Ontology-mediated Data Integration for real-time Antibiotics Resistance Surveillance

SWAT4LS 2014, Berlin

Dr. Daniel Schober  www.ipb-halle.de
Projects Clinical Context

- Steep increase in bacterial antibiotics resistance
- Nasty nosocomial hospital infections
• **DebugIT**
  – Detecting and Eliminating Bacteria Using IT
  – ‘ITbiotics’ approach
    • Enable cross-country resistance-data comparison
  – Using ‘semantic data’ to exploit distributed clinical data
    • Apply data in resistance monitoring

• Why are existing surveillance networks insufficient?
Reviewing existing surveillance efforts

- **Paul Ehrlich Gesellschaft (PEG) surveillance**
  - Resistance data from <30 German clinics (not covering EU)
  - 3 year interval (low time resolution)

- **European Antimicrobial Resistance Surveillance (EARS-NET)**
  - European resistance comparison
  - 1 year interval
  - Limited to specific clinical isolates, i.e. blood culture samples

- **German Network for Antimicrobial Resistance Surveillance (GENARS)**
  - Only German sites

- **German Krankenhaus Infektions Surveillance-System (KISS)**
  - Only looks at intensive care units
  - Few pathogens

- **SARI system**
  - Only German sites
  - Low time resolution
**Drawbacks → Requirements for DebugIT Surveillance**

**Limited access to local data-sets**
→ Parallel integrated access to EU-wide hospital data

**Limited timely availability & coarse time resolution**
→ Real-time data access

**Limited automatisation**
→ Automatic real-time feed of hospital data into vCDR

**Limited data coherence, dissemination & re-use**
→ Ontologies formalize exchange syntax & domain semantics

**Limited data coverage & granularity**
→ Granular domain coverage through expressive ontologies
Methods

- Real-time Data Federation
  - Lexical normalization
  - ETL Process
  - D2R conversion

- Ontologies
  - Integrating geographically distributed, multilingual syntactic & semantically heterogeneous data
    - DDO (Data Definition Ontologies) in N3
    - DCO (DebugIT Core Ontology) in OWL-DL, SKOS
    - Terminological mapping rules, N3

- Reasoning
  - Tableau based DL reasoning (Hermit), T-Box
  - N3 rule-based coherent logics reasoning (Euler-EYE), A-Box

- SPARQL querying
  - On DDO layer for multiple clinical endpoints
  - On DCO layer for vCDR endpoint
**DebugIT SIP Architecture**

1. **Local CIS (DB)**
   - UKLFR CIS
   - INSERM CIS
   - HUG CIS
   - D2R mapping calls
   - D2R Mappings
   - Lexical Normalization

2. **Local wrapper (DDO) for Local CDR**
   - Freiburg DDO Endpoint
   - Paris DDO Endpoint
   - Gueneva Endpoint
   - Syntactic Normalization

3. **Semantic Integration: Global Integration Layer**
   - Clinical Analysis SPARQL Query (CASQ)
   - Common DO Endpoint, Virtual CDR
   - Semantic Normalization

4. **Syntactic Integration: Local Mediation Layer**
   - Data Set SPARQL Query (DSSQ)
   - Euler Eye Result Rewriting
   - DDO
   - DDO2DO Conversion Local N3 Rules

5. **User access: GUI/Tool Layer**
   - Clinical Researcher
   - Question Authoring Tool
   - Result Display
   - Artemis Monitoring Dashboard

6. **Query specification**
   - User query
   - Query Template
   - Result Set
Ontology Layers in WP1a

7 Data Definition Ontologie (DDO) average 40 Entities
- **Mediation layer** closing the ‘formality’-gap
- Describing **site-specific local** CIS data models in RDF
- For SPARQL data access to local hospital data (DSSQ)

13 Operational ontologies (OO, e.g CAO) average 35 Entities
- **Mediation & Syntaxtic Integration layer**
- Implementation, module crosstalk, data mining, evidences, maths, units, …
- OWL-Full → Coherent Logic reasoning (e.g. rule-based)

