CONCEPT

DebugIT works complementary to approaches such as applied in life sciences, molecular biology or health policy to reinforce and support the war against infectious pathogens. This includes the detection and prevention of resistances, as well as improved management strategies and healthcare workflow. Everyday, large amounts of data is generated, collected and stored in numerous clinical environments. However there is only very little secondary use of this information to improve the quality and safety of care. While the amount continues to grow, there is urgent need to develop a comprehensive system to manage and understand this data, including influencing the way it is collected, stored, shared and analysed. DebugIT specifically addresses many of the present obstacles to exploiting infection management data, including

1. **The lack of technical interoperability:** Integrating proprietary and heterogeneous clinical and non-clinical databases and systems from distributed sources is still a highly challenging task.
2. **The lack of semantic interoperability** is of even greater concern. Various pieces of information and data must have the same meaning in order to be analysed together.
3. **The missing transparency about quality and reliability:** Any analysis of the data should imply the accessibility of information beforehand which characterises the source. This will allow the assessment of its quality, reliability, timeliness and relevance.
4. **The low quality:** The intrinsic character of real-world clinical data is full of missing values, errors and noise. This is rarely taken into account in limited research studies.
5. **The multimedia barrier** and the ability to merge several different sources and modalities of information. One of the complex characteristics of data in life sciences is the multiplicity of media types, ranging from numbers and simple text to images to audio and video.
6. **The contradicting stakeholder interests:** This is a rarely mentioned, though important barrier. This is due to the many different and sometimes opposite interests of stakeholders active in this domain, especially when it comes to operational analysis, such as resource usage, efficiency and costs.
7. **The security, privacy and confidentiality barriers:** the power gained in aggregating data from many sources can lead to risks for the patient's or the citizen's privacy rights.

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.

We are improving the detection of harmful patterns and trends in clinical and operational information from Clinical Information Systems (CIS). This is done through an analytic 'view' of a virtualised Clinical Data Repository (CDR) specifically tailored for knowledge discovery and featuring ethically sound transparent access to data. Text, image and structured data mining from individual patients, as well as populations are analysed. The new knowledge generated feeds a Medical Knowledge Repository and is being systematically combined with existing information from external sources, such as clinical guidelines and medical evidence. After editing and validating, this knowledge is used in the clinical environment. This is improving issues pertaining to infectious diseases which prevent patient safety issues.

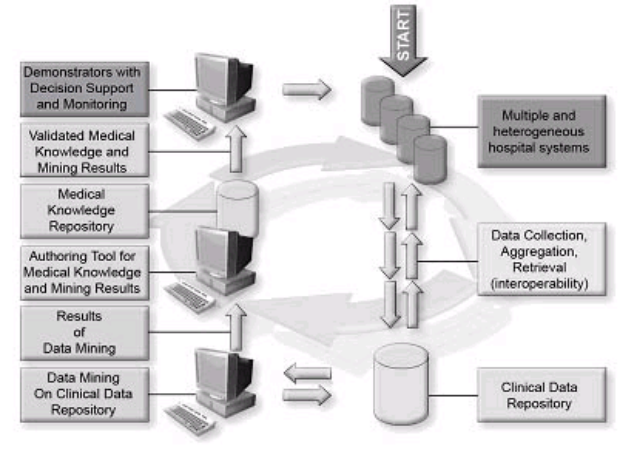

Figure 1 Overview of the project

Information technology is a far-ranging domain, extending from educational aspects covered by eLearning to communication and policy implementation issues tackled by eGovernment tools. The conceptual framework of this project will focus primarily on the following three (iterative) process steps (each with two sub steps):

- **Aggregation:**
  - **Data:** Integrate data from a wide variety of clinical sources and feeder systems, available

in various modalities, via advanced and commonly agreed data models (minimal data sets), standards and mapping algorithms.

- **Knowledge:** Consolidate multiple knowledge sources, including medical, workflow and public health knowledge, to standardize and simplify access and use.

- **Research and analysis:**

  - **Detect:** Identify safety-relevant events, issues and patterns based on multi-media and multi-source analyses of (integrated, merged) clinical databases, using advanced statistical methods.
  - **Learn:** Obtain new hypotheses and evidence by validating the results obtained from the preceding step; merge these with existing medical and biological knowledge and information in order to build new consolidated knowledge and decision support, usable at various stages, from research to clinical care, hospital management, and Public Health.

