DebugIT: Ontology-mediated layered Data Integration for real-time Antibiotics Resistance Surveillance

Daniel Schober[a,*,1], Remy Choquet[b], Kristof Depraetere[c], Frank Enders[d], Philipp Daumke[d], Marie-Christine Jaulent[b], Douglas Teodoro[e], Emilie Pasche[e], Christian Lovis[e], Martin Boeker[a]

[a] Center for Medical Biometry and Medical Informatics, Medical Center - University of Freiburg, Germany
`dschober@ipb-halle.de, martin.boeker@uniklinik-freiburg.de`

[b] INSERM, LIMICS, (UMR_S 1142) F-75006 Paris Université, France
`remy.choquet@gmail.com, marie-Christine.Jaulent@crc.jussieu.fr`

[c] Advanced Clinical Applications Research Group, Agfa HealthCare NV, Gent, Belgium
`kristof.depraetere@agfa.com`

[d] AVERBIS, Averbis GmbH, Freiburg, Germany
`enders@averbis.com, daumke@averbis.de`

[e] Division of Medical Information Sciences, University Hospitals of Geneva, Switzerland
`dhteodoro@gmail.com, emilie.pasche@gmail.com,`
`christian.lovis@hcuge.ch`

---

[1] Present address: Leibniz Institute of Plant Biochemistry, Dept. of Stress and Developmental Biology, Weinberg 3, Tel. +49 (0) 345 5582 – 1476, 06120 Halle, Germany

Abstract. Antibiotics resistance poses a significant problem in today's hospital care. Although large amounts of resistance data are gathered locally, they cannot be compared globally due to format and access diversity.

We present an ontology-based integration approach serving an EU project in making antibiotics resistance data semantically and geographically interoperable. We particularly focus on EU-wide clinical data integration for real-time antibiotic resistance surveillance. The data semantics is formalized by multiple layers of terminology-bound description logic ontologies. Local database-to-RDF (D2R) converters, normalizers and data wrapper ontologies render hospital data accessible to SPARQL queries, which populate a mediator layer. This semiformal data is then integrated and rendered comparable via formal OWL domain ontologies and rule-driven reasoning applications. The presented integration layer enables clinical data miners to query over multiple hospitals which behave like one homogeneous 'virtual clinical information system'. We show how cross-site querying can be achieved across borders, languages and different data models. Aside the drawbacks, we elaborate on the unique advantages over comparable previous efforts, i.e. tackling real-time data access and scalability.

Keywords: Ontology, Semantic Web, Data Integration, Interoperability, Antibiotics Resistance, Infection Monitoring, Data Linkage, Public Health Surveillance, Epidemiological Monitoring, Antibacterial Drug Resistance, Antibiotics

1 Introduction

As reflected in the appearance of recent infection control projects such as Infection-Control 2020[2], antibiotics resistance still poses a significant health risk with high economic impact [1]. One of the main obstacles to effectively counteract the development of multi-resistant bacteria is the failure to provide clinical data on infectious diseases from different sites in real time. The **DebugIT** project (**D**etecting and **E**liminating **B**acteria **U**sin**G** **I**nformation **T**echnology, http://www.debugit.eu), a large scale data integration effort funded within the EU 7th Framework Program, aimed to use semantic web technologies to fight the increase in antibiotics resistances by means of EU wide monitoring and surveillance [2].

In order to access and compare such distributed patient data over various nations, languages, formats and databases, DebugIT had to find a way to harmonize and integrate this heterogeneous data. To be able to monitor and alert on resistance development, a software architecture had to be set up to access, communicate and compare clinical data from different European hospitals. We here derive the requirements of our integration architecture by reviewing the drawbacks of existing resistance monitoring programs and show how real time data access and comparison as well as sec-

---

[*] corresponding author

[2] www.infectcontrol.de

ondary data usage is facilitated in the DebugIT system. Data within such a semantic interoperability platform (SIP) has to be uniform with respect to its intended meaning to ensure coherent interpretation by humans and processing tools.

The aim of this work is to present and analyse a layered ontology-architecture approach serving the DebugIT SIP and its Artemis resistance monitor [3]. Each formalization-layer represents a step further from a concrete local and context-dependent data source towards an integrated global, formal and unambiguous data semantics. As the core ontologies' creation, maintenance and design principles have been described earlier [4], we here focus on the ontology-based data integration approach and elaborate on how the ontology layers are used within the DebugIT application chain at runtime. We show how ontological expressions are used in querying locally dispersed resistance-data and how this data is exploited in surveillance applications feeding clinical monitoring and real-time resistance alert systems.

