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)

WHO-UMC and ROCHE is actively involved in pilot studies

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
Motivation I

- We address the interoperability gaps between clinical research and clinical care systems for post market safety studies.
- Clinical trials are focused and not adequate to ensure comprehensive drug safety:
  - Limited size and scope
    - Patients with co-morbidity excluded
    - Mostly no co-medication considered
  - Designed to pick-up immediate common problems not rare adverse events
  - Cannot detect long-term adverse events
Motivation II

- Post market safety studies address this problem, but
  - Reactive based on spontaneous case safety reports
    - Signal detection algorithms run by SRSs (such as WHO UMC) on top of these voluntarily sent reports
  - Medical professionals do not always see reporting a priority & detecting adverse events may not always be straightforward
    - It is estimated that medical practitioners report only about 5% of harmful drug side effects
  - Approximately 5% of all hospital admissions in Europe are due to an adverse drug reaction (ADR), and ADRs are the fifth most common cause of hospital deaths
    - An impact assessment carried out for the European Commission has estimated that ADRs cause 197,000 deaths per year in the EU, at a total cost of €79 billion
Objectives

- Enable effective integration and utilization of electronic health record (EHR) data to improve post-market safety activities on a proactive basis
  - EHR covers extended parts of the underlying medical histories, include more complete information on potential risk factors, and not restricted to patients who have experienced a suspected ADE
    - Denominator is missing in SRS data
- Aim to create the necessary infrastructure to enable secondary use of EHRs in an efficient and effective way for reinforcing the post market safety studies
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

• Screening of heterogeneous EHR data for adverse event signals detection
• Carrying out outcome research to identify 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 discovery
- 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
Semi-automatic Notification of Suspected ADEs and Reporting ADEs within a Hospital

- Adverse drug event reporting is
  - Uncoordinated
  - Manual
  - Resulting in underreporting

- SALUS supports two scenarios addressing these needs:
  - Electronic Notification of Suspected ADEs to the healthcare professional
  - Semi-automatic Reporting of ADEs to the health authorities
Notification of Suspected ADEs

- Deployment of an ADE Notification Tool within a hospital (or region)
  - On top of the Hospital EHR and the Lab System, a standard based SALUS interface
    - To specify data to be collected
    - To seamlessly collect (at regular interval or when triggered) specified data
  - SALUS Semantic Interoperability Component
    - To assess clinical data, although using different codes for diagnostic data (for example ICD 10 and SNOMED CT)
    - To assess semantic differences in the data model between the ADE Notification Tool and the EHR data model used in TUD.
- The physicians are automatically notified about possible adverse drug events that needs to be reported, once notification tool deployed
  - Integrated into the hospital information system
  - Based on a predefined set of rules
  - Upon confirmation/refutation of a suspicious ADE, the decision is stored in the local EHR database to facilitate training of more advanced algorithms for finding ADEs

16/01/2013
Semi-automatic ADE Reporting

- Reporting can be
  - either after the ADE is automatically detected
  - or after the ADE is manually detected by the treating Physician

- ICSR Reporting tool
  - automatically pre-fills the individual case safety report in E2B format
    - Using SALUS standard based query interface
    - Using SALUS Semantic Interoperability Component when necessary
  - presents it to the physician so that s/he can fill in the important missing values, e.g. free text description of the case

- Filled ICSR report in E2B XML format is submitted through normal reporting facilities to
  - to the local/international regulatory authority
  - and/or to a Pharmacovigilance.
Characterizing the cases and contrasting them to a background population

- During investigation of Nifedipine and myocardial infarction at the Uppsala Monitoring Centre 20 out of 82 cases were found to originate from SALUS connected health facilities
  - The analyst logs in to SALUS as an authorized user validated to see summarized statistics from the health care facility data connected through the SALUS architecture
  - The query is sent through the SALUS framework specifying that summarized statistics for Nifedipine and myocardial infarction contrasted against Nifedipine in general is to be returned
  - SALUS automatically highlights events and covariates that are substantially more (or less) common in patients with myocardial infarction after Nifedipine than in patients on Nifedipine in general. This allows the analyst to:
    - Identify potential risk factors and confounders like age, smoking/alcohol habits, prior prescriptions of other medications associated with medical conditions such as diabetes etc.
    - Find predisposing factors and co-morbid conditions like deep vein thrombosis indicating cardiovascular disease by comparing the medical events prior to the medical event of interest for the 20 cases to the medical events occurring in close relation to the drug prescription for all patients.
    - Characterize the outcome and course of the disease by looking at the events occurring after the medical event in question.
Characterizing the cases and contrasting them to a background population.
Temporal pattern discovery

