e-Health – For Continuity of Care
C. Lovis et al. (Eds.)

doi:10.3233/978-1-61499-432-9-657

User Needs and Requirements Analysis for Big Data Healthcare Applications

Sonja ZILLNER[a,b,1], Nelia LASIERRA[c], Werner FAIX[b], and Sabrina NEURURER[c,d]
[a] *Corporate Technology, Siemens AG, Munich, Germany*
[b] *School of International Business and Entrepreneurship, Steinbeis University, Berlin, Germany*
[c] *Semantic Technology Institute Innsbruck, University of Innsbruck, Austria*
[d] *Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Austria*

Abstract. The realization of big data applications that allow improving the quality and efficiency of healthcare care delivery is challenging. In order to take advantage of the promising opportunities of big data technologies, a clear understanding of user needs and requirements of the various stakeholders of healthcare, such as patients, clinicians and physicians, healthcare provider, payors, pharmaceutical industry, medical product suppliers and government, is needed. Our study is based on internet, literature and market study research as well as on semi-structured interviews with major stakeholder groups of healthcare delivery settings. The analysis shows that big data technologies could be used to align the opposing user needs of improved quality with improved efficiency of care. However, this requires the integrated view of various heterogeneous data sources, legal frameworks for data sharing and incentives that foster collaboration.

Keywords. Big data, user needs, requirements analysis

Introduction

Recent publications [1-4] highlight the potential impact of big data applications and health data analytics for the healthcare sector. However, as of today, big data technology is rarely used within health care settings. The aim of this paper is to present and discuss the results of a big data related *user needs and requirements analysis* that was accomplished in the course of a big data technology road map development for the healthcare domain. Big data technologies establish means to aggregate and analyze large sets of heterogeneous data sources, and allow generating new insights and knowledge. Thus, one essential step for developing a technology road map for big data in the healthcare domain is to determine the particular information needs of the involved user groups. The context of our work is the Big Data Public Private Forum[2] Project, which aims towards developing a technology road map for big data technologies for the European healthcare market[3].

[1] Corresponding Author: sonja.zillner@siemens.com

[2] http://www.big-project.eu/

3 Beside the healthcare sector, several other industrial sectors, such as Energy, Transport, Finance, Manufacturing, Retail and the Public Sector are addressed within the project but not focus of this publication.

The goal of our study was to identify the most relevant information needs of the various stakeholder groups in the healthcare delivery domain. They could be addressed by aggregating and analyzing large and heterogeneous data sets. Particular requirements that need to be addressed in order to foster the implementation of big data applications in the healthcare domain were identified as well.

1. Methods

As depicted in Figure 1, our study consisted of two different stages.

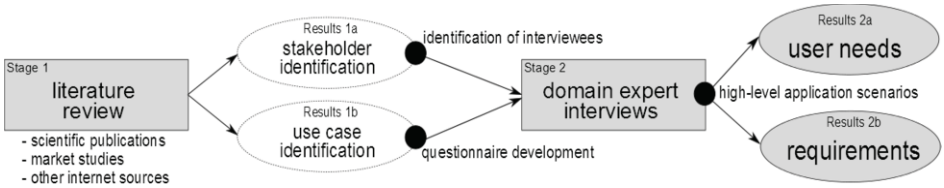

Figure 1. Two-stage approach of the presented study.

The goal of the first stage was to identify both stakeholders and use case applications of big data. Therefore, we conducted a literature review including scientific reviews, market studies and other internet sources. This knowledge allowed us to identify and select potential interviews partners, and to develop the questionnaire for the domain expert interviews. The interview guide consisted of 12 introductory questions with open response options which were clustered into three parts: a) direct inquiry of specific user needs b) indirect evaluation of user needs by discussing the relevance of the use cases identified at Stage 1 as well as any other big data applications they were aware of c) reviewing constraints that need to be addressed in order to foster the implementation of big data applications in healthcare. At Stage 2, we conducted 13 semi-structured interviews with an average length of 75 minutes. At least one expert of each stakeholder group identified in Stage 1, such as patients, clinicians, hospital operators, pharmaceutical industry, research and development (R&D), payors, and medical product providers was interviewed. To derive the user needs from the collected material respectively interview transcripts, we aggregated the most relevant and frequently mentioned use cases into high level application scenarios. Our data collection and analysis strategy was inspired by the triangulation approach [5]. Reviewing and quantitatively assessing the high-level application scenarios (see [6]), we derived a reliable analysis of user needs (see Section 0), and by examining likely constraints of big data applications the relevant requirements that need to be addressed were identified (see Section 2.2).

