Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth
A. Ugon et al. (Eds.)

doi:10.3233/978-1-61499-852-5-875

Implementing Systems Medicine: A Medical Informatics Perspective

Matthias GANZINGER[a,1], Matthias GIETZELT[a], Christian KARMEN[a], Blanca FLORES[a] and Petra KNAUP[a]
[a]*Heidelberg University, Institute of Medical Biometry and Informatics, Heidelberg, Germany*

Abstract. Systems medicine is a paradigm for translating *in silico* methods developed for modelling biological systems into the field of medicine. Such approaches rely on the integration of as many data sources as possible, both in the dimension of disease knowledge and patient data. This is a challenging task that can only be implemented in clinical routine with the help of suitable information technology from the field of Medical Informatics. For the research project "Clinically-applicable, omics-based assessment of survival, side effects, and targets in multiple myeloma" (CLIOMMICS) we developed a prototypical systems medicine application system. It is based on a three-level-architecture covering data representation, decision support, and user interface. The core decision support component is implemented as a case-based reasoning engine. However, the architecture follows a modular design that allows to replace individual components as needed.

Keywords. Systems medicine, medical informatics, clinical decision support, case-based reasoning

1. Introduction

In the last years, the development of systems medicine has made progress. Many research projects were performed with focus on the translation of systems biology approaches towards clinical application. Thus, one definition of the term *systems medicine* is: "[…] the application of systems biology approaches to medical research and medical practice. Its objective is to integrate a variety of biological/medical data at all relevant levels of cellular organization using the power of computational and mathematical modelling, to enable understanding of the pathophysiological mechanisms, prognosis, diagnosis and treatment of disease." [1] However, this transformation has to be paralleled in the advancement of clinical information systems. In this paper we present the results of a systems medicine research project, as part of which a prototype information system was developed.

Medical Informatics research was performed as the sub-project *Multi-level data management and IT architecture* of the project *Clinically-applicable, omics-based assessment of survival, side effects, and targets in multiple myeloma (CLIOMMICS)*. The

[1] Corresponding Author, Matthias Ganzinger, Heidelberg University, Institute of Medical Biometry and Informatics, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany; E-mail: matthias.ganzinger@med.uni-heidelberg.de.

overall aim of the project was to improve prognostication and response to treatment by using novel molecular data. Multiple Myeloma (MM) is a cancer of plasma cells producing monoclonal antibodies and accumulating in bone marrow. Incidence of MM is 4 to 6 per 100.000 people per year with a median age at diagnosis of 65 to 70 years [2]. The specific aim of the Medical Informatics subproject is to research decision support systems and building a prototype for the clinical context. In this paper we will show our approach of implementing such a systems medicine application prototype.

2. Methods

For the data and information and technology management subproject of CLIOMMICS we defined the following six work packages:

1. Data protection
2. Architecture for Multiple Myeloma system medicine
3. Standard Data Preparation Procedure
4. Visualization Concept
5. Prototypical Information System
6. Archiving Concept

These work packages were planned for a time frame of three years and an extension of two years in total.

3. Results

3.1. Data protection

In the first step, the data sets provided by the project partners were analyzed to determine protection requirements for the whole system. As a result, a base cohort of approximately 3.500 patients was considered for systems medicine research, but an additional level of de-identification was deemed necessary. Thus, a new random identification number was assigned to all patients resulting in two-level pseudonymization.

3.2. Architecture for Multiple Myeloma system medicine

As a generic architecture for systems medicine software systems, we designed a modular three-level architecture as shown in Figure 1 [3]. In level one, data are collected and made available for systems medicine. In our project we use a research data warehouse based on the open-source software i2b2 [4] as the basic central data store.

The second level is comprised of components for decision support. We primarily rely on the concept of patient similarity that has been leveraged for decision support systems in the well-known approach of case-based reasoning (CBR) [5, 6]. In our system, the case base is provided by the data warehouse. Of great importance for the appropriate quantification of similarity between patient cases is the selection of a suitable similarity measure for each attribute type. Since in our project case information is derived from clinical trial data to a great extent, we developed novel similarity measure that is based on survival data (corresponding paper under review). In addition to the experience-based

approach of CBR, rule-based modules can also be added to level two and as a result, the architecture is capable of a hybrid decision support solution [7].