- During investigation of Vancomycin and acidosis, 4 out of 12 cases were found to originate from SALUS connected health facilities.
- As a complement to the information on those four reports, an analysis of the temporal association between first prescriptions of Vancomycin and acidosis is carried out in the relevant SALUS data to inform the analyst whether there is an association also at the level of the exposed cohort.
- To evaluate the temporal pattern of Vancomycin and acidosis the analyst sends a query to the SALUS framework by specifying the drug at substance level (either using the ATC classification or the WHO-DD code system) and the event using MedDRA codes.
- The SALUS framework processes the query and returns the selected subset of medical summaries of the eligible patients in OMOP CDM format.

The Temporal pattern discovery method to be developed by UMC

- Calculates the number of first prescriptions of Vancomycin with an event acidosis in different time periods relative to the prescription.
- Expected counts are computed for each time period based on the number of patients at risk and the frequency of acidosis in different time periods relative to other first prescriptions (e.g. an active comparator, or the set of all possible prescriptions).

The result is used by the analyst to create a visual representation of the empirical association to enable assessment of the potential signal.

- As it turns out, the temporal pattern of Vancomycin and acidosis show increased occurrence of acidosis in the first month after new prescriptions of Vancomycin compared to what is expected for in the database and what is seen in the same cohort prior to the first prescription.
- This pattern suggests that there is an association between Vancomycin and acidosis also at the level of the cohort, and the analyst decides to communicate the safety signal to the regulatory authorities for further evaluation.
Temporal association screening

- Electronic Health Records may be a valuable complement to collections of spontaneous reports as a source of information to detect potential safety signals
- To this end, open-ended statistical pattern discovery may be deployed on collections of EHRs within the SALUS framework
- To realize the full potential of prospective surveillance, the drug exposure and event experience should be evaluated as it accumulates from the EHRs
  - Set up of a central clinical data repository of specified clinical data to be collected from the underlying EHRs
    - The clinical data fed through SALUS defined functional interoperability protocols are converted to the native data model used by the algorithms employed by UMC (OMOP CM)
    - The available algorithms for signal detection (as developed in OMOP Project) then can be continuously executed against EHR data
  - Automatic identification of emerging patterns of temporal association between drug prescriptions and medical events for detailed clinical review
    - Through temporal pattern discovery, cases series characterization, and clinical review of relevant medical history
Using EHRs as secondary data sources for pharmaceutical safety studies

- CHF is of particular concern in diabetic patients in whom incidence rates are two to five times greater than those in the general population
  - Several risk factors of CHF in diabetic patients have been identified
    - duration of diabetes, history of ischemic heart disease, renal function, hypertension, diabetes treatments and HbA1C
  - However, the incidence of CHF in diabetic patients with a recent acute coronary event is not fully known
  - In particular, no estimates of CHF for different treatment regimens are available in these patients.

- Roche is conducting clinical trials in both acute coronary syndrome (ACS) patients and in ACS patients with diabetes
  - Whilst the trials are blind, it is important to compare the observed overall incidence rate of an important adverse event like CHF in the trials with that in similar background populations
  - Such a comparison helps provide a context to the observed incidence and enables us to identify any potential safety concerns earlier on (eg if the observed incidence in the trial is greater than the background)

- **Objective**
  - Estimate incidence rates of CHF in diabetic patients with a recent acute coronary syndrome (ACS) event.
  - Secondly, to estimate incidence rates of CHF in patients on different diabetic medications.