2 Backgrounds

2.1 Requirements for an EU wide surveillance system

The DebugIT consortium attempts to set up an antibiotics surveillance system to be used by medical IT personnel that alleviates the drawbacks of existing approaches. We reviewed earlier efforts for such drawbacks (see Supplementary Material[3]) and derived the following requirements for our particular use case:

a) Enabling parallel access to EU-wide hospital data.
b) Allowing for highly granular domain coverage by means of expressive ontologies of sufficient domain coverage.
c) Granting real-time data access by means of fully automatic *semantic web* toolchains.
d) Defining ontologies serving as common format for exchange syntax and domain semantics.
e) Setting-up an automatic feed of real hospital data into knowledge bases with formal semantics, allowing for secondary data exploitation and pattern discovery in a timely fashion.

3 Material and Methods

3.1 The layered ontology-based data integration architecture

DebugIT is set up as a Service Oriented Architecture (SOA), where all framework components invoke services from each other and communicate in ontologically defined semantics, using ETL mechanisms. We choose an ontology-based approach [5],

---

[3] Please find supplementary material, indicated per manuscript section, on our webpage at msbi.ipb-halle.de/msbi/debugit

to formalize our data semantics and used the BioTop[4] formal upper level ontology for description logics (DL) constraint inheritance. Access to geographically distribute and semantically heterogeneous data is integrated via semantic web technologies like OWL and Notation 3 (N3) ontologies plus rules[5]. SPARQL endpoints are used for data querying [6]. At the root stands a data access mechanism related to the federated data warehouse model approach described in [7]. The interoperability backbone represents a wrapper-mediator architecture implemented in RDF syntax, which allows for data re-usage in an Open Linked Data approach [8].

The overall ontology based interoperability architecture is based on the W3C Health Care and Life Science (HCLS) Linked Data Guide[6], but in order to bridge the semantic gap from informal database entries to ontological descriptors in formal DL, we choose a hybrid ontology approach as described in [9], mapping local ontologies to a global ontology for scalability reasons. A stepwise data conversion approach over two ontology layers with different degrees of formality is applied (Fig. 1 and Supplementary Material).

The complete flow chain comprises of three levels of data integration, each of which consist of a data representation layer and an associated mapping and query step. Thus, the complete integration chain consists of a stack of six communication artefacts. These levels are here sequentially identified from local relational databases to the highest level of semantic integration on formal knowledge representations. The data representation layer of each level is indicated by Roman numbers I - III, the directly associated query step on top of this representation with Arabic numbers 1-3, and the mappings between the layers are indicated via Greek letters α, β, γ.

On the first level of data integration, the relational database data (I) is lexically normalized via mappings to medical terminologies and morphosemantic mapping employing the Averbis Morphosaurus software[7]. We enrich ambiguous local data with ontological expressions in OWL on the levels II and III of the semantic integration framework. A D2R mapping call (1) exploits a D2R mapping assignment (α, to populate a local but internet-accessible RDF wrapper in form of a SPARQL endpoint (II). This level II employs so-called Data Definition Ontologies (DDO) [10], which bridge the gap from local information models to semi-formal data on the local mediation layer, serving syntactic integration and the ETL process,[8]. Here the SOA services request the local RDF converted data (I) via SPARQL queries (2), which we call Data Set SPARQL Queries (DSSQ).

In the next integration step (2) the local DDO data (II) is mapped onto the DebugIT core ontology (DCO [4]), and Operational Ontologies (OO) (III) via DDO2DCO mapping rules (β) in N3 format using Simple Knowledge Organization Structure (SKOS) mappings. The particular formalization approach is chosen depending on the

---

[4] http://bioportal.bioontology.org/ontologies/BT

[5] enforced via the coherent logics reasoner Euler Eye: http://eulersharp.sourceforge.net/README#eye, last accessed 03.03.14,

[6] http://www.w3.org/2001/sw/hcls/notes/hcls-rdf-guide/, last accessed 03.03.14

[7] http://www.freidok.uni-freiburg.de/volltexte/4932/pdf/diss_daumke.pdf, last accessed 03.03.14

[8] http://en.wikipedia.org/wiki/Extract,_transform,_load, last accessed 03.03.14

datatypes appearing in the original Clinical Information Systems (CIS). For free text data, we exploit a rule-driven formalization, e.g. manually generated database to DDO (D2R) mappings (α) followed by local-to-global DDO to DCO N3 mapping rules. For terminologies/codes, we also use a chain of SKOS terminology mappings, e.g. ICD-10 to SNOMED-CT and SNOMED-CT to DCO.

In accordance to [11] on the semantic integration level (III), the data is now globally accessible by means of so-called Clinical Analysis SPARQL Queries (CASQ (3)), addressing one common Domain Ontology formalized endpoint and representing an integrated virtual Clinical Data Repository (vCDR).