1 **DebugIT Core Ontology (DCO)** ~ 1720 Entities
- **Semantic Integration layer**, mapped to DDOs & external Terminologies
- Rooted in Biotop upper level ontology
- **Global**, clinical domain of infectious diseases
- OWL-DL → DL & Coherent Logic reasoning
- For SPARQL data access on vCDR level (CASQ)
Describing real world (independent of data)

Describing data

DDO~DCO \rightarrow \text{Local to Global rules to later create SPARQL CONSTRUCT/WHERE clauses}
DL Reasoning for DCO maintenance
(Infering BloodSample is a BodyLiquidSample)

Stated Facts

BodyLiquidSample =

```
- Sample and derivesFrom some BodyLiquid
```

BloodSample =

```
- Sample and derivesFrom some Blood
```

Asserted Hierarchy (flat list)

Inferred Hierarchy (more structure)

Logics Reasoner
Artificial Intelligence helps DCO modeling

- DL semantics used to spot modelling errors
  ‘antimicrobial susceptibility test’ shown to be inconsistent by reasoner
Example Inference (N3 rule)

```
{  ?episodeOfCare a dco:EpisodeOfCare.
    ?bacterialAntibiogramAnalysis
       a dco:BacterialAntibiogramAnalysis;
       biotop:processualPartOf ?episodeOfCare;
       biotop:hasParticipant [biotop:encodes [a ?Antibiotic]];
       dco:hasResultDateTime ?abgResultTime.
    ?abTherapy
       a dco:AntibioticTherapy;
       biotop:processualPartOf ?episodeOfCare;
       biotop:hasParticipant [a ?Antibiotic];
       dco:hasStartDateTime ?therapyStart.
    ?dif math:greaterThan "P0D"^^xsd:duration; math:lessThan "P2D"^^xsd:duration}
=>
{  ?abTherapy
    a dco:AntibioticTherapy;
    event:basedOn ?bacterialAntibiogramAnalysis.
    ?bacterialAntibiogramAnalysis a dco:BacterialAntibiogramAnalysis}.
```

Check if AntibioticTherapy is Antibiogramm based, i.e. started within 2 days after an antibiogram result for same EpisodeOfCare, Bacterium & Antibiotic
Open an existing analysis

1. What is the percentage of EColi cases, cultured from UrinaryTractInfection collected by a UrineSample is resistant to Fluoroquinolone in the period from 1 July 2007 to (not including) 30 September 2007 at https://debugit.spim.jussieu.fr?
   Last modified on Jul 25, 2011 12:30:34 PM
   https://debugit.agfa.net/analysis/resource/clinicalAnalysisQuery/7bce4957-b977-4b04-a5eb-e4ee8299eaf9?this - properties - cancel

2. What is the percentage of EColi cases, cultured from some sample type collected by a some sample collection type is resistant to Fluoroquinolone in the period from 1 January 2007 to (not including) 15 January 2011 at https://babar.unige.ch:8443/cdr?
   Last modified on Jun 8, 2011 10:50:42 AM
   https://debugit.agfa.net/analysis/resource/clinicalAnalysisQuery/85a7e508-8d64-4a4c-8328-60da6c0feeda9?this - properties - cancel

3. What is the percentage of antibiotic therapies using Trimethoprim in the period from 2 May 2010 to (not including) 30 May 2011 at http://debugit.kontrax.bg?
   Last modified on May 30, 2011 8:19:50 AM
   https://debugit.agfa.net/analysis/resource/clinicalAnalysisQuery/02d24634-0e0-44af-bd5a-56c5d2c566b?this - properties - cancel

4. What is the percentage of EColi cases, cultured from some sample type collected by a some sample collection type is resistant to SAureus in the period from 1 July 2007 to (not including) 30 September 2007 at https://debugit.spim.jussieu.fr?
   Last modified on Apr 20, 2011 10:37:16 AM

Create new analysis

1. Clinical Analysis Query 1 Template with a list of antibiotics and loci
   Thu Mar 31 2011 18:16:02 GMT+0200 (Romance Daylight Time)
   What percentage of some bacteria, cultured from some sample, is resistant to some antibiotics in a certain period at a certain location? properties