- **Methodology:**
  - Give eligibility criteria
  - Identify the data sets to be collected
  - Analyze the data collected to estimate incidence
Challenges to be addressed in SALUS

- **The problem of Interoperability**
  - **Syntactic and Semantic**
    - The ability to exchange information
      - *access*
    - The ability to use the information once it has been exchanged
      - *understand*
  - **Security and privacy**
  - **Intelligent tools to analyze the collected content**
Challenges and corresponding SALUS WPs

- Achieving syntactic and functional interoperability between EHR Systems and clinical research systems
  1. To seamlessly query heterogeneous EHR systems for analysing and detecting possible ADEs, pre-filling case safety reports and for enabling signal follow-up studies to trace the safety reports back to the related EHRs
  2. To seamlessly specify the target eligible patient group for enabling set up of continuous safety studies that screen EHRs
  3. To specify the requested clinical data by intelligent data analysis tools for the selected group of patients
  4. To transfer the specified de-identified clinical data to the clinical data registries for the selected patients for safety analysis

- SALUS WP5: Functional Interoperability Toolkits for secondary use of EHRs in Post Market Safety Studies
  - Task 5.1 Subscription Based Interoperability Profiles and Open Source Toolsets
  - Task 5.2 Query Based Interoperability Profiles and Open Source Toolsets
  - Task 5.3 Interoperability Profiles and Open Source Toolsets for Reporting Activities for Post Market Safety Studies

Transaction definitions!
Content agnostic
Current Progress in WP5

- Tasks 5.1 & 5.2 (Subscription/Query Based Interoperability Profiles)
  - Related available interoperability approaches have been examined
    - HL7 CRFQ
    - IHE QRPH Profiles: IHE RFD, IHE CRD
      - Form based interaction, not query/subscription based, focusing on case safety reports
    - IHE PCC Profiles: IHE QED, IHE CM
      - Subscription/query based, yet not specialized for population based queries
    - Representing eligibility queries:
      - HL7 HQMF queries
      - HL7 Study Design Message
  - Now, the Consortium is working on the extensions to IHE QED and CM Profiles to pass population based queries
  - In parallel with this, we will be actively involved in IHE DEX Profile as co-author of volume I
    - Exploitation of metadata registries for flexible mapping of medical summaries to research data required by Clinical research systems (like CRF forms, safety reports)
    - Based on our IEEE TITB publication…
Current Progress in WP5

- Tasks 5.3 (Interoperability Profiles and Open Source Toolsets for Reporting Activities for Post Market Safety Studies)
  - Will be based on IHE DSC profile
    - We have examined this profile in detail
  - We have examined E2B guidelines
    - Identified the data elements which can be readily taken from EHRs
    - Identified the mappings with HL7 CCD based templates
  - We will build a semantically enabled ICSR reporting tool
    - Which will collect the medical summary from the underlying EHRs through the Interoperability profiles proposed in T5.1 & T5.2 or through SPARQL Endpoints
    - Medical summaries will be RDFized through SALUS Semantic Interoperability Layer
      - Semantically query the RDFized medical summaries (including terminology reasoning) to extract the data fields required in E2B
      - Prepare the E2B form, present to medical professionals for filling the missing fields, and send the pharmacovigilance center through the required interfaces
  - In parallel with this, we will be actively involved in IHE DEX Profile as co-author of volume I
    - Exploitation of metadata registries for flexible mapping of medical summaries to research data required by Clinical research systems (like CRF forms, safety reports)
Challenges and corresponding SALUS WPs

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

- SALUS WP4 Semantic Interoperability Framework for Post Market Safety Studies
  - Task 4.1: Developing content models for post market surveillance studies (ODM, CDA, templates, archetypes)
  - Task 4.2: By analysing these content models, developing the common core data element set as meaningful fragments → requirements for SALUS Common Ontology
  - Task 4.3: Developing SALUS semantic resource set including these common data element set and already existing domain ontologies and terminology systems
  - Task 4.4: Developing a semantic mediation framework
Proposed Semantic Mediation Approach

- Two complementary approaches will be followed:
  - By providing a semantic interoperability layer on top of the functional interoperability profiles to be developed in WP5: clinical research and clinical care systems can communicate through using well accepted standards like HL7 CDA, CEN EN 13606 archetypes, and CDISC ODMs within the scope of well defined transactions, yet be able to meaningfully interpret these syntactically different but semantically similar content models.
  - By enabling the development of semantic interfaces on top of the clinical information sources, so that clinical data exchange among clinical care and research systems can be handled based on a common semantic model.
- Provide a migration path from clinical care and research systems that can communicate through semantically enhanced functional interoperability profiles to clinical care and research systems that support full-fledged semantic systems enabling semantic interfaces through our harmonized patient safety ontology.
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
SALUS Semantic Interoperability Platform aims to build a semantic architecture, where data will be mediated to one another through common models, created as a set of semantic resources (common data elements, domain ontologies, terminology systems).
Building a part of the Common Models as a CDE Set

