eHealth Beyond the Horizon – Get IT There   641
S.K. Andersen et al. (Eds.)
IOS Press, 2008

DebugIT for Patient Safety – Improving the Treatment with Antibiotics through Multimedia Data Mining of Heterogeneous Clinical Data

Christian LOVIS[a,1], Dirk COLAERT[b], and Veli N. STROETMANN[c]

[a]University Hospitals of Geneva, University of Geneva, Switzerland
[b]Agfa HealthCare, Belgium
[c]empirica GmbH, Germany for the DebugIT consortium

Abstract. The concepts and architecture underlying a large-scale integrating project funded within the 7th EU Framework Programme (FP7) are discussed. The main objective of the project is to build a tool that will have a significant impact for the monitoring and the control of infectious diseases and antimicrobial resistances in Europe; This will be realized by building a technical and semantic infrastructure able to share heterogeneous clinical data sets from different hospitals in different countries, with different languages and legislations; to analyze large amounts of this clinical data with advanced multimedia data mining and finally apply the obtained knowledge for clinical decisions and outcome monitoring. There are numerous challenges in this project at all levels, technical, semantical, legal and ethical that will have to be adressed.

Keywords. Infectious disease, patient safety, semantic inter-operability, multimedia data mining, decision support, clinical outcome monitoring

Introduction

Building a safer and more efficient care system has become the most shared goal of all actors involved in healthcare. From a historical perspective, there has been an impressive shift towards awareness of the impact of errors in medicine in the last 25 years. In the early nineties, research papers and reports about patient safety, incident reporting and initial order-entry systems were published, mostly originating from academic settings. At about the same time, the first reports of the US Institute of Medicine (IOM) on computerized patient record systems stressed the ability of ICT-based solutions to improve the quality of care [1]. Ten years later, by the end of the nineties, a famous report of the IOM called attention to the wide prevalence of errors in healthcare [2]. While medical errors are under the spotlight, (re-)emerging infectious

---

[1] Corresponding Author: Service of Medical Informatics, University Hospitals of Geneva; E-mail: christian.lovis@hcuge.ch

diseases are becoming major challenges. Among them, the rapid development of anti-microbial resistances [3], the spread of nosocomial and other infections [4], the inadequate care and missing appropriate tools to lead the care system facing these new emergent problems [5] are major concerns. The issues around infectious diseases are strongly interrelated and have immediate and important effects on safety, quality of care and efficiency. In half a century of antibiotic use, new challenges have emerged: fast emergence of resistances among pathogens, misuse and overuse of antibiotics. Antimicrobial resistance results in escalating healthcare costs, increased morbidity and mortality and the (re-)emergence of potentially untreatable pathogens.

1. The project

Dedicated to infectious diseases, the DebugIT (Detecting and Eliminating Bacteria UsinG Information Technology) project aims at (1) detecting patient safety related patterns and trends, (2) acquiring new knowledge and (3) using this for better quality healthcare. A consortium of eleven partners has been built in order to gather scientific competencies in all domains involved, as well as to assure access to specific information of more than 2 millions clinical records.

Table 1. Consortium

Belgium	Agfa HealthCare
Bulgaria	GAMA/SOFIA Ltd.
Czech Republic	IZIP - Internetový Přistup Ke Zdravotním Informacím Pacienta
France	Institut National de la Santé et de la Recherche Médicale (with European Hospital George Pompidou)
Germany	empirica GmbH
Germany	University Medical Center Freiburg
Greece	Technological Educational Institute of Lamia
Sweden	Linköping University
Switzerland	University Hospital of Geneva
Switzerland	University of Geneva, Computer Science Department
United Kingdom	University College London

The project has a strong clinical lead guaranteed by a Clinical Advisory Board and a Scientific Advisory Board with European and American experts of the infectiology field and the scientific domains involved.

Outcomes and benefits, in clinical and socio-economic terms, will be measured. Results will be integrated into Clinical Information Systems (CIS) of participating European hospitals, industry and their clients, and become available globally through a European or global Disease Control Centre/Public Authority, also as Open Source solution. Advanced ICT applications and innovations concern the virtualization of the Clinical Data Repository through ontology and terminology binding and mediation, advanced data mining techniques, the use of machine reasoning related to real, point-

of-care patient data, as well as consolidation of all these techniques in a comprehensive but open framework. Output will be applicable to other clinical fields.

The concept developed as foundation for the DebugIT project addresses all of these issues in an operational manner with the ultimate goal to develop a new, highly advanced and pre-eminent tool aiming at producing a new and efficient weapon for the war against infectious pathogens across all health system actors and levels.

