reported Age Value

Components: Person | Reported | Age | Value

CUI: C25190 | C25375 | C25150 | C25712

Term: person
CUI: C25190
UMLS ID: C0030705
Synonyms: person, participant

Term: report
CUI: C25375
UMLS ID: C0684224

Term: age
CUI: C25150
UMLS ID: C0001779

Term: value
CUI: C25712
UMLS ID: C052797

Integration with other MDRs

• ICD9, ICD10
• SNOMED CT
• LOINC
• RxNorm
• WHO ART
• MedDRA

Diagram adopted from http://nicicbtraining.nci.nih.gov/TPOnline/TPOnline.dll/Public%20Course/COURSENO=COUR2006121515230703800967

29/11/2012
@ SALUS 2012
Organizations using ISO/IEC 11179

- Australian Institute of Health and Welfare - METeOR
- US Department of Justice - Global Justice XML Data Model GJXDM
- US Environmental Protection Agency - Environmental Data Registry
- US Health Information Knowledgebase (USHIK)
- Ohio State University - open Metadata Repository (openMDR)
- Minnesota Department of Education Metadata Registry (K-12 Data)
- Minnesota Department of Revenue Property Taxation
- The Census Bureau Corporate Metadata Repository
- Statistics Canada Integrated MetaDataBase
- The Environmental Data Registry
Why Jena based ISO 11179?

- There are lots of MDR instances out there
  - Most of them are based on ISO/IEC 11179
    - have the chance to interoperate semantically
  - ISO/IEC 11179 ontology
    - common vocabulary for meta-data level
  - Manage all items, classifications, inter-relations and links to the external world (terminology systems, taxonomies, vocabularies)
    - in a triple-store
    - easily expose as RDF
    - easily import as RDF
Interoperable through LOD

ISO/IEC 11179

  - eXtended MetaData Registry (XMDR) Project
  - not active since 2006
  - in collaboration with ISO/IEC TR 20943
    - Part 6: Framework for generating ontologies based on ISO/IEC 11179-3 Ed.
    - under development

- In collaboration with the ISO/IEC 20944 family of standards
  - describe codings, APIs, and protocols for interacting with an ISO/IEC 11179 metadata registry (MDR).
  - under development
ISO/IEC 11179-3 Ontology

- We started from scratch, implemented an ISO11179-3 ontology manually.
- one-to-one correspondence to ISO11179-3
  - No change in naming and modeling

- Every Entity in ISO11179-3 model is an ontological Concept $\rightarrow$ owl:Class
- Each property of an Entity
  - owl:ObjectProperty if the target is another Entity
  - owl:DataTypeProperty if the target is a literal
- Hierarchies of Entities
  - rdfs:subClassOf
ISO/IEC 11179-3 Ontology
Design & Implementation

Common Data Element (CDE) Repository

- CDE Repository Web GUI
- UML Model Importer
- Semantic Model Importer
- Schema Model Importer
- CDE Knowledge Base

29/11/2012 @ SALUS 2012
Design & Implementation

```java
public DataElement createDataElement(String name, String definition,
    DataElementConcept dec, ValueDomain vd);
```

```java
public DataElementResource createDataElement(
    AdministrationRecordResource dataElementAdministrationRecord,
    String representationClassQualifier, Integer dataElementPrecision,
    StewardshipRelationshipResource administeredBy,
    SubmissionRelationshipResource submittedBy,
    RegistrationAuthorityResource registeredBy,
    AdministeredItemContextResource having,
    DataElementConceptResource expressingDataElementConceptExpression,
    ValueDomainResource representedByDataElementRepresentation)
```

CDE Knowledge Base

Java API

MDR API
(Easy-to-use Semantic ISO 11179 Mapping)

Semantic MDR

Semantic Data Manipulation API
(Pure ISO 11179 Mapping)

JENA RDF/OWL API

Triple Store
(Jena TDB | Virtuoso)
Design & Implementation

- Easy-to-use high-level API for ISO 11179-3
- A methodological way for AdministeredItem maintenance for each Context

```
Context
  getDataElements();
  createConcept("name", "definition");

DataElement

Concept
  getObjectClass();
```
Design & Implementation

- Each resource is uniquely identifiable and accessible
  - Linked Common Data Elements
- Common Data Elements have links to external semantic resources
  - BioPortal or any other resource within LOD
- `asMDRResource();`
  - automatically have the Jena model of any resource
  - natively RDF
- Native SPARQL support
Design & Implementation

- **SPQRQL Queries**

  ```sparql
  PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
  PREFIX mdr:<http://salus-project.eu/iso11179-3/mdr#>
  SELECT ?oc WHERE {
    ?oc rdfs:subClassOf mdr:ObjectClass .
  }

  PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
  PREFIX mdr:<http://salus-project.eu/iso11179-3/mdr#>
  SELECT ?property WHERE {
    ?dec rdfs:subClassOf mdr:DataElementConcept .
  }
  ```

- Semantic search
  - Keyword search on semantic indexes
  - Semantic index: indexing through LOD
- Intelligent CDE matching
Other ISO standardization efforts

- **Relation with ISO/IEC 20943-6**
  - Procedures for achieving metadata registry content consistency — Part 6: Framework for generating ontologies based on ISO/IEC 11179-3 Ed. 3
  - Managing ontology concepts within 11179 MDRs → one of our implicit objectives
  - different mapping models between 11179 Entities and Ontological constructs
    - easily convertible

- **Relation with ISO/IEC 20944**
  - Metadata Registry Interoperability and Bindings (MDRIB)
  - Different approach: We try to provide the interoperability through LOD with RDF and SPARQL

  *Our main objective is to provide interoperability between different content models through CDE annotations*
Future Work

- Importers for CDE identification
  - XML Schema, UML (v1.x and v2.x), RDFS/OWL based ontologies
- Mining through content models
- Automatic CDE identification
- Semi-automatic CDE annotation
  - mappings to existing CDEs
- Guiding users through highly interactive Web based GUI
  - HTML5 with no server-side rendering
Thank you for listening...

Questions