For data normalization and DDO instance generation on layer II, results from text-mining normalization approaches encoded in N3 and medical terminology mappings (encoded via SKOS) were integrated, linking the database schema and contained data to a common underlying individual databases. The links between data and entities in the ontologies employed here combine ‘shallow annotation’, i.e. annotating per database schemata fields, and ‘deep annotation’, i.e. annotating per database values) [12]. These two local-to-global formalization approaches are applied in parallel to create formalized class instances in the ontologized layers II and III. The described mapping and integration approaches are exemplified with concrete code examples in the Supplementary Material.

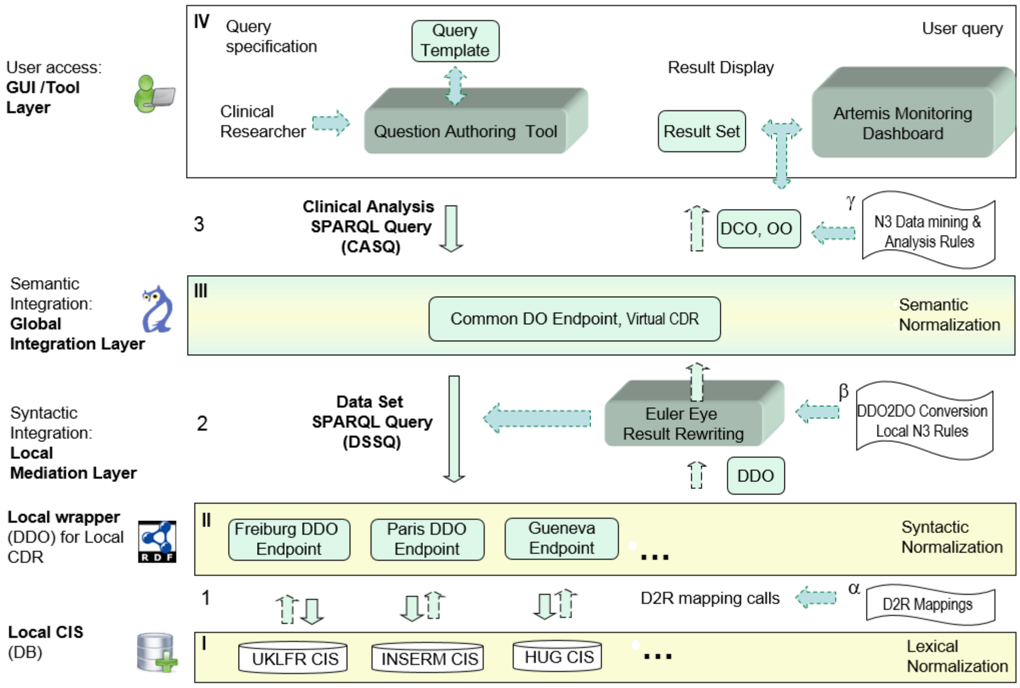


Fig. 1. Layered DebugIT mediator architecture. A schematic overview of the data formalization layers serving distributed data integration in the DebugIT interoperability platform. Clinical data in local relational databases (I) can be accessed in real-time via SPARQL querying the local endpoints via a DSSQ using DDOs on an Extract Transform Load wrapper (II). Integrated cross-site queries can be stated as CASQ using DCO and Operational Ontologies (OO) (III). The DDO (local) to DCO (global) binding is done via N3 mapping rules enforced by the Euler Eye reasoner. CASQ results are fed back to the clinical researcher or into applications e.g. a resistance monitor. For each data representation layer I - III a layer-specific mapping process had to be employed which binds each representational layer to the succeeding layer above. I.e., the first mapping α (D2R mappings) binds the underlying relational database layer (I) to the RDF representation layer (layer II), the next mapping β (N3 and SKOS) binds the RDF layer II to the Domain Ontology (DO) layer III. These stacked mappings enable each semantic layer to 'interpret' the results provided by the layer one level below with a fixed semantics and hence allow a user or a monitoring tool to query distributed clinical data integratively.

4 Results

4.1 Generated Ontologies

The built ontologies cover a conceptual space from patient data and infectious diseases to antibiotics resistance measurements. After axiomatization of these elements, the ontologies were expanded in an iterative use-case-driven manner: We successively iterated through the agreed set[9] of competency questions, adding representational units needed to answer these.

Seven **Data definition ontologies (DDOs** were generated locally and are accessible under their respective endpoints. They define between 10 and 25 classes each. **Operational ontologies (OOs)** deliver semantic identifiers for the implementation of the DebugIT framework itself, such as query building, data mining, decision support, evidences, units, quantities, SKOS schemes and datatypes. **Domain ontologies (DOs** like DCO and clinical analysis ontologies (CAO)) add description logic expressivity SRIQ(D) [13] to the antibiotics terminology. The Hermit DL reasoner[10] was used, which takes ~10 min on an average PC to check 668 DCO classes including its 400 top level BioTop classes for consistency.