The overall project outcome will not only be a theoretical work and proof of concept, but also a practical implementation of a highly improved and advanced computerised system in the field of infectious disease treatment and antibiotics usage. This application, which, due to its generic conceptual base, should be easily expandable and adaptable to other similar medical application fields, will initially be evaluated by participating project partners, but should be made publicly available to other healthcare organizations soon after.

2. The Conceptual Framework

The conceptual framework of this project is an ever continuing iterative cycle, implementing the principle of translational medicine and true Evidence Based Medicine (EBM). Translational medicine makes the connection between medical research and clinical care by providing to research clinical data and providing the results of the research – the medical knowledge – as input for clinical care. While medical research is often focusing on prospective and tightly controlled studies, retrospective studies with access to huge amounts of data, just waiting to be analyzed, are a welcome addition to clinical research.

The framework can be broken down into several distinct steps (figure 1):

1- **Collect Data**. Clinical data will be collected and aggregated across different hospitals, countries, languages, information models and legislations, via advanced and commonly agreed data models (minimal data sets), standards and mapping algorithms.

2- **Learn**: Advanced data mining techniques on multimodal, multi-source, structured and unstructured data will detect patterns, relevant for patient safety and the treatment of infectious diseases such as: resistance of bacterias, adverse events and operational practices. This will result in new knowledge and new evidences for existing knowledge.

3- **Store Knowledge**: This knowledge will be stored, visualized, validated and aggregated together with pre-existing medical and biological knowledge (guidelines, regulations) to achieve a consolidated view on the needed knowledge, to be applied in the next step

4- **Apply**: Software tools will be integrated into the available clinical and public health information systems. Decision support tools will apply the generated knowledge and help the clinician to provide clinical care (choice, dose and administration of antibiotics for example). The knowledge will also be used to monitor the ongoing care activities and even predict future outcome to give additional feedback, both on individual patient and cohort level. This will allow healthcare providers and decision makers to take appropriate actions at various level of the healthcare system, including point-of-care, management or policy, and subsequently influence the future development of our health systems. Integration in existing CIS will enable

to record activities and results and thus make sure the necessary data are generated for a next cycle.
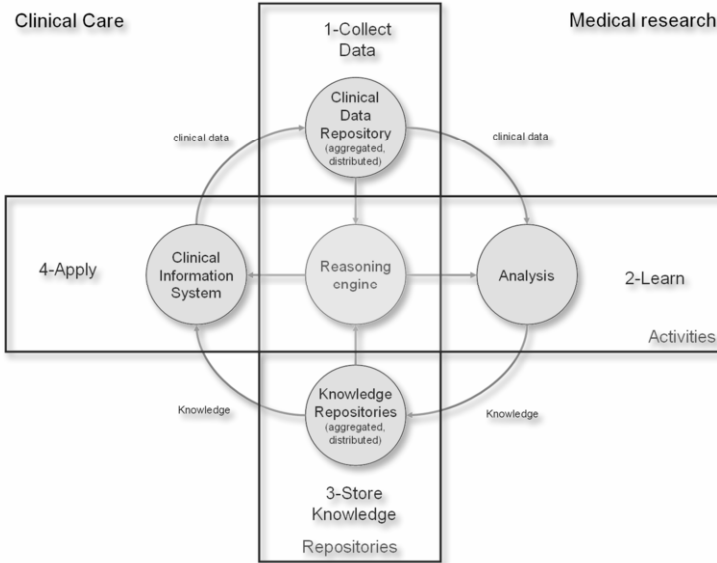

Figure 1. The Conceptual Framework

3. Technology and Architecture

To achieve his goals, the DebugIT project will make extensive use of clinical and operational information originating from running Clinical Information Systems (CIS) across the EU. The processes and entities are described in figure 2. Clinical data are accessible through a virtualized, fully integrated Clinical Data Repository (CDR).
The CDR will feature transparent access to the original CIS and provide data aggregations in local stores. The CDR is specifically tailored for knowledge discovery, featuring ethically sound, transparent access to data at or from the original CIS and/or collection and aggregation of data in a local data store.

Multimodal Data Mining (MDM) will have a strong focus on new fields of research doing mining on distributed storage, using highly advanced new text, image and structured data mining on individual patients as well as on populations.

New knowledge will be fed into a Medical Knowledge Repository (MKR) and mixed with domain knowledge coming from external sources (guidelines and scientific evidence). Innovative and user friendly knowledge representation paradigms will be developed in order to enable not only knowledge engineers but also clinicians to use the repository

After validating, this knowledge will be used by a decision support module (DSM) and monitoring tool in the clinical environment to prevent patient safety issues and report on them, both at the population and at the patient level for direct care.