2. Results

In order to derive user needs and requirements from the collected data, we aggregated the most relevant and frequently mentioned use cases (53 of the 67 use cases that we had discussed with our interviewees) into six high-level application scenarios: 1) Comparative effectiveness research aiming to compare clinical and financial effectiveness of clinical care services 2) Clinical Decision Support assisting

the decision making process of clinicians 3) Clinical operation intelligence aiming to optimize clinical processes 4) Secondary Usage of health data aiming to discover new knowledge by means of data analytics 5) Public health analytics relying on comprehensive disease management of chronic and severe disease and 6) Patient engagement platforms that foster the active engagement of patients in the care process. The six high-level application scenarios established the basis for our user needs and requirements analysis.

2.1. User Needs

User needs in the context of big data can be captured by identifying the most relevant information needs of the various user groups. Within all application scenarios, any information unit that helped the involved users to improve the quality of care without increasing the costs was of great value. For instance, individualized treatment paths cannot be standardized and thus are likely to become very labor and cost-intensive without means of big data-based analytics. Thus, any information that could help to improve the quality AND the efficiency of care at the same time was indicated as most relevant and useful for the user groups. In general, those high impact insights can only be realized if the data analytics is accomplished on heterogeneous data sets encompassing data from the clinical, administrative and financial domain.

2.2. Requirements

Within our study, we identified several requirements that need to be addressed in order to foster the implementation of big data healthcare applications. We distinguished requirements a) that are business-related (BR), b) technical-related (TR) and c) requirement that are business and technical-related (BTR):

High investment needed (BR): The majority of big data applications in the healthcare sector rely on the availability of large-scale, high-quality and longitudinal health care data. The collecting and maintaining of such comprehensive data sets require not only high investments, it usually takes several years until the data sets are comprehensive enough for producing insightful analytics results. In general, such expensive and long-term based investments can rarely be covered by one single party, such that the conjoint engagement of multiple stakeholders, often including the government, is needed.

Value-based system incentives needed (BR): Current system incentives enforce “high number” instead of “high quality” of treatments. Although, it is very obvious that nobody wants to pay for treatments that are ineffective, this is still the case in many medical systems. In order to avoid low-quality reimbursements, the incentives of the medical systems need to be aligned with outcomes and thus foster cooperation between stakeholders.

Business cases with several partners needed (BR): Business cases for big data-bases solutions are difficult to identify. Several partners with diverging points of interests need to cooperate. Often, the one who is benefiting from the solution, is not the one who is in the position to drive the solution or able to pay for the complete solution. For instance, the implementation of data analytics solutions using clinical data requires high investments and resources to collect and store patient data [1]. Although, it seems to be quite obvious how the involved

stakeholder could benefit from the aggregated data sets, it remains unclear whether the stakeholder would be willing to pay or drive such an implementation.

*Data security and privacy (BTR)***:** As of today, legal frameworks defining data access, security and privacy issues and strategies are missing, and the seamless sharing and exchange of data is hindered. Simply because the involved parties lack procedures for sharing and communicating relevant findings, important data and information often remains within one department, group or organization.

Data quality (BTR): Although there exist many big data applications that are mainly looking for patterns in data, and thus do not need clean data, this is not the case for the healthcare domain. In order to derive reliable insights for health-related decision, high data quality standards need to be fulfilled. For instance, the features and parameter list used for describing the patient health status need to be standardized in order to enable the reliable comparison of patient or population data sets.

*Data digitalization (TR)***:** *S*till today, a high percentage of health related data is documented in paper-based form. However, to derive the maximum benefit from health-related data analytics, the data needs to be available in digital format. And, it needs to be complete and in good quality. In order to fulfill this requirement, the demand for standardization as well as technologies supporting the documentation process (e.g. context sensitive information extraction) is pointed out.