4.2 Applying DebugIT on the client side

At the moment, two end-user applications are set up to exploit the ontology-integrated data: the monitoring dashboard for the DebugIT SIP with the knowledge authoring tool[11] and the Antimicrobial Resistance Trend Monitoring System (Artemis, an automatic resistance comparison and visualization tool [3]. The SIP dashboard (Fig. 2) is

---

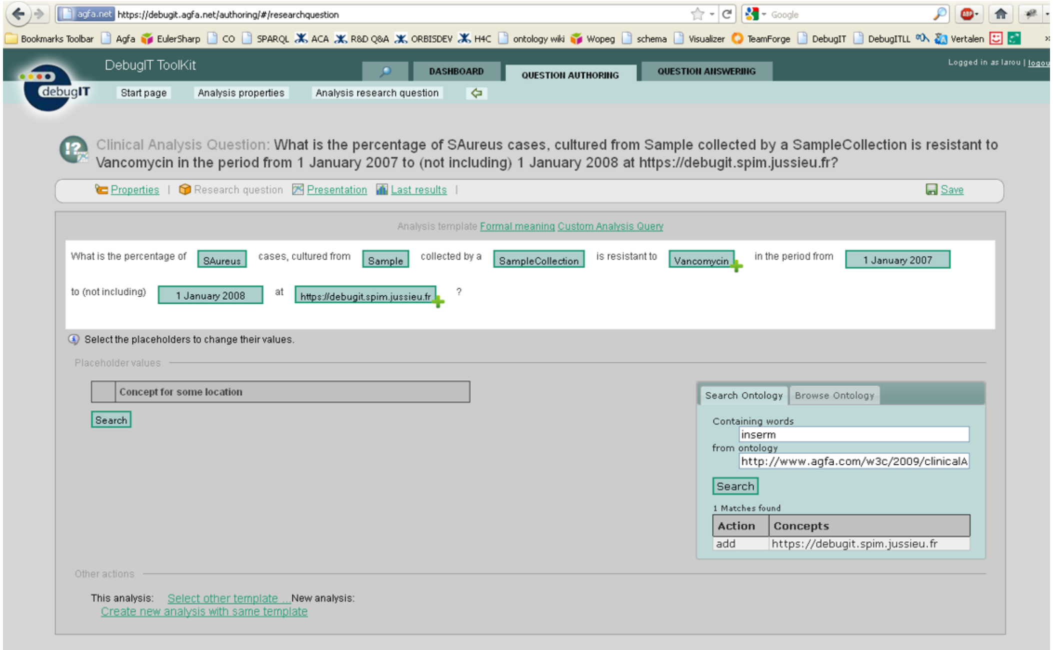
[9] http://www.debugit.eu/progress/documents/DebugIT%20D1.1b%2020091214%202145.pdf, Page 47, last accessed 03.03.14

[10] http://hermit-reasoner.com/, last accessed 03.03.14

[11] https://debugit.agfa.net/authoring/, last accessed 03.03.14

an end users interface to interact with the DebugIT knowledge base and ask knowledge-driven clinical questions to its participating clinical sites. The user has three options reflected as GUI tabs, closable depending on the type of data he is interested in:

1. **Select and run an existing predefined CASQ query** with all template fillers provided as bound variables. The user searches in a list of predefined CASQ in a given query-library, searching queries by query-type, keywords, author's name, validation status or result-age. No particular skill prerequisites are demanded on the user side.
2. **Align and refine an existing CASQ query** template to specific user-needs by changing filler values (selecting specified ontology terms, Fig. 2.). To do so, users need to understand at least the taxonomic *is_a* hierarchy of the domain ontologies.
3. **Generate a new CASQ query** with new overall semantics and store it in the repository as a new custom analysis. To do this the user needs to have a deeper understanding of semantic web technologies like N3 formatted subject-predicate-object triple statements.



https://debugit.agfa.net/authoring/#/researchquestion

Logged in as larou | logou

Clinical Analysis Question: What is the percentage of SAureus cases, cultured from Sample collected by a SampleCollection is resistant to Vancomycin in the period from 1 January 2007 to (not including) 1 January 2008 at https://debugit.spim.jussieu.fr?

Analysis template Formal meaning Custom Analysis Query

What is the percentage of SAureus cases, cultured from Sample collected by a SampleCollection is resistant to Vancomycin in the period from 1 January 2007 to (not including) 1 January 2008 at https://debugit.spim.jussieu.fr ?

Select the placeholders to change their values.

Placeholder values

Concept for some location

Containing words: inserm

from ontology: http://www.agfa.com/w3c/2009/clinicalA

1 Matches found

Action	Concepts
add	https://debugit.spim.jussieu.fr

Other actions

This analysis: Select other template ...New analysis: Create new analysis with same template

Fig. 2. Question Authoring Tool. An already formalized query template can be aligned according to a specific research question, by binding it to concrete actual variables by selecting appropriate DO classes.