Co-ordination and steering of both the analysis and the care process will be done by a performant and versatile reasoning engine.
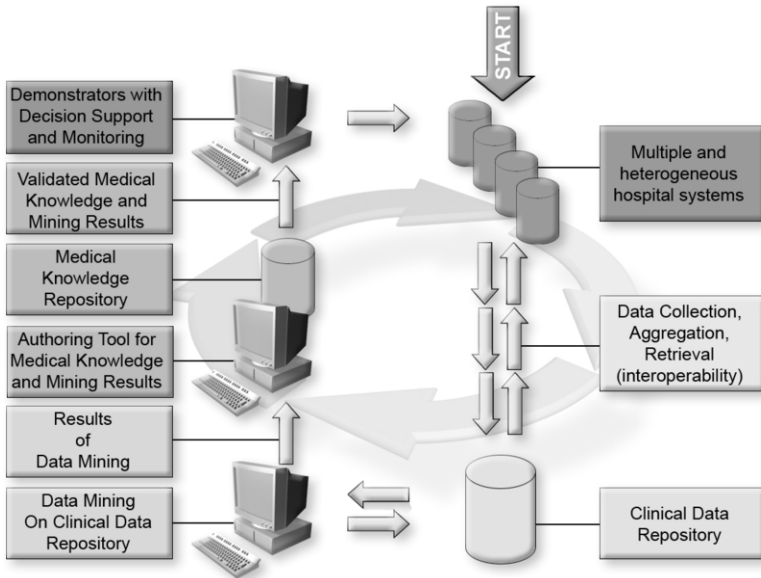
Figure 2: Overview of the project

4. Privacy

Strong attention is given to privacy concerns, taking into account the various legal and ethical frameworks to be met. Therefore, privacy is made a central part of the project *by design,* using a virtualized data repository without dealing directly with the original data. Identification elements provided by the clinical data repositories can be carried all along the process, blindly, in order to allow the original clinical information system to feed back decision support without need for patient identification.

5. Conclusion

There are huge challenges in the concepts and architecture, needing new and original solutions. Building a shared and distributed data repository with a common semantic infrastructure, dealing with privacy, etc, which will have to be addressed? The infectious model disease has been chosen as it allows addressing these issues with a pretty well defined domain of medicine. Infectious diseases have commonly fast course, identified causes and symptoms and well-defined treatments.

The DebugIT project is focused on using large existing and heterogeneous clinical datasets covering hundreds of thousands of patients from several clinical information systems in different European countries. DebugIt proposes to build an interoperability platform to populate a pertinent dataset about the infectious domain to achieve a very large common shared virtualized clinical repository that enables knowledge-driven data mining. This "semantic mining" will be based on innovative methodologies to deal with the characteristics of real world clinical data. A knowledge repository will drive

the data mining and serve as storehouse for the results. Finally, a decision support engine will exploit the aggregated knowledge to loop-back to the real world.

To achieve this system, several aspects will have to reach the frontiers of current state-of-the-art and beyond. Two strategies can be chosen for that. The first one is to invent something radically new. The second one consists of using all existing knowledge and methods, putting them together, and trying to build upon this base. For most of its research, the second strategy is the one chosen in this project, because operational results for clinical information systems must be available and sustain the DebugIT outcomes after the end of the project.

In order to meet these requirements, the project has been organized according to architectural component-based considerations:

- Interoperability Platform (IOP);
- Clinical Data Repository (CDR);
- Multimodal Data Mining (MDM);
- Medical Knowledge Repository and associated Knowledge Authoring Tool (MDR)
- Decision Support and Monitoring engine (DSM);
- Clinical applications.

This scientific and technical framework, associated with access to large amounts of clinical databases and led by experts in the medical field will lead to a serious advance in building a large IT infrastructure aiming at creating new knowledge in the field of monitoring, surveillance and efficient measures to fight infectious diseases.

6. References

[1]   IOM. The computer-based Patient record: An essential technology for health care. Institute of Medicine report. 1991, revised 1997.
[2]   Medicine Io, editor. To Err is Human. Building a safer Health System; 1999. National Academy Press.
[3]   Frazee BW. Update on emerging infections: news from the Centers for Disease Control and Prevention. Severe methicillin-resistant Staphylococcus aureus community-acquired pneumonia associated with influenza--Louisiana and Georgia, December 2006-January 2007. Ann Emerg Med. 2007 Nov;50(5):612-6.
[4]   O'Brien J A, Lahue BJ, Caro JJ, Davidson DM. The Emerging Infectious Challenge of Clostridium difficile-Associated Disease in Massachusetts Hospitals: Clinical and Economic Consequences. Infect Control Hosp Epidemiol. 2007 Nov;28(11):1219-27.
[5]   McAlonan GM, Lee AM, Cheung V, Cheung C, Tsang KW, Sham PC, et al. Immediate and sustained psychological impact of an emerging infectious disease outbreak on health care workers. Canadian journal of psychiatry. 2007 Apr;52(4):241-7.