- **Innovation and diffusion:**

  - **Implement:** Link the new eHealth tools and decision support system developed within the research and analysis framework and integrate it into the available clinical and public health information systems. 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.
  - **Monitor and predict:** Build new monitoring systems to help hospitals prevent and fight sources of infections, to support public health policy makers in managing the political healthcare process and to help industries in their ability to provide and foresee adequate answers in the fight against infectious pathogens.

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 diseases 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 organisations soon after.

The conceptual approach as described earlier is illustrated in Figure 2. It will serve as the basic framework for developing the scientific and technical work flow of this project. The prevention of adverse events is the overall goal of this project. This is indicated by the outer most circle that embraces the three "pillars" of Aggregation, Research & Technological Development and Innovation. Each of these three steps in an iterative process of system and therefore quality improvement is further subdivided into two sections, depicted by six segments. Data collection and knowledge discovery are part of the aggregation step, pattern detection and evidence learning underpin RTD and finally, application implementation and event monitoring form the innovation process.

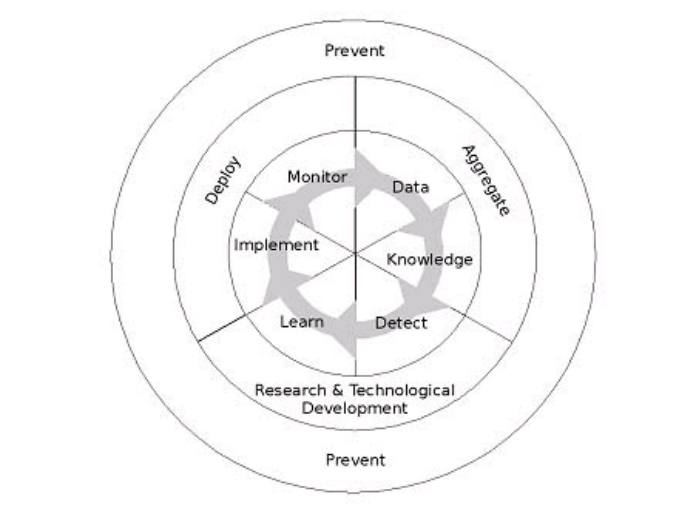

Figure 2 Generic conceptual approach

Data that are important for the discovery of clinical events, such as adverse drug events are typically collected and stored in Clinical Information Systems (CIS) maintained by multiple institutions such as hospitals or other national healthcare providers. In general, a varying number and combination of heterogeneous techniques and technologies is being used, including proprietary as well as open-source based platforms, hard- and software. In order to establish a coherent and systematic data exchange, a rich data set is needed, harmonised across the DebugIT sites and CIS systems, that defines required data items for the medical domain targeted, including information about patients, pathogens, drugs and workflow aspects, as well as their syntax and semantic. DebugIT will adopt a three-stage methodology to this challenge.

1. Specific interfaces will convert from the proprietary data set to standards-based formats of the **Interoperability Platform**. The Interoperability Platform provides a de-identification process as well as cleaning of the data from artefacts. The results are used to populate the Clinical Data Repository.

2. The **Data Mining Engine** processes data from the Clinical Data Repository together with fed-back data from the Medical Knowledge Repository to create new/additional medical knowledge which is being stored in the Medical Knowledge Repository.
3. The **Decision Support System** links health observations from the Clinical Information Systems with health knowledge from the Medical Knowledge Repository to support the fact finding process on a patient, at a local, regional, national or international level.

This three-stage approach provides the project with great methodological data while keeping empirical power for further improving safety issues and risk management in this specific domain. For example, the "beyond the state-of-the-art" multimedia surveillance system will support the discovery of unsuspected, helpful patterns of nosocomial infections and antimicrobial resistance.

A key aspect of this new system will be to support improved antimicrobial prescribing and quality of care. This will translate into medico-economic efficiency and a decrease in antimicrobial selection pressure on the acquisition and transmission of multi-resistant pathogens.

Both in- and outpatient clinical data sources are available within the project, extending the scope beyond the hospital environments. This increases the power to develop generic methods and tools. As a consequence, the developed system will be easily adaptable to other medical and clinical topics and environments.