Towards a Technology Roadmap for Big Data Applications in the Healthcare Domain

Sonja Zillner[1,4], Heiner Oberkampf[1], Claudia Bretschneider[1], Amrapali Zaveri[3], Sabrina Neururer[2,5]
[1]Corporate Technology, Siemens AG, Munich, Germany
[2]Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Austria
[3]Institute of Computer Science, University Leipzig, Germany
[4]School of International Business and Entrepreneurship, Steinbeis University, Berlin, Germany
[5]Semantic Technology Institute Innsbruck, University of Innsbruck, Austria

***Abstract*—Big Data technologies can be used to improve the quality and efficiency of healthcare delivery. The highest impact of Big Data applications is expected when data from various healthcare areas, such as clinical, administrative, financial, or outcome data, can be integrated. However, as of today, the seamless access to the various healthcare data pools is only possible in a very constrained and limited manner. For enabling the seamless access several technical requirements, such as data digitalization, semantic annotation, data sharing, data privacy and security as well as data quality need to be addressed. In this paper, we introduce a detailed analysis of these technical requirements and show how the results of our analysis lead towards a technical roadmap for Big Data in the healthcare domain.**

Keywords—Big Data; technical requirements; data digitalization; semantic annotation; data integration; data privacy and security; data quality

I. Introduction

The healthcare domain faces tremendous productivity challenges. Due to the changing patient demographics as well as the increasing healthcare costs, there is a clear need for cost efficiency, improved quality of care, and broader healthcare services.

Recent studies [1-4] highlight that Big Data technologies and health data analytics are being used to address the efficiency and quality challenges in the healthcare domain. For instance, by aggregating and analyzing health data from disparate sources, such as clinical, financial and administrative data, the outcome of treatments in relation to the resource utilization can be monitored. This aggregation in turn helps to improve the efficiency of care. Moreover, the identification of high-risk patients and predictive models leading towards proactive patient care allows to improve the quality of care.

After performing a comprehensive analysis of domain needs and requirements, we found that the highest impact of Big Data applications in the healthcare domain is achievable when it becomes possible to not only acquire data from one single but various data sources such that different aspects from the various sectors can be combined to gain new insights. Therefore, the availability and integration of all related health data sources, such as clinical data, claims, cost and administrative data, pharmaceutical and R&D data, patient behavior and sentiment data as well as the health data on the web, is of high relevance [5].

However, as of today, the access to health data is only possible in a very constrained manner. In order to enable seamless access to healthcare data, several technical requirements need to be addressed such as: 1) health data is documented in digitalized manner without imposing extra-effort for physicians 2) the content of unstructured health data (such as images or reports) is enhanced by semantic annotation 3) data silos are conquered by means of efficient technologies for semantic data storage and exchange 4) technical means backed by legal frameworks ensure the regulated sharing and exchange of health data, and 5) means for assessing and improving the data quality are available.

The main aim of this paper is to describe and analyze these technical requirements in detail as well as indicate the state-of-the-art and related future research questions. The context of our work is the Big Data Public Private Forum[1] Project, which aims towards developing a technology roadmap for Big Data technologies in the healthcare domain[2]. The detailed analysis of enabling technologies is an important step for developing the technology roadmap, as it helps to identify key requirements for building a Big Data economy within the healthcare sector.

In the following section, we give an overview of our methodological approach before describing the technical requirements in further detail. In Section IV, we analyze the state-of-the-art as well as open research questions for each of the mentioned technical requirements and detail a technology roadmap. We conclude the paper with a discussion of results

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

[2] Beside several other industrial sectors, such as Energy, Transport, Finance, Manufacturing, Retail and the Public Sector

and detailing the next steps for finalizing the technology roadmap development.