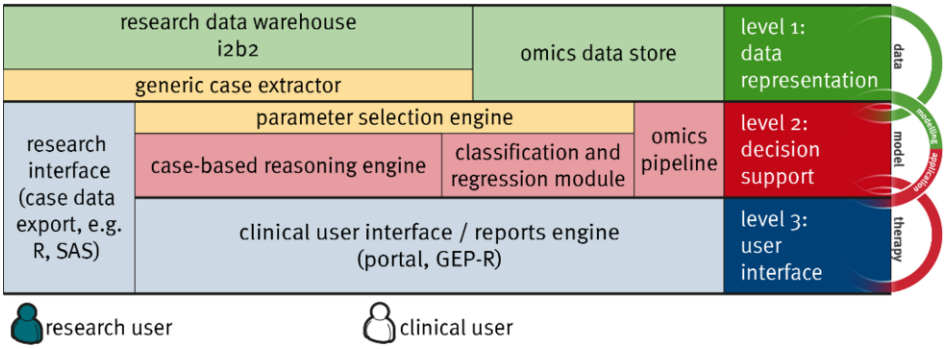

Figure 1. CLIOMMICS three layer architecture

Level three is dedicated to the development of user interfaces for both researchers and clinical users. For researchers, we provide access to the research data warehouse which can be used to select a specific sub-cohort based on the attributes available. Corresponding data for the selected cases can be extracted and analyzed with tools such as R [8]. For clinicians, we offer two types of user interface:

- A report about the patient's prognosis is generated. It can be used like other letters commonly exchanged among health care sectors [9].
- A prototype of an interactive, web-based user interface is provided for accessing results of the decision support information generated in level two.

The module slots of the architecture have been filled with sample components to provide a running prototype of a complete systems medicine application stack.

3.3. Standard Data Preparation Procedure

Our case base in the data warehouse is designed for continuous growth as new patient cases are added and follow-up information for existing patients is becoming available. Further, data originate from different data sets with differing data definitions. Consequently, data need to be harmonized to establish a set of attributes for describing cases that is as comprehensive as possible. We applied a generic data harmonization process, resulting in a harmonized set of common data elements in the form of a harmonization table [10]. Since this table is both human and machine readable, it is used for automated configuration of extract, transform, and load processes (ETL). With these ETL processes, data have been initially loaded into the core data warehouse. Follow-up data are imported using the same processes without change. Additional data sources can be added in the future by extending the harmonization table and generating the corresponding ETL processes.

3.4. Visualization Concept

Visualizing complex systems medicine information with respect to the clinical needs is crucial for successfully implementing such a system. These requirements are closely

related to those for statistical graphics. Among others, graphics should make large data sets coherent, reveal the data at several levels of detail, and be closely integrated with statistical descriptions of a data set [11]. To allow for a flexible user interface, have implemented the web portal software Liferay as a framework for visually integrating different modules for decision support as so-called portlets [12]. One portlet is used for generating the CLIOMMICS report as a PDF document to access patient data. Other portlets are designed to provide interactive access to the results, such as recommendations of the case-based reasoning system. In terms of technology we use client-side JavaScript libraries for interactive graphics generation such as D3 [13].

3.5. Prototypical Information System

Currently, all modules shown in our architecture for systems medicine are integrated into a prototype of a systems medicine information system. With this tool, first tests in a clinical context are possible. Usability of the system for clinical practice can be assessed and user feedback will be the basis for future developments.

3.6. Archiving Concept

All data that are handled by our ETL processes are also archived. For reproducibility of results, data are captured daily using a virtual snapshot concept that only copies new or changed data blocks and only references the remaining ones for efficient use of storage.

For bioinformatics pipelines used for analyzing RNA-seq data we developed a concept for conserving the analytical steps in addition to the data themselves [14]. The complete pipeline with all software components is stored in a snapshot that can be reactivated to reproduce results.