2. Clinical Analysis Query 1 Template with locus
   Mon Mar 28 2011 15:31:02 GMT+0200 (Romance Daylight Time)
   What percentage of some bacteria, cultured from some sample, is resistant to some antibiotic or another antibiotic in a certain period at a certain location? properties

3. Clinical Analysis Query 20 Template
   Mon Mar 28 2011 14:40:03 GMT+0200 (Romance Daylight Time)
   What percentage of some sampletype during a certain period from some location from which some bacteria is cultured? properties

Filter: [Blank] Search

Custom analysis

New custom analysis
Clinical Analysis Question: What is the percentage of EColi cases, cultured from UrineSample collected by a UrineSampleCollection is resistant to Ciprofloxacin in the period from 1 November 2006 to (not including) 1 December 2006 at https://babar.unige.ch:8443/cdr?
Configurable Resistance Monitoring Dashboard

- Collated data is analyzed and visualized graphically displaying selected query results as freely configurable diagrams.
- Visualization portlets, called gadgets, show the results of the CASQ SPARQL queries for the selected hospital sites.
Result Evaluation

• DebugIT vs. HEGP resistance trends for 2001-2007

<table>
<thead>
<tr>
<th>E. Coli vs CEFIXIME: sensitive @ hegp vs debugit</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image-url" alt="Graph showing percentage comparison between HEGP and DebugIT over years" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HEGP (internal report)</strong></td>
<td>73%</td>
<td>85%</td>
<td>90%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td>89%</td>
</tr>
<tr>
<td><strong>DebugIT (SPARQL endpoint)</strong></td>
<td>73%</td>
<td>86%</td>
<td>89%</td>
<td>91%</td>
<td>92%</td>
<td>91%</td>
<td>91%</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td>1244</td>
<td>2404</td>
<td>2853</td>
<td>2780</td>
<td>2838</td>
<td>2850</td>
<td>2727</td>
</tr>
<tr>
<td><strong>ddo:Sensitive</strong></td>
<td>911</td>
<td>2074</td>
<td>2553</td>
<td>2525</td>
<td>2607</td>
<td>2596</td>
<td>2479</td>
</tr>
</tbody>
</table>

• Good alignment between DebugIT Results & local check by Microbiologists
Conclusion I

- Archived federation of heterogeneous local data into coherent formalized ‘virtual CDR’

- Created semantically & geographically interoperable resistance data
  - Integration across 6 languages over 7 EU hospitals

- Feasibility of semantic integration approach
  - Open & scalable
  - Real-time access
  - Deep-Annotation rather than solely shallow (schema based) annotation
Conclusion II

• Stepwise semantic formalization
  – Bi-layered hybrid formalization approach
    • Bridge semantic gap between local DDO RDF data & global formal DCO OWL integration layer
    • Rule reasoning for layer-binding (DDO to DCO mapping)
    • DL-reasoning for DCO engineering & quality assurance

• Good end user compliance
  – Complexity shielded from user
  – Good tool usability measured via a short questionnaire
    • 10 clinicians, using 5-point Likert scale
Resources & Acknowledgements

DebugIT Resources
• http://www.debugit.eu

Acknowledgements
• Remy Choquet, Hans Cools, Kristof Depraetere, Christian Lovis, Douglas Teodoro, Emilie Paasche, Giovanni Mels, Frank Enders, Philipp Daumke, Daniel Karlsson, Steffen Neumann, Stefan Schulz, Ilinca Tudose, Maren Kechel, Jessica Bullenkamp, Patrick Ruch, Jos De Roo, Marie-Christine Jaulent, Dirk Coalert, Martin Boeker

• Funded by EU 7th Framework Program grant agreement ICT-2007.5.2-217139
Reasoner

• **Formal logical inferring**
  • Adhering coherent logic (close to first order logic)

• **EYE: Euler Yap Engine**
  • Open source: [http://eulersharp.sourceforge.net/](http://eulersharp.sourceforge.net/)
  • Inference engine supporting logic based proofs
  • Backward-chaining reasoner enhanced with Euler path detection
  • Implementation in Java/Prolog (logic programming language)
  • **60 million lips** (logical inferences / second) on an Intel Core Duo 2.2 GHz