II. Methodology

For developing the technology roadmap for Big Data applications in the healthcare domain, we followed a process of several steps:

1. In the first step, we analyzed all the needs and requirements in the healthcare domain that could be addressed by means of healthcare analytics and Big Data technology. For that, we accomplished a *review of available literature, internet sources and market studies* that guided us in developing a comprehensive questionnaire.
2. The questionnaire was used to conduct *12 semi-structured interviews* with representatives of all stakeholder groups of the healthcare domain, such as patients, clinicians, hospital operators, pharmaceutical companies, R&D, payors, and medical product providers. The questionnaire focused on three aspects: a) we asked the interviewees about user needs that could be addressed by means of healthcare IT, b) we asked the interviewees to evaluate a list of precompiled Big Data application scenarios (which we found within our review) as well as to describe other promising Big Data scenarios they are aware of and c) we reviewed with them a list of possible constraints that are hindering the successful implementation of Big Data scenarios in the healthcare domain.
3. By *clustering and ranking the discussed use case scenarios*, we derived a set of six high-level application scenarios. By analyzing these application scenarios, we identified relevant constraints and requirements that need to be in place for the successful implementation of the scenarios. By aligning this initial list of constraints and requirements with the input from our interviews, a final list of constraints/requirements was compiled.
4. In our further analysis, we distinguished between technical and business-related requirements[3]. The technical requirements (which we also call *enabling technologies*) were analyzed in further detail by describing the AS-IS and TO-BE[4] situation, by investigating the required functionalities, the available technologies as well as by identifying open R&D questions. When needed, we conducted further expert interviews. The summary of this analysis is described in the following two sections[5] III and IV.
5. Besides our analysis of enabling technologies, the investigation of *future opportunities* associated with Big Data applications in the healthcare domain is needed. This is done by analyzing the technology required for implementing the selected high-level scenarios.
6. The final step within the roadmap development is the temporal alignment and ranking of technical requirements described in Steps 4 and 5. As the adoption of new technologies depends on the degree to which the identified business requirements can be addressed, we determine how the business requirements can be influenced and by whom they need to be addressed.

Step 5 and 6 is part of our future work.

[3] Three business-related requirements were identified: a) lack of promising business cases, b) the need for high investments and c) need for value-based incentive system.

[4] TO-BE Situation refers to the year 2020

[5] The complete analysis can be accessed under http://www.big-project.eu/

III. Enabling Technologies

In order to establish the basis for wide-spread usage of Big Data applications in the healthcare domain, several technical requirements need to be addressed. We labeled these technologies as “enabling”, because they establish the technical foundation for subsequent Big Data applications. In this way, *enabling technologies* cover all health-specific data management technologies that ensure that the various heterogeneous health data pools can be easily accessed and that the health data is integrated and available. Within our analysis of technical requirements, we distinguish enabling technologies from *value creating technologies* that are needed to elaborate concrete Big Data-based business opportunities. The differentiation between enabling and value-creating technologies is needed for our further analysis in order to indicate how and to which extent the various technologies depend on each other: Enabling technologies often require long-term investments from various partners without immediate business potential, while value-creating technologies relate to concrete business opportunities that assume that the various data sources are available (by means of enabling technologies).

A. Data Digitalization

Some years ago, data digitalization was a huge problem, but today it is progressively becoming a less important problem from a technical point of view. Nevertheless digital data is still not available everywhere. The extent to which healthcare Information Technology (IT) systems are in use differs across and also within countries. A study showed that on average 55% of healthcare providers in Germany use healthcare IT within primary care settings and 60% in secondary care settings [9]. For the United Kingdom, the numbers are different: 63% of providers rely on healthcare IT in primary and only 15% in secondary healthcare [9].

In a perfect scenario, data collected for primary care settings should be available in an annotated, curated and high quality manner for secondary use.

B. Semantic Annotation

When working with health-related data in general and Big health-related Data in particular, one is facing the challenge of data heterogeneity (reports, lab reports, images, sensor data, etc.). In addition, large amounts of information are captured in unstructured formats (e.g. reports or images). The International Data Corporation (IDC) market research institute estimates that in the upcoming years 90% of health data will be provided in unstructured formats [6]. Semantic annotation is described as a possible solution for processing heterogeneous unstructured data seamlessly.

Besides a small set of standardized metadata, such as the DICOM header, that provides the basis for the exchange and management of documents, the content of unstructured information is in general not provided in standardized formats. For example, reporting in radiology is still conducted as free text. The reading and interpretation of such data is accomplished manually by individual clinicians. Without semantic annotations, it is not possible to process the content of unstructured data automatically. Therefore, a holistic analysis of the patient's status is hindered.

The envisioned impact of semantic annotation is very promising. In order to automatically align related data sets, their content needs to be represented explicitly and consistently. This means, semantic annotation relies on commonly used vocabularies or ontologies.