4. Discussion

Systems Medicine is often considered as the transfer of methods from Systems Biology to the level of patient treatment. Such approaches are valuable to gain a better understanding of diseases by models describing disease-specific mechanisms. Beyond that, Systems Medicine strives for an individualized prognosis and treatment of patients on the basis of heterogeneous data including omics data as well as phenotype data, patient lifetime environment or individual preferences. Complex diseases like cancers are especially suited to be addressed by Systems Medicine, since reference cohorts and individual patients have to be described in many dimensions. Therefore, Systems-Medicine-based patient care is only possible with the help of appropriate information technology (IT). Such IT systems leverage Systems Biology models by applying them on an individual patient's data.

Integrated representation of disease related data, as shown in our application, is only the first step in the pipeline of an IT system for Systems Medicine. These data have to be made available for the clinical decision making process to bring benefits for individual patients. IT-based clinical decision support systems (DSS) can be used for this task, but so far they usually do not cover all available data types. In our project we developed a generic IT architecture for Systems Medicine. As replaceable DSS component we investigate the application of a CBR system on Multiple Myeloma data including phenotype and omics data. We prepared a phenotype case base and achieved first results

on using gene expression data with CBR. Our architecture is ready to be transferred to other disease areas and most of the tools we have developed are already available as open-source software.

As next steps we will investigate further similarity measures to optimize the identification of similar patients. Correspondingly, we plan to research if treatment recommendations can be improved by applying data mining methods to sub cohorts of the case base similar to a patient. Finally, we plan to develop a new validation approach for complex DSS based on artificially designed data sets.

Acknowledgements

CLIOMMICS is funded by the German Ministry of Education and Research within the e:Med initiative. Grant id: 01ZX1609A.

References

[1] C. Auffray, R. Balling, M. Benson, M. Bertero, H. Byrne, M. Cascante et al., From Systems Biology to Systems Medicine; Workshop report 14-15 June 2010. Brussels: European Commission, DG Research, Directorate of Health; 2010. Available from: URL:http://ec.europa.eu/research/health/pdf/systems-medicine-workshop-report_en.pdf
[2] T. Moehler, H. Goldschmidt, Multiple Myeloma, Springer-Verlag Berlin Heidelberg. Berlin, Heidelberg, 2011. (Recent Results in Cancer Research; vol 183)
[3] M. Ganzinger, M. Gietzelt, C. Karmen, D. Firnkorn, P. Knaup, An IT Architecture for Systems Medicine. Stud Health Technol Inform 210 (2015), 185–189
[4] T. Ganslandt, S. Mate, K. Helbing, U. Sax, H.U. Prokosch, Unlocking Data for Clinical Research - The German i2b2 Experience. Appl Clin Inform 2 (2011), 116–127
[5] S.-A. Brown, Patient Similarity: Emerging Concepts in Systems and Precision Medicine. Front Physiol 7 (2016)
[6] A. Aamodt, E. Plaza, Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI communications 7 (1994), 39–59
[7] M. Ganzinger, C. Haux, C. Karmen, T. Wetter, P. Knaup, Systems Medicine for Multiple Myeloma: A Review on Decision Support Systems. Stud Health Technol Inform 216 (2015), 951
[8] R Core Team, R: A Language and Environment for Statistical Computing. Vienna, Austria, 2016. Available from: URL:https://www.R-project.org/
[9] T. Meissner, A. Seckinger, T. Rème, T. Hielscher, T. Möhler, K. Neben et al., Gene expression profiling in multiple myeloma—reporting of entities, risk, and targets in clinical routine. Clin Cancer Res 17 (2011), 7240–7247
[10] D. Firnkorn, M. Ganzinger, T. Muley, M. Thomas, P. Knaup, A Generic Data Harmonization Process for Cross-linked Research and Network Interaction. Methods Inf Med 54 (2015), 455–460
[11] E.R. Tufte, The visual display of quantitative information, Graphics Press. Cheshire, Conn., 2001
[12] R. Sezov, Liferay in action: The official guide to Liferay Portal development, Manning. Shelter Island, NY, 2012
[13] D3: Data-Driven Documents [cited 2017 Nov 7]. Available from: URL:https://d3js.org/
[14] B. Flores, D. Hose, A. Seckinger, P. Knaup, M. Ganzinger, From Bench to Bedside: a View on Bioinformatics Pipelines. Stud Health Technol Inform in print (2017)