The Debug IT semantic integration platform has been evaluated against the set of competency questions and responded to all competency questions successfully. In the

Supplementary Material, we provide an exemplary view on the semantic integration process for a selected competency question, corresponding query template and mapping layers. All intermediate formal artefacts are shown and explained there.

4.3 DebugIT usage within EU hospitals

The DebugIT dashboard is used in different European hospitals to update their CIS with the new DebugIT knowledge. The University Hospital of Geneva (HUG) for example, consults DebugIT generated results through the Artemis monitor, although not driving antibiotics prescription decisions directly, because of the tight quality procedures that must be fulfilled to ensure functionality in the clinical production system. Microbiologists at the Hôpital Européen Georges Pompidou (HEGP, Paris) use DebugIT to drive the yearly reporting on resistance patterns in a semiautomatic fashion[12]. General tool usability was measured via a short questionnaire with ten clinicians using standardized interviews on a 5-point Likert scale[13], which indicated that the tool lived up to the expectations.

4.4 Applications for secondary data usage of integrated data

The formalized instances are now ready for secondary data usage in an Open Linked Data fashion and amenable to intelligent data mining techniques. External semantic web tools can now be applied on the integrated data, e.g. in our case, collated data is analyzed and visualized graphically in our Pan-European resistance monitoring dashboard, displaying selected query results as freely configurable diagrams (Fig. 3).

---

[12] http://www.biomedcentral.com/content/pdf/1753-6561-5-S6-P320.pdf

[13] http://www.debugit.eu/progress/documents/DebugIT_D7_2_20110214-Dipak.pdf, last accessed 03.03.14

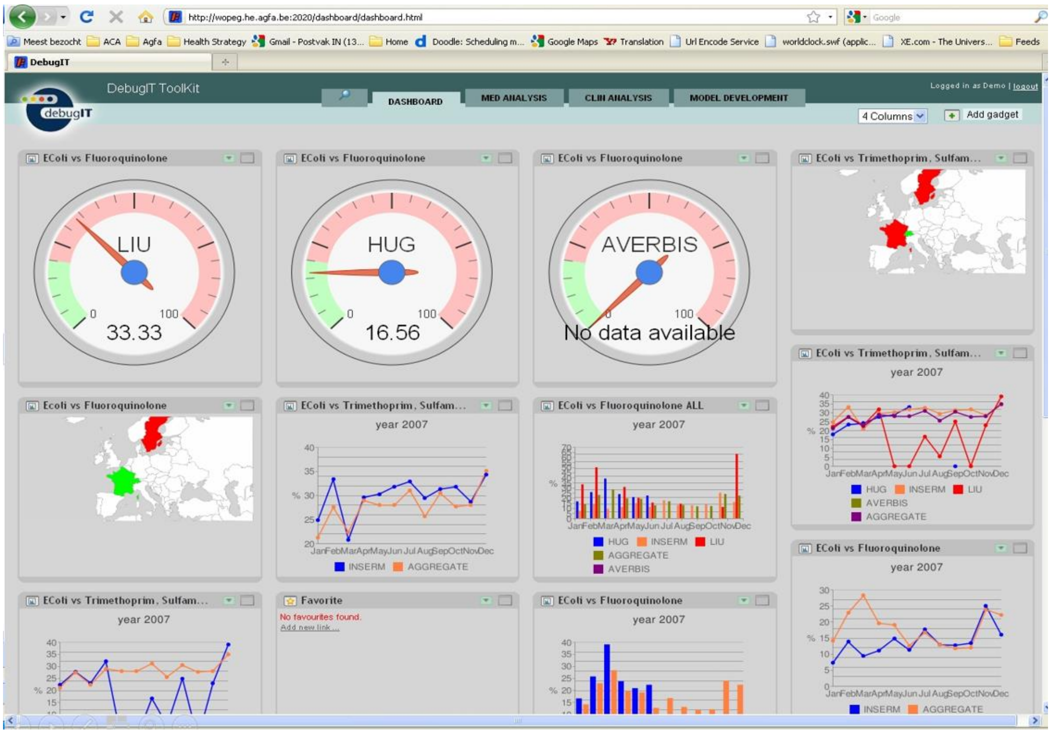


Fig. 3. DebugIT bacterial resistance monitor dashboard. Population monitoring is here build around a parametrisable dashboard, where individual visualization portlets, called gadgets, show the results of the CASQ SPARQL queries for the selected hospital sites at Linköping University Hospital (LIU), University Hospital of Geneva (HUG), University Clinic Freiburg (Averbis) and on selected additional variables. New gadgets can be dragged in, according to each user's needs and preferences.

