Alexander Rossmann, Alfred Zimmermann (eds.): Digital Enterprise Computing 2017,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2017 107

Data Architecture for Digital Health Insurances

Jutta Degele[1], Julia Hain[2], Valeria Kinitzki[3], Sascha Krauß[4], Peter Kühfuß[5] and Natascha Sigle[6]

Abstract: The increasing number of connected mobile devices such as fitness trackers and smartphones define new data sources for health insurances, enabling them to gain deeper insights into the health of their customers. These additional data sources plus the trend towards an interconnected health community, including doctors, hospitals and insurers, lead to challenges regarding data filtering, organization and dissemination. First, we analyze what kind of information is relevant for a digital health insurance. Second, functional and non-functional requirements for storing and managing health data in an interconnected environment are defined. Third, we propose a data architecture for a digitized health insurance, consisting of a data model and an application architecture.

Keywords: Data Architecture, Data Model, Digital Health Insurance, ArchiMate 3.0, Big Data and Analytics, Requirements Engineering

1 Introduction

As a result of advances in sensor system technologies new possibilities in health analytics arise, changing the traditional business models of health care providers.

The base for a digitization of health insurances is the more conscious and active lifestyle of an increasing percentage of the population. By using wearables much data is generated, though not evaluated efficiently enough yet. Enhanced data analysis enables health insurance companies to move away from the current reactive model of care towards a preventive one. Through proactive offers, such as fitness courses, support for therapies or rehabilitation, or simply reminders for medication or conscious food intake, the customer experiences direct added value. The customer also contributes to increasing the knowledge about disease risk and the effectiveness of interventions.

In this paper, we will describe a scenario which is divided in four use cases. The scenario is a simplified showcase for the possibilities emerging from the use of fitness trackers and sets the foundation for an exemplary data architecture of a digital health insurance.

[1] Reutlingen University, Herman Hollerith Zentrum, jutta.degele@student.reutlingen-university.de
[2] Reutlingen University, Herman Hollerith Zentrum, julia.hain@student.reutlingen-university.de
[3] Reutlingen University, Herman Hollerith Zentrum, valeria.kinitzki@student.reutlingen-university.de
[4] Reutlingen University, Herman Hollerith Zentrum, sascha.krauss@student.reutlingen-university.de
[5] Reutlingen University, Herman Hollerith Zentrum, peter.kuehfuss@student.reutlingen-university.de
[6] Reutlingen University, Herman Hollerith Zentrum, natascha_sarah.sigle@student.reutlingen-university.de

Participating roles are the insurance company, the customer himself and a doctor.

Main interactions between the participants are:

1. The individual calculation of insurance fees using a customer's current fitness status.

2. The customer can access an aggregated fitness data overview on the insurance portal.

3. The customer is heavily burdened with heart diseases and agrees to a pulse monitoring by the health insurance. In case of irregularities the responsible doctor will be informed.

4. By comparing data sets of the insured with the analytic database of the health insurance, undetected diseases may be discovered.

Sources analyzed in a preliminary literature review focused either solely on the business context of health insurances, mostly without including health personnel, or on collecting data from fitness trackers and other smart devices without consequential business models. The intended data architecture is supposed to fill the gap and include both business and technical prospects. Therefore, the specific research question is:

RQ How can a data architecture for a health insurance with a digital business model look like?

To approach this question, two further sub-questions were defined:

RQ1 What data is relevant for the digital health insurance and what insights about customers can be gained from this data?

RQ2 What are functional and non-functional requirements for health insurance data and its organization?

The paper is structured accordingly in the following chapters. Chapter 2 starts off with the methodology used. Then, chapter 3 covers the results of the sub-questions. The relevant data for the chosen scenario and the insights which can be gained from that data are presented in chapter 3.1. The following subchapter then covers the requirements that data yields towards both a data and application architecture before chapter 3.3 continues to introduce different data architecture modeling approaches and rate their suitability for the given use cases and requirements. In chapter 4, the chosen architecture approaches are modeled based on the results of the sub-questions. A discussion with input of a health professional and a big data specialist follows in chapter 5. Finally, the outcomes of this paper are summed up and an outlook is given in chapter 6.

2 Research Approach

In order to draw conclusions from the given use cases in the context of digital health insurances, we have performed extensive literature research. We started with several

research questions ranging from existing digital health insurance approaches via data and application requirements for such a one to the possibilities of Service Oriented Architectures in this context. In an initial keyword search in five electronic databases (IEEE, ACM, Springer, EDDI and Google Scholar), over 100 relevant documents were retrieved, including scientific journal papers, industry proposed data models for health care and insurance and documents describing the application structure of Big Data software. This basis allows further limitations to the scope of the initial research topics, leading to a focus on the data architecture of a digital health insurance. Designing the architecture will be performed in two steps. Firstly, the most important findings on data relevant to a digital health insurance and the ensuing functional and non-functional requirements identified from literature are aggregated into a suitable data model. Secondly, a corresponding ArchiMate 3.0 application architecture will be proposed.

3 Preliminary results

3.1 Relevant Data and Insights

Approaching