- We aim to create the **a part of the Common Models 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 (such as BRIDG), and domain ontologies to CDE Repository (in future CDISC Share elements?) to create the base CDEs.
  - 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: CDEs will be the basic requirements of our common ontology: We will build our Common Ontology based on that
  - Step 6: Define semantic mediation rules between content models (annotated with CDEs) and Common Ontology

- This extracted ontology will be linked with the available domain ontologies in the healthcare domain (including terminology systems) (through the linked data approach) → **SALUS Semantic Resource Set**
- 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**
Proposed Semantic Mediation Approach

1. Case Series Characterization Tool
   - SALUS Functional Interoperability profile for subscribing specific subset of clinical data available in EHRs for a specific target patient cohort

2. Eligibility Criteria based on OMOP CDM

3. Ontology Normalization Component
   - OMOP CDM-based Criteria Ontology Instance
   - ELIGIBILITY CRITERIA REDefined through HQMF Model
   - SALUS Functional Interoperability profile for subscribing specific subset of clinical data available in EHRs for a specific target patient cohort

4. EHR based Signal detection tools (Semantic aware)
   - Eligibility Criteria Semantic Query based on SALUS Common Ontology
   - Alignment Rules
     - Ontology Alignment through Reasoners

5. SALUS Semantic Mediation Framework
   - Ontology De-normalization Component
   - SALUS Semantic Resource

6. EHR System used in clinical care
   - EHR System used in clinical care (Semantic aware)

7. Semantic Interface of ORBIS
   - ORBIS DB
Current Progress in Task 4.1

- Formally defined the data requirements of all of the selected SALUS Pilot cases
  - Identify candidate CDEs

- Defined content modes
  - To represent subsets of medical summaries to cover these data requirements: SOURCE
    - As HL7 CCD Templates
    - As 13606 Archetypes
  - To represent the research data requested by different research systems: TARGET
    - E2B format for ICSR Reporting Tool
    - OMOP CDM for Case Series Characterization/Temporal Pattern discovery and Temporal Association Screening tools

- Defined mappings among them conceptually
  - CCD ↔ CDEs
  - CDEs ↔ E2B
  - 13606 archetypes ↔ CDEs
  - CDEs ↔ OMOP CDM
## Data Requirements of each pilot. As candidate CDEs

<table>
<thead>
<tr>
<th>Selected SALUS Scenarios/Related EHR Sections</th>
<th>Selected SALUS Scenarios/Related EHR Data Items</th>
<th>Enabling Notification of Suspected ADIs</th>
<th>Enabling Semi-automatic ADE Reporting</th>
<th>Characterizing the cases and contrasting them to a background population</th>
<th>Temporal pattern characterization</th>
<th>Running Exploratory Analysis Studies over EHRs for Signal Detection</th>
<th>Calculating incidence rates of CHF in diabetic patients with a recent acute coronary syndrome (ACS) event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
<td>Patient Name or Initials</td>
<td>☑️ (o)</td>
<td></td>
<td>☑️ (Psuedonym)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>☑️ (Age is to be derived)</td>
<td>☑️ (Year of Birth is Adequate)</td>
<td>☑️ (Year of Birth is Adequate)</td>
<td>☑️ (Year of Birth is Adequate)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date of Birth</td>
<td>☑️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>☑️</td>
<td>☑️ (o)</td>
<td>☑️</td>
<td>☑️</td>
<td>☑️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>☑️</td>
<td>☑️ (o)</td>
<td>☑️</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth Place (Region or City)</td>
<td>☑️</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient registration date</td>
<td></td>
<td>☑️ (o)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient de-registration date</td>
<td></td>
<td>☑️ (o)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>☑️ (o)</td>
<td>☑️ (o)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provider ID</td>
<td>☑️ (o)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provider Organisation ID</td>
<td>☑️ (o)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Address</td>
<td>☑️ (o)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### HL7 CCD Templates