5 Discussion

We have presented an ontology-based distributed data integration approach to serve the communication channel in the DebugIT EU project, hereby making antibiotics resistance data semantically and geographically interoperable. Although ontological data integration was achieved, semantic formalization had commenced in a stepwise manner. We showed how a bi-layered hybrid formalization approach can bridge the semantic gap between local RDF converted clinical data and the common formal integration layer. Domain ontologies representing the terminological domain of interest in a computer-interpretable way ensured a common interpretation, increased its robustness and suitability for secondary data usage. For the layer binding, we choose a rule-driven DDO to DCO mapping method, over SPARQL Construct-to-Where clause mappings. This decision was taken, although an existing limited performance analysis[14] highlighted SPARQL Construct-to-Where clause mappings [4] as the most performant binding method. A key argument in favor of the N3 rule-mapping approach was the envisioned real-time handling of high-throughput data volumes. In accordance with previous findings [14], processing of OWL axioms was considered as too slow for the envisioned real-time handling of large data volumes. Another reason for selecting N3 rules over SPARQL bindings was that rules were used within the remainder of the SIP already and the burden of writing correct rules could be alleviated partly by checking generated rules automatically.

5.1 Comparison to other ontology-based integration efforts

In [15] a knowledge base (KB) is described serving a rule-driven clinical decision support system (CDS) for guiding antibiotics prescription. Although its main scope is error-detection in patient-centric hospital care, its underlying ontology captures concepts overlapping with the DebugIT scope, i.e. 'antibiotics coverage range'. This CDS is however site-specific and only considers local medical data whereas the DebugIT System considers resistance-centric data from all over Europe and is set up in an extensible way allowing multiple new sites to participate in a seamless and scalable manner. Another difference is the data gathering method. Whereas for the CDS, the instance data was fed into the KB manually, in DebugIT we set up an automatic Extract, Transform, Load (ETL) data feed from the site-specific production databases, rendering the accessible data up-to-date over time. For the above reasons and due to the fact that the whole system is implemented as plugin for the Protégé 4 ontology editor, the CDSs' general setup is less complicated. On the downside, however medical doctors have to work with quite a complex tool and GUI, whereas in DebugIT these could be kept simple and easy as they were developed proprietarily [3] shielding the end-user from underlying complexity.

In its general set-up, our approach is similar to the OpenFlyData project[15] in that it uses semantic web technologies, integrating data for a specific domain. OpenFlyData also integrates distributed data via ETL, D2R servers and SPARQL endpoints and provides query templates. OpenFlyData however uses a single ontology layer for terminological mapping and tackles the entirely different domain of Fly gene mapping and expression analysis.

Regarding the implementation of the hybrid ontology layer approach the DebugIT SIP architecture closely resembles that of the NASA "SIMA: SemanticIntegrator for Mobile Agents" project[16], which integrated multiple heterogeneous sources via wrappers, a data source mediator and rule-enabled ontology integration.

As in the Advancing Clinico-Genomic Trials on Cancer (ACGT) project [16], which aims at improving post-genomic clinical trials by providing seamless access to

---

[14] comparing rules with OWL axiom and SPARQL query based cross-data mappings

[15] http://intranet.cs.man.ac.uk/dils09//presentations/2009_dils_flyweb.pdf, slide 8, last accessed 03.03.14

[16] http://ti.arc.nasa.gov/m/pub-archive/1221h/1221%20%28Keller%29.pdf, last accessed 03.03.14

integrated clinical, genetic, and image databases, we use information model-derived mediator artefacts and SPARQL to resolve syntactic and semantic heterogeneities when accessing wrapped databases. Like DebugIT, both the latter projects maintain separate SPARQL endpoint for each data source, which makes data updating easier than in data warehouse approaches. We wanted to be able to cope with dynamics and changes on all levels, e.g. updates to data, evolving schemas, and new CIS sources to be added. Classic DB based data warehouses were hence no option.

The functionalities of the web-based ResistanceMap[17] for interactive exploration of antimicrobial surveillance indicators are partly overlapping with that of the DebugIT dashboard. Whereas ResistanceMap mainly covers maps visualizing data from the U.S., our dashboard concentrates on European data. Both resources allow for national resistance comparisons, and provide graphics, which can be re-embedded on blogs or websites. However, besides DebugITs richer query answering capabilities, we believe our real-time surveillance can identify trends in pathogen incidence and antimicrobial resistance much faster than monitoring set-ups with larger time delays. An additional value of the DebugIT ontology based data integration is its eased secondary data usage and re-purposing.

5.2 Advantages and Limitations of our approach

The known overall drawbacks of our wrapper-mediator architecture are the query complexity and diminished source transparency. Although it is possible to achieve similar results using non-ontological ad hoc transformations of the source data, this kind of solution will always be context dependent and not explicit nor generic. Furthermore its results cannot be automatically proven to be correct, which is, at the end the most important issue in our case, as DebugIT attempts to enhance patient safety.