Semantic annotation (TR): Health data consist of very heterogeneous data, such as lab reports, medical images, clinical reports, sensor data or gene test results. Only a small percentage of this data is documented in a structured or standardized manner (e.g. International Classification of Diseases[4] (ICD) codes for diagnoses, laboratory data). It is estimated that in the upcoming years, 90% of health data will be provided in of unstructured format (e.g. medical reports) [7]. Semantic annotation facilitates automated content processing of unstructured data. This establishes the data foundation for a holistic analysis of the patient's status or the processing of complex research questions. For semantic annotation it is necessary to rely on standardized and commonly used vocabularies, terminologies or ontologies or coding schemes.

Data sharing (TR): Today a large amount of health-related data is stored in data silos. The efficient and automated data sharing is hardly possible as it faces multiple media disruptions. Although first approaches facilitating data interoperability (e.g. Health Level 7[5], OpenEHR[6], Integrating the Healthcare Enterprise[7]) are available, additional work and research is still required. At present, the health data exchange is mainly based on individualized solutions. This problem could be resolved by standardized clinical data models, commonly agreed terminologies and coding systems. Although, different coding systems are available, they are mainly used in country-specific adaptations (e.g. ICD-10) or lack usability (e.g. SNOMED Clinical Terms[8]).

[4] http://www.who.int/classifications/icd/en/
5 https://www.hl7.org/
6 http://www.openehr.org/
7 http://www.ihe.net/
[8] http://www.ihtsdo.org/snomed-ct/

3. Discussion

Our investigations showed that big health data applications indicate a high potential for improving the efficiency and quality of care delivery. However, we could identify only a limited number of already implemented big data application scenarios. Although non-advanced healthcare analytics applications - such as analytics for improved accounting or quality control - are available in a wide-spread manner [1], those do not yet make use of the potential of big data technologies. To use this potential, it is necessary that the *various dimensions of health data*, such as a) the clinical data describing the patient's history and health status b) the administrative data c) the knowledge about diseases as well as related population data d) the knowledge about changes in time are incorporated in the automated health data analysis. Hence, the highest clinical impact of big data approaches for the healthcare domain can be achieved when aggregating, comparing and relating data from the four dimensions. Therefore, big data technologies could help to produce new insights enabling personalized treatments [6].

This leads to one major problem: Health data cannot be easily accessed. High investment and efforts would be needed. As a consequence, convincing business cases are difficult to identify as the burden of initial required investments strongly reduces profit expectations [6]. In other words, one of the biggest challenges for the realization of big health data applications is the need for high investments, standards and frameworks as well as new supporting technologies to make health data available for subsequent big data analytics applications [8]. Technical and non-technical challenges need to be addressed to foster the seamless access to health data. Within our future work, we plan to elaborate these requirements in further detail.

Acknowledgment

This research has been supported in part by the Big Data Public Private Forum, a project that is co-funded by the European Commission within the 7th Framework Programme under the grant number 318062. The responsibility lies with the authors.

References

[1] Frost and Sullivan. U.S. Hospital Health Data Analytics Market [Internet]. Frost and Sullivan; 2012 [updated 2012 Aug 7; cited 2014 May 6]. Available from: http://tinyurl.com/myb2f3c.
[2] McKinsey and Company. Big data: The next frontier for innovation, competition, and productivity [Internet] McKinsey and Company; 2011 [updated May 2011; cited 2014 May 6]. Available from: http://tinyurl.com/cplxu6p.
[3] Kayyali B, Knott D, Van Kuiken S. The 'big data' revolution in healthcare [Internet]. McKinsey & Company; 2013 [updated 2013 April, cited 2014 May 6]. Available from: http://tinyurl.com/pq4dzrn.
[4] Porter M, Teisberg OE. Redefining Health Care: Creating Value-Based Competition on Results. Boston: Harvard Business Review Press; 2006.
[5] Flick U. Triangulation: Eine Einführung, Wiesbaden: VS Verlag,; 2011.
[6] Zillner S, Rusitschka S, Munne R, Lippell H, Vilela FL, Hussain K, et al. D2.3.1 First Draft of Sector's Requisites. Public Deliverable of the EU-Project BIG; 2013.
[7] Lünendonk Gmbh. Lünendonk Trendpapier: Big Data bei Krankenversicherungen: Bewältigung der Datenmengen in einem veränderten Gesundheitswesen [Internet]. 2013 [cited 2014 May 6]. Available from: http://tinyurl.com/ph6bwmh.
[8] Lobillo F. et al. D2.4.1.First Draft of Sector's Roadmap. Public Deliverable of the EU-Project BIG; 2013.