<table>
<thead>
<tr>
<th>Common Data Element Name</th>
<th>Cardinality</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Type</td>
<td>1..1</td>
<td>CD</td>
</tr>
<tr>
<td>Comments / text describing Problem</td>
<td>0..1</td>
<td>ED</td>
</tr>
<tr>
<td>Start Date</td>
<td>0..1</td>
<td>TS or IVL&lt;TS&gt; (for effective time)</td>
</tr>
<tr>
<td>End Date</td>
<td>0..1</td>
<td>TS or IVL&lt;TS&gt; (for effective time)</td>
</tr>
<tr>
<td>Problem Code (can also indicate Death.cause of death)</td>
<td>0..1</td>
<td>ED</td>
</tr>
<tr>
<td>Problem Status</td>
<td>0..1</td>
<td>CD</td>
</tr>
<tr>
<td>Severity</td>
<td>0..1</td>
<td>CD</td>
</tr>
<tr>
<td>Date of Death</td>
<td>0..1</td>
<td>TS</td>
</tr>
</tbody>
</table>

### Corresponding Common Data Elements

- **Xpaths to CCD elements**
- **Location in CDA Document in the specified sections**
  - `<cda:entry/cda:act[@root='1.3.6.1.4.1.19376.1.5.3.1.4.5.2']/ cda:effectiveTime][@typeCode='SUBJ']/cda:observation/ cda:templateId[@root='1.3.6.1.4.1.19376.1.5.3.1.4.5']`
  - **Condition**
    - Problem Type
    - Comments / text describing Problem
    - Start Date
    - End Date
    - Problem Code (can also indicate Death.cause of death)
    - Problem Status
    - Severity
    - Date of Death
  - **Cardinality**
    - 1..1
    - 0..1
  - **Data Type**
    - CD
    - ED
    - TS or IVL<TS> (for effective time)
Current Progress in Task 4.2

- We are building an ISO/IEC 11179 based Metadata Registry
  - A semantic metadata registry on top of a triple store (currently both Jena and Virtuoso are supported)
  - An ontology of ISO/IEC 11179 model is created
  - 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

16/01/2013 @ SALUS -January 2013
Current Progress in Task 4.2

- 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
Current Progress in Task 4.3 & 4.4

- Logical architecture of SALUS Semantic Services and Semantic Mediation Layer is completed
- Implementation of the RESTful Semantic Service architecture is ready

- Athena Entity Services
  - Consumes/produces RDF data (text/turtle)
  - Entity configured using
    - N3 rules: inference
    - N3 queries: projection
  - Uses EYE to reason on data with configured rules, projects using configured queries

- Mapping definitions in EYE between
  - TUD DDO- Common Ontology
  - CCD Ontology- Common Ontology
  - OMOP CDM Ontology- Common Ontology

16/01/2013 @ SAL
Challenges and corresponding SALUS WPs

- Safeguarding the security and privacy of the medical data in the context of secondary use
  - Anonymization
  - Secure Exchange
  - Accountability
    - Audit mechanisms
  - Authorization and Authentication
- SALUS WP5: Functional Interoperability Toolkits for secondary use of EHRs in Post Market Safety Studies
  - Task 5.4 Interoperability Profiles and Open Source toolsets for Security and privacy
Current Progress in Task 5.4

- Examined policies and available standards
  - EU Directive 95/46/EC
  - Opinion 4/2007 on the concept of personal data
  - ISO/TS 25237:2008 Health Informatics – Pseudonymization
  - HITSP Anonymization guidelines
  - NHS Pseudonymization Implementation Project (PIP)
  - IHE IT Infrastructure Healthcare Pseudonymization Handbook
  - Pommerening approach by TMF Germany

- Established the Security architecture
  - Services identified
  - Exchange of medical data between different zones are analyzed

16/01/2013
Current Progress in Task 5.4

- De-Identification Service is implemented in a flexible and extensible way
  - Redaction: Removing an atomic data element
  - Fuzzing: Adding “noise” to an atomic data element
  - Generalization: Making an atomic data element less specific
  - Longitudinal consistency: Modifying data so that it is shifted by a specific amount
  - Text Processing: Special considerations for free-format text
  - (Recoverable) Substitution: Changing one data element into another data element
  - Pass-through: No change