C. Data Sharing

As of today, a lot of health data is stored in data silos. A seamless exchange and aggregation of the data often relies on individualized solutions due to the lack of standards and flexible interfaces as well as the heterogeneous nature of the data. In comparison to the degree of healthcare IT adoption, the incorporation of seamless information exchange is far less advanced [9]. On average, for instance, in Germany less than 23%, in the UK less than 46% and in the US less than 36% of the healthcare providers use healthcare information exchange technology [9].

In terms of analyzing the current state of health data sharing, several deficiencies have to be emphasized: First, the data exchange within one healthcare provider is complicated due to the usage of different information systems in different departments. Although it is feasible from a technical point of view to exchange health data by e.g. using HL7[6] CDA (Health Level 7 – Clinical Document Architecture), as of today health data is hardly shared across organizations due to non-technical reasons. Moreover, the healthcare domain lacks internationally accepted coding systems. Even the ICD[7] (International Classification of Diseases), which is broadly accepted, is used in country-specific adaptations only. Other promising systems still lack acceptance (e.g. SNOMED Clinical Terms[8]). In terms of underlying means for data representation, existing Electronic Health Record (EHR) systems mainly provide a case-centric instead of a patient-centric view, which hinders longitudinal health data integration.

To enable seamless health data exchange, standardized data models for clinical data as well as coding schemes for labeling content need to be agreed upon.

D. Data Privacy and Security

As health data is private data, its processing needs to follow high data security and privacy constraints. Therefore, there is a strong need for technical infrastructures, legal processes for data sharing and communication as well as for the implementation of suitable organizational processes that enable secure and transparent health data sharing.

Since the importance of data and especially of Big Data is increasing and becoming a competitive company value, data security, risk management and data privacy requirements are becoming more and more important. For ensuring transparent and secure data sharing, challenges on four different levels, namely the legal, technical, organizational and social level, have to be taken into account: 1) In the European Union the huge importance of (health) data privacy and security is underlined by Directive 95/46/EC [7]. Based on this minimal *legal* standard, all member states were called to implement a national data protection law. This is one of the major issues concerning data privacy and security within the European Union. There does not exist one common legal framework for all member states. Each member state issued its own data protection law. 2) From a *technical point of view*, the implementation of secure Big Data applications with respect to data privacy is supposed to be possible with the available technologies (e.g. Integrating the Healthcare Enterprise (IHE) initiative[9]). Other privacy enhancing techniques like anonymization or pseudonymization of personal health-related data are shown and described in [8]. 3) From an *organizational point of view*, the storage, processing, access, and protection of Big health Data has to be regulated on different levels (e.g. institutional, regional, national, international). 4) From a *social point of view*, the huge benefits of Big health Data need to be communicated and promoted.

When talking about anonymization in Big health-related Data context, one always has to keep in mind the possibility of unintentional re-identification of individuals, when aggregating anonymized Big health Data from different sources. This problem results from the additional knowledge gained by the merging of different data sources.

In order to resolve the mentioned issues, a common legal framework within the European Union regulating the intramural and extramural exchange and processing of health-related (Big) Data is needed.

E. Data Quality

The data quality in Big health Data application depends on the quality of data of the original data sources. Usually medical product providers are not responsible for the quality (e.g completeness, accuracy) of the data collected and documented in hospitals. Nevertheless, they provide tools for data collection, documentation and analysis. These tools can help to support hospitals and healthcare providers to ensure that data is complete. Additionally, the tools can help to improve data quality (e.g. plausibility checks, mandatory items). A major issue regarding data quality, which directly links it to data digitalization, are media disruptions. The more media disruptions exist in a process chain, the worse the data quality gets.

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

Poor data quality is a major limitation to the benefit of (Big) Data analysis and the quality of healthcare itself. Therefore, before analyzing data two assessment steps have to be carried out: a) the quality of the data sets used for Big Data applications has to be evaluated and b) quality has to be improved by using appropriate approaches (e.g. multiple data imputation).