The DebugIT overall application scope, concentrating on clinical data, can still be seen as limited, as a comprehensive system would need tackle the domain in a holistic manner, i.e. need to cover veterinary resistance occurrences, as well as other, sometimes unexpected, but relevant hot spots of resistance occurrence, like drainages of pharmaceutical companies in developing countries [17]. With our research, momentarily we only show feasibility and practicability of the technical approach, before going for larger domain coverage. Providing direct evidence on the effectiveness of the general framework to patient centered outcomes must be the objective of future research.

An exemplary result comparison to existing surveillance efforts, - the German Paul Ehrlich Society (PEG) Monitoring and the French InVS surveillance studies, has shown evidence for good alignment of resistance trends shown by DebugIT with the ones indicated in the investigated surveillance efforts (see Supplementary Material). With respect to the reviewed surveillance set-ups, the DebugIT approach revealed the following key advantages:

---

[17] http://www.cddep.org/resistancemap, last accessed 03.03.14

- Most existing antibiotics monitoring systems suffer from low timely resolution. The DebugIT monitoring tools access hospital data in real-time, hence allowing for early trend discovery and opportunities for early interventions, as was demanded in [18].
- DebugIT is not limited to a certain selected set of bacteria or sampling methods, but potentially includes all bacteria and strains as found in the hospital data that can be mapped to a DCO/Uniprot taxon. This advantage can be generalized to a potentially higher overall data granularity, allowing for more concise data analysis.
- DebugIT stores data in a re-useable, semantically formal and traceable way, addressing the main reason for the current lack of information re-use namely data heterogeneity and inaccessibility. Current hospital data is often stored in proprietary and unconnected data silos and in diverse formats and languages.

The heterogeneity in the resistance measurement results in difficulties to compare international resistance data and introduces a blur into the integrated data quality. The main factor hindering harmonization of monitoring efforts is a lack of a commonly agreed definition for resistance and epidemiological cut-off values, as these values are compound and methodology dependent [19]. A look at an analysis on the validity and potential bias of surveyed resistance data [20] can prove useful to avoid inclusion of low impact data and will foster agreement what to include in the respective ontologies, rendering data validity more robust.

Besides the clinical queries to be solved as demanded by our clinical advisory board, the twelve identified quality indicators defined by the European Surveillance of Antimicrobial Consumption Network (ESAC-Net)[18] should serve validation. The importance of query contents for specific countries for example should be scored for their relevance on resistance and public health policies.

6 Conclusion

We have developed an IT framework to address the integrated management of clinical antibiotics data and detect patient safety related patterns and trends. To fight the epidemic spread of resistant pathogens due to citizen mobility we focus on EU wide and potentially global data inclusion. We have outlined an ontology-driven approach, leveraging on semantic web technologies, to create an open expandable and large-scale antibiotic resistance surveillance system across internet resources. We showed how ontologies provide computer-interpretable semantics and standard exchange syntax to be exploited for heterogeneous data integration, abstract querying, data comparison and reuse in an EU-wide antibiotics monitoring and alert system, easing secondary data usage. We showed in real world examples how cross-site data access, exchange and comparison is made possible via a three step layered integration approach, bridging from a local RDF representation over local data definition ontologies

---

[18] http://www.ecdc.europa.eu/en/activities/surveillance/ESAC-Net/publications/Documents/antimicrobial-consumption-ESAC-Net-reporting-protocol-2014.pdf

towards formal domain ontologies and rules as semantic integrators. The DebugIT ontologies were shown to drive a semantic interoperability platform to federate heterogeneous data from different hospital information systems into one formalized resource called ‘virtual CDR’. We exemplified feasibility of our layered data integration approach in an ‘ITbiotics’ application.

Although the presented multi-layered data integration approach is complex and requires considerable technological experience, we believe it is feasible. So far, it represents the only scalable solution, enabling seamless integration of additional hospital sites while maintaining local autonomy of present data sources.

In general, we have contributed to access the deep web[19], as we make hidden data web-searchable and we hope the DebugIT approach will serve as a model for a European bio-surveillance networks providing real-time monitoring tools.

7 Authors’ contributions

DS lead the DebugIT WP1a, developed the DebugIT core ontology, contributed to the SIP and prepared the manuscript. RC and MB contributed to the SIP, helped developing the DCO and contributed to the manuscript. KD implemented the DebugIT dashboard and contributed to the SIP architecture. FE and PD were responsible for the lexical mappings. DT, EP and CL implemented the Artemis monitor. All authors contributed to the manuscript.

8 Acknowledgements

This research is funded by the DebugIT project of the EU 7th Framework Program grant agreement ICT-2007.5.2-217139. We thank Hans Cools, Dirk Colaert and Giovanni Mels for their leading participation in the DebugIT project.

9 Conflict of Interest

We declare that there are no conflicts of interest.