- Performed first tests with CDA documents

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Algorithm</th>
<th>Longitudinal Consistency</th>
<th>Precision Loss</th>
<th>distortions</th>
<th>format</th>
<th>RD</th>
<th>Failure of Blinding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redaction</td>
<td>Delete Attribute Name and Value</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Delete Value</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fixed Length Substitution Value</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Original Length Substitution Value</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data Type Labelling</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>Substitution</td>
<td>Substitute Attribute Name and Value</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Substitute Meaningful Value</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Substitute Meaningless Value</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>Fuzzing</td>
<td>Numerical (statistical algorithm)</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>Generalization</td>
<td>Code Set (Random)</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Week values</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Remove day</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CODE to age</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Statistical adjustment</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time of day (such as w/t meals)</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Specialized time (weekend vs week day)</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shift by random offset</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Move to day of week</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Move to day of month</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Remove day of week</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time of Day Generalization</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shift by offset</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geographical locations</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Consistent algorithmic substitution of coded value</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Longitudinal Consistency</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Date/Time Generalization</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Recoverable) Substitution</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sequential Values</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provided by Third Party</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pass-through</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test Processing</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
</tr>
</tbody>
</table>

16/01/2013 @ SALUS -January 2013
Challenges and corresponding SALUS WPs

- Developing intelligent data analysis algorithms
  - To detect possible ADEs within a healthcare institute (or region) to notify medical professionals to increase ICSR reporting rates
  - To achieve open-ended temporal pattern discovery on top of the electronic health records collected from disparate, distributed, heterogeneous EHR Systems
    - While large collections of electronic health records are a standard resource for epidemiological confirmatory studies, their use in the context of exploratory data analysis is still limited

- SALUS WP6: Integrating Clinical Research with Clinical Care
  - Task 6.1 Enabling ADE Detection on EHRs based on temporal patterns
  - Task 6.2 Enabling Exploratory analysis for Signal detection and signal qualification studies on EHRs based on temporal patterns
Current Progress in WP6

- **ADE Notification Tool**
  - Current progress: Identification of indicators of adverse drug events, e.g. liver-, kidney-, muscle-parameters or bone-marrow- and electrolyte-values.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Normal range (before prescription)</th>
<th>Ade detection rule (after prescription)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALAT (Liver)</td>
<td>Male: 10 - 50 U/l</td>
<td>2 x normal value (value before drug prescription)</td>
</tr>
<tr>
<td></td>
<td>Female: 10-35 U/l</td>
<td></td>
</tr>
<tr>
<td>Myoglobin (Muscle)</td>
<td>Male: 20 – 70 µg/l</td>
<td>2 x normal value (value before drug prescription)</td>
</tr>
<tr>
<td></td>
<td>Female: 16 – 60 µg/l</td>
<td></td>
</tr>
<tr>
<td>Creatinine (Kidney)</td>
<td>Blood: Male: 74 – 110 µmol/l</td>
<td>2 x normal value (value before drug prescription)</td>
</tr>
<tr>
<td></td>
<td>Female: 58 – 96 µmol/l</td>
<td></td>
</tr>
</tbody>
</table>

- **Temporal Association Screening and Temporal pattern discovery**
  - Design of the algorithms/methods
    - Case Series Characterization Tool
    - Temporal Pattern Discovery tool
      - What differs between the patients having a myocardial infarction within two weeks of Nifedipine intake to all the other patients taking Nifedipine?
  - Temporal Association screening on EHRs
    - What does the Medical Event profile look like for Nifedipine?
    - Are the any drugs that might be associated with causing Myocardial Infarction?
    - Open ended analysis, no prior hypothesis
    - Generates associations that might become signals
Project objectives in a nutshell..

- Provide an interoperability architecture to increase the scalability of post market safety studies
  - Enable communication with various different type of EHRs
    - Through CDA based interfaces, I3606 based interfaces, or SPARQL endpoint
  - Enable different safety analysis tools to query SALUS Platform, and easily consume data
    - Eg. OMOP Methods
    - On top of that, we will develop semantic enabled safety analysis tool to show the power of semantic data processing
      - Eg. ICSR reporting tool, ADE Notification Tool

- In a recent study, the ability to detect ADEs over historical EHR data is evaluated where the excess risk of acute myocardial infarction of celecoxib users compared to naproxen users is assessed through sequential analysis of the collected EHR data
  - The signal is identified at month 34 by screening the historical clinical data set of 7 million people
  - However it is expected that such a signal would have occurred by 2nd or 3rd month if 100 million people had been observed
  - **SALUS interoperability architecture aims to enable the signal exploratory studies to scale to such wide EHR sources so that safety signals can be detected as early as possible.**
Thank you for listening...

Questions

Gokce B. Laleci