IV. Technical Roadmap Description

In this section, we will analyze the required technologies for addressing the above mentioned challenges. However, one needs to keep in mind that the availability of technology is not sufficient to solve the mentioned challenges, but dedicated processes, standards or frameworks need to be available and implemented. As these non-technical aspects might influence the temporal axis of our technical roadmap, we decided to integrate those aspects within our analysis, but to indicate them explicitly. Figure 1 provides a consolidated view of the main enabling and value-creating technologies, which are described in the following sections. These technologies may be already available or still open R&D topics. Enabling technologies are referred to as technologies, which establish the technical foundation for subsequent Big Data applications, while value creating technologies create a concrete business value. For a better understanding, we highlighted (italic) the catchwords also appearing in Figure 1 throughout the text and tagged them (e.g. *A1*) in order to find them quicker in the figure.

A. Data Digitalization

Digitization of medical data has a huge impact on every aspect of the healthcare domain: healthcare delivery, health management, healthcare policy making, and administration. Digital health-related data supports integrated healthcare delivery, highly automated data processing, research, as well as medical routine. Various technologies for the digital capturing of health data, such as Hospital Information System (HIS), Radiology Information System (RIS), Picture Archiving and Communication System (PACS), etc. are already in place.

In order to improve the health data documentation process by reducing the extra effort for physicians, several technologies in the domain of m*edical speech recognition (A1)* systems provide means for implementing non-paper based documentation.

B. Semantic Annotation

To enable semantic annotation for medical data, some major requirements have to be fulfilled. Annotation requirements should be based on medical context understanding. Therefore, a fully covered medical terminology with an attached semantic classification is needed as well as *medical information extraction (B1)* techniques for the medical domain, which allow to integrate external domain knowledge. Also an annotation process, which makes the semantic of the health data explicit, is required.

As of today, there are some technologies available, which meet the requirements mentioned above. Nevertheless they are not yet widely used and/or accepted on a broader basis. *Semantic medical terminologies (B2)* (e.g. Unified Medical Language System (UMLS), SNOMED Clinical Terms (SNOMED CT)), which contain a lot of medical knowledge and information have several shortcomings (e.g. lack of multilingual concepts [10]). *Semi-structured reporting standards (B3)* (e.g. RSNA reporting standards for Radiology) are not commonly used.

There are still some (partially) open Research and Development (R&D) questions concerning semantic annotation. First *standardized medical annotation frameworks (B4)* and *enriched semantic medical terminologies (B5)* are required. Available terminologies have problems and need to be extended regarding, quality and/or quantity and semantic description of concepts. Also *semi-structured reporting algorithms (B6)* and *image-understanding algorithms (B7)* should be targeted by research.