10 References

1. Kaier K, Wilson C, Chalkley M, et al. Health and economic impacts of antibiotic resistance in European hospitals—outlook on the BURDEN project. *Infection* 2008; 36: 492–94.
2. Lovis C et al. DebugIT for patient safety - improving the treatment with antibiotics through multimedia data mining of heterogeneous clinical data. *Stud Health Technol Inform.* 136 (2008), 641-6

---

[19] https://en.wikipedia.org/wiki/Deep_Web, last accessed 03.03.14

3. Teodoro D, Pasche E, Gobeill J, Emonet S, Ruch P, Lovis C. Building a Transnational Biosurveillance Network Using Semantic Web Technologies: Requirements, Design, and Preliminary Evaluation. *Journal of Medical Internet Research.* 2012 May 29;14(3):e73.
4. Schober D, Boeker M, Bullenkamp J et al. The DebugIT core ontology: semantic integration of antibiotics resistance patterns. *Stud Health Technol Inform.* 2010;160(Pt 2):1060-4.
5. Martin L, Anguita A, Maojo V, Bonsma E, Bucur A, Vrijnsen J, Brochhausen M, Cocos C, Stenzhorn H, Tsiknakis M, Doerr M, Kondylakis H (2008) Ontology based Integration of Distributed and Heterogeneous Data Sources in ACGT. HEALTHINF 2008, Funchal, Portugal.
6. Teodoro D, Choquet R, Schober D, Mels G, Pasche E, Ruch P, Lovis C. Interoperability driven integration of biomedical data sources. *Stud Health Technol Inform.* 2011; 169:185-189
7. Stolba N, Towards a Sustainable DWH Approach for Evidence-Based Healthcare; PhD Thesis, Reviewers: A. Tjoa, T. Mück; Institut für Softwaretechnik und Interaktive Systeme, 2007; Rigorosum: 20.11.2007.
8. Linked Data - Connect Distributed Data across the Web, http://linkeddata.org/, last accessed 02.20.2012
9. Wache H, Voegele T, Visser U, Stuckenschmidt H, Schuster G, Neumann H, et al. Ontology-based integration of information-a survey of existing approaches. IJCAI-01 workshop: ontologies and information sharing. 2001. p. 108–17.
10. Assélé Kama A, Primadhanty A, Choquet R, Teodoro D, Enders F, Duclos C, u. a. Data Definition Ontology for clinical data integration and querying. *Stud Health Technol Inform*. 2012;180:38–42.
11. Wache H. Towards Rule-Based Context Transformation in Mediators. *EFIS Proceedings*. 1999. p. 107–22.
12. Handschuh S, Staab S, Volz R, On deep annotation, *Proceedings of the 12th international conference on World Wide Web*, 1-58113-680-3, Budapest, Hungary, 431-438, 2003, 10.1145/775152.775214, ACM
13. Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF. The Description Logic Handbook: Theory, Implementation and Applications. 2. Aufl. Cambridge University Press; 2008.
14. Boran A, Bedini I, Matheus CJ, Patel-Schneider PF, Keeney J, Choosing between Axioms, Rules and Queries: Experiments in Semantic Integration Techniques, in *Proceedings of the Eighth International Workshop on OWL: Experiences and Directions* (OWLED 2011), San Francisco, California, USA, June 5-6, 2011, 2011
15. Bright TJ, Furuya EY, Kuperman GJ, Cimino JJ, Bakken S. Development and Evaluation of an Ontology for Guiding Appropriate Antibiotic Prescribing. *J Biomed Inform*. February 2012;45(1):120–8.1.
16. Weiler G, Brochhausen M, Graf N, Schera F, Hoppe A, Kiefer S: Ontology based data management systems for post-genomic clinical trials within a European Grid Infrastructure for Cancer Research. *Conf Proc IEEE Eng Med Biol Soc* 2007, 2007:6435-6438.
17. Kristiansson E, Fick J, Janzon A, Grabic R, Rutgersson C, et al.: Pyrosequencing of Antibiotic-Contaminated River Sediments Reveals High Levels of Resistance and Gene Transfer Elements. 2011 *PLoS ONE* 6(2): e17038. doi:10.1371/journal.pone.0017038
18. Giske CG et al., Supranational surveillance of antimicrobial resistance: The legacy of the last decade and proposals for the future, *Drug Resistance Updates*, Volume 13, Issues 4–5, August–October 2010, Pages 93–98 doi:10.1016/j.drup.2010.08.002

19. Silley P, de Jong A, Simjee S, Thomas V, Harmonisation of resistance monitoring programmes in veterinary medicine: an urgent need in the EU?, *Int J Antimicrob Agents*. 2011 Jun;37(6):504-12. Epub 2011 Feb 3
20. Rempel OR, Laupland KB. Surveillance for antimicrobial resistant organisms: potential sources and magnitude of bias. *Epidemiol Infect.* 2009 Dec;137(12):1665-73. Epub 2009 Jun 4. Review. PMID: 19493372