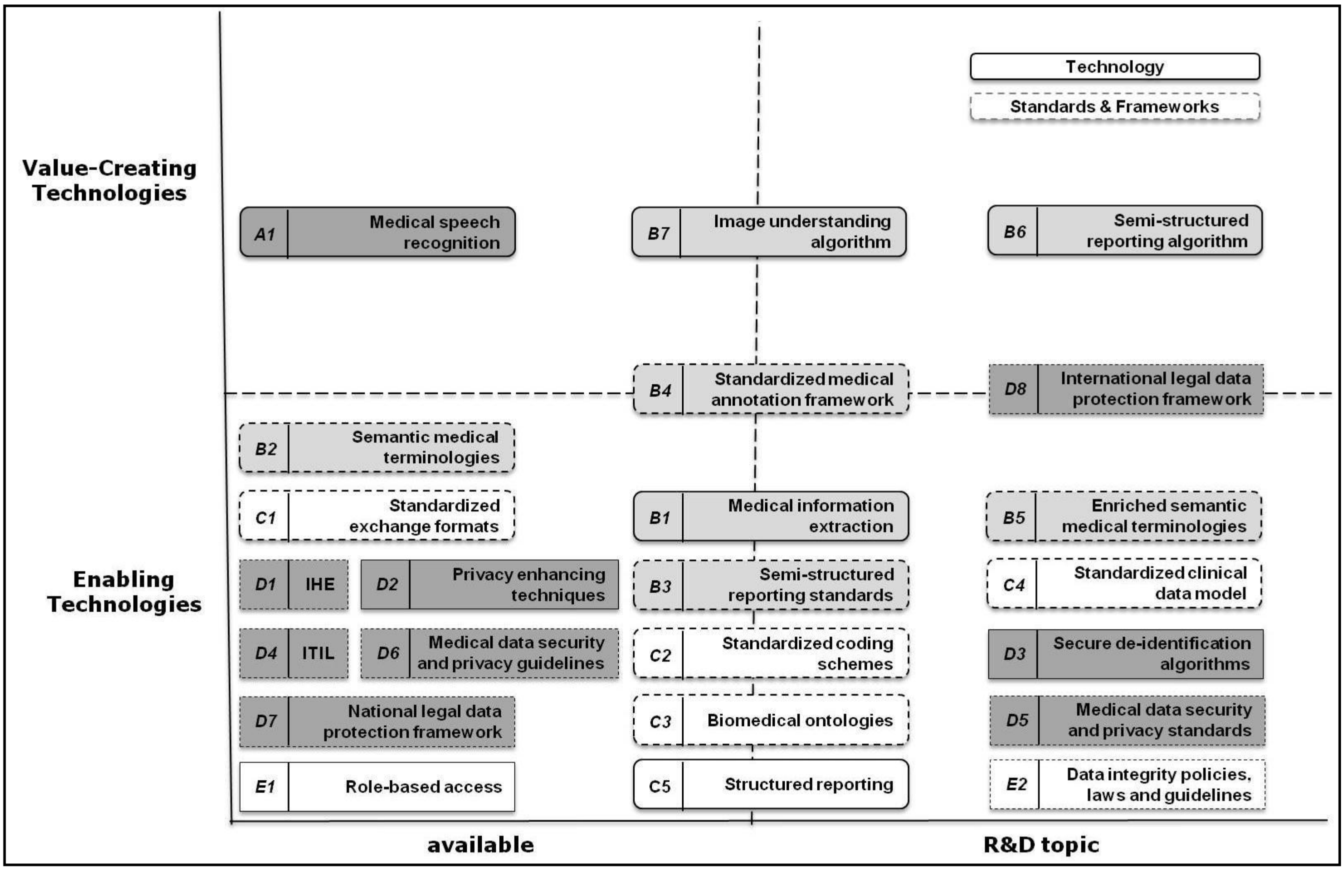

Figure 1 Enabling and Value creating Technologies for Big Data applications in the healthcare domain (different colour-codings / shapes indicate the five technology areas)

C. Data Sharing

To enable semantic data integration and data sharing, common standards for representing data are strongly required. Therefore, standardized data storage and exchange formats as well as a structured and standardized representation of health data are needed. Commonly accepted and used coding systems, a complete, patient-centric and longitudinal patient data representation as well as proper semantic annotation algorithms are required to facilitate and support seamless data sharing and exchange.

There are already technologies available, which meet the requirements mentioned above: Health Level 7 (HL7) provides commonly accepted *standardized exchange formats (C1)* (HL7v2.x and v3.0) as well as document standards (HL7 Clinical Document Architecture (CDA)). To provide a common meaning and understanding of medical data, *standardized coding schemes (C2)* and terminologies are required. They allow a common representation of structured data (e.g. ICD or LOINC[10]). Coding systems for more granular health data like SNOMED CT are available even though with several open issues regarding consistency, performance and unambiguous representation [11-15]. *Biomedical ontologies (C3)* (e.g. Gene-Ontology[11]) as well as *standardized clinical data models (C4)* (e.g. HL7 Reference Information Model (RIM)[12] are already available.

A majority of the technologies mentioned above have several issues, which need to be targeted by further research: Most of the existing ontologies are available in English but lack in multilingual representations and, therefore, in usability. Regarding standardized data models we note that HL7 RIM has certain issues concerning the ontological consistency [16]. Research activities for the development of an integrated patient data model on the basis of well-defined ontologies are on-going (e.g. Model for Clinical Information [17]). There are numerous other initiatives developing data models for improved data sharing in the context of clinical studies. Prominent examples are the Translational Medicine Ontology [17] or models developed by the Clinical Data Interchange Standards Consortium[13] or within the EHR4CR project[14]. Another important target for research are tools facilitating *structured reporting (C5)* which do not put extra work for clinicians and can be seamlessly integrated in the clinical workflow.

D. Data Security and Privacy

The Cloud Security Alliance [18] listed major Big Data security and privacy challenges, which include secure computations in distributed programming frameworks, secure data storage and transaction logs, real-time security and compliance monitoring, scalable and composable privacy-preserving data managing and analysis approaches and granular access control and audits.

[10] http://loinc.org/
[11] http://www.geneontology.org/
[12] http://www.hl7.org
[13] http://www.cdisc.org/
[14] http://www.ehr4cr.eu/

As of today, some technologies enhancing data privacy and security are already available. The Integrating the Healthcare Enterprise (*IHE (D1)*)[15] initiative enables plug-and-play and secure access to health information whenever needed. IHE also promotes the use of well-established and internationally accepted standards, such as DICOM or HL7. [19] lists some *privacy enhancing technologies (D2)* (e.g. Peterson Approach, Pommerening Approaches, Electronic Health Card). Some of these approaches use pseudonymization techniques, which is important especially for longitudinal medical research. Here it is sometimes necessary to re-identify study subjects in order to, for example, communicate important study results. As mentioned earlier, unintentional re-identification of individuals becomes a real problem, when integrating linked data from various sources. Therefore, the real dimension of anonymity has to be evaluated and *secure de-identification algorithms (D3)*, such as the k-anonymity approach have to be used or developed. To support IT service management, *ITIL*[16] *(D4)* is a widely accepted approach that provides a cohesive set of best practices including guidelines for data security and privacy.

For enhancing data privacy and security as well as for the other requirements listed in this article, internationally accepted *medical data security and privacy standards (D5)* are needed. Therefore existing *medical data security and privacy guidelines (D6)* can serve as a point of reference. As there will always be efforts for cracking the existing algorithms in order to collect personal health data, secure data de-identification approaches need to be continuously evolved. One major requirement for enhancing data privacy and security is a rigid and proper *national legal data protection framework (D7)* as well as *international legal data protection framework (D8)*.

E. Data Quality

According to [20], in order to improve data quality, the following requirements need to be met: a) a controlled, *role-based access (E1)* can hinder unauthorized reading and writing actions b) data dictionaries have to be used to achieve a common terminology c) standardized formats ensure consistency and follow directly from the last point mentioned d) *data integrity policies, laws and guidelines (E2)* are needed and have to be followed.

V. Discussion

In order to identify the timeline for the technology roadmap, we need to answer the following three questions: a) Is the needed *technology* available? b) Are *standards and legal frameworks* that enable the reliable exchange of data in place? and c) Are the relevant *organizations* in the healthcare domain, such as hospitals, pharmaceutical companies, payers, etc., ready to *implement* the dedicated processes, standards and frameworks to foster the adoption of the Big Data technology.

Within our analysis, we found that the *availability of technologies* is not the major issue. For sure, extensions, adaptations or new functionalities need to be developed, however, the majority of technological challenges can already

[15] http://www.ihe.net/
[16] http://www.itil-officialsite.com/

be described in a well-defined manner and first research results do already exist.

The *availability and usage of standards* is an issue that applies to all discussed technological areas: Either standards are not used at all or local adaptations of standards (e.g. country-specific versions of the ICD) are favored. Considering the fact that the healthcare market in general is very much affected by local constraints and processes, this insight is not surprising. However, it is important to keep in mind that local imprints of healthcare markets limit the economic scale of new technological developments due to the rather small (local) market sizes. This again reduces the need for global standards and their implementation, which again hinders the development and implementation of standards and frameworks. In a nutshell, the availability of standards and frameworks fostering the seamless health data sharing are needed for Big Data applications, however their development, adoption and implementation triggers a classical hen and egg problem.

As the *implementation of standards* is often a high cost factor with unclear ROI[17] values, organizations usually only decide to implement standards if it is obligatory, it is incentivized or if a critical number of market players have already adopted the standard, which again puts pressure on the remaining ones. Thus, regulation and legislation has a strong influence on fostering the adoption of standards in healthcare, and thus, the implementation of the described promising opportunities of Big Data applications in the healthcare domain.

VI. Conclusion And Next Steps

Our study has shown that the impact of Big Data applications is very promising, but relies on the availability of healthcare data. Thus, several technical challenges, such as data digitalization, semantic annotation, data sharing, data quality and data privacy and security, need to be investigated. Within our study, we identified the state-of-the-art as well as future research topics within five technical domains. However, it is important to keep in mind that the availability of healthcare data is not only a technical challenge, but relies to a large extent on the availability and implementation of standards and frameworks specifying the seamless and reliable data exchange.

In future work, we will finalize our work on the development of a technical roadmap for Big Data technologies. Therefore a detailed analysis of identified business requirements, needed value-creating technologies, and identified use case scenarios will be carried out. Afterwards the comprehensive findings will be aggregated into a compact roadmap representation.

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 for this publication lies with the authors.

[17] Return on Investment

References

[1] Frost and Sullivan, U.S. Hospital Health Data Analytics Market, 2012.

[2] McKinsey and Company. Big data: The next frontier for innovation, competition, and productivity. 2011.

[3] P. Groves, B. Kayyali, D. Knott and S. Van Kuiken. The ‘big data’ revolution in healthcare. McKinsey & Company. 2013

[4] M. Porter and E.Olmstead Teisberg. Redefining Health Care: Creating Value-Based Competition on Results. Boston: Harvard Business Review Press, 2006.

[5] S. Zillner et al. *D2.3.1 First Draft of Sector’s Requisites*. Public Deliverable of the EU-Project BIG (318062; ICT-2011.4.4), 2013.

[6] Lünendonk Gmbh. Trendpapier 2013: Big Data bei Krankenversicherungen. Bewältigung der Datenmengen in einem veränderten Gesundheitswesen. online available, 2013.

[7] European Parliament and the Council of the European Union. Directive 95/46/EC of the European Parliament and of the Council 1995, [http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML], 1995.

[8] International Organization for Standardization ISO/TS 25237:2008 Health informatics - Pseudonymization. 1.edition, Geneva, 2008

[9] Accenture. Connected Health: The Drive to Integrated Healthcare Delivery. Online: www.acccenture.com/connectedhealthstudy, 2012.

[10] J. Ingenerf. Die Referenzterminologie SNOMED CT : von theoretischen Betrachtungen bis zur praktischen Implementierung. Neu-Isenburg, Germany, 2007.

[11] A. Rector, S. Brandt and T. Schneider. Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications. Journal of the American Medical Informatics Association : JAMIA, 18(4), 432–40, 2011.

[12] R. Cornet. Definitions and qualifiers in SNOMED CT. Methods of information in medicine, 48(2), 178–83. doi:me10.3414/ME9215, 2009.

[13] S. Schulz, R. Cornet, and K Spackman. Consolidating SNOMED CT’s ontological commitment. Applied Ontology, 6, 1–11. doi:10.3233/AO-2011-0084, 2011.

[14] C. Martínez-Costa, and S. Schulz. Ontology-based reinterpretation of the SNOMED CT context model (pp. 1–6). In Proceedings of the International Conference on Biomedical Ontology, 2013.

[15] D. Markwell, L. Sato, and E. Cheetham. Representing clinical information using SNOMED Clinical Terms with different structural information models. In K. Spackman & R. Cornet (Eds.), Proceedings of the 3rd international conference on Knowledge Representation in Medicine (KR-MED 2008) (pp. 72–79). Retrieved from http://ceur-ws.org/Vol-410/Paper13.pdf, 2008.

[16] B. Smith, L. Vizenor, and W. Ceusters. Human Action in the Healthcare Domain : A Critical Analysis of HL7 ’ s Reference Information Model. Retrieved from http://ontology.buffalo.edu/12/HL7-and-BFO.pdf, 2012.

[17] J.S. Luciano et al. (2011). The Translational Medicine Ontology and Knowledge Base: driving personalized medicine by bridging the gap between bench and bedside. *Journal of Biomedical Semantics*, *2 Suppl 2*(Suppl 2), S1. doi:10.1186/2041-1480-2-S2-S1

[18] H. Oberkampf, S. Zillner, B. Bauer and M. Hammon. An OGMS-based Model for Clinical Information (MCI). In Proceedings of International Conference on Biomedical Ontology, Montreal, Canada, 2013.

[19] Cloud Security Alliance. Top Ten Big data Security and Privacy Challenges. Online: http://www.isaca.org/Groups/Professional-English/big-data/GroupDocuments/Big_Data_Top_Ten_v1.pdf, 2011..

[20] Neubauer T, Kolb M. "An Evaluation of Technologies for the Pseudonymization of Medical Data". In R. Lee, G. Hu, H. Miao (Eds.): Computer and Information Science, SCI 208, Springer, 2009.

[21] AHIMA. "Assessing and Improving EHR Data Quality (Updated)." Journal of AHIMA 84, no.2 (March 2013): 48-53 [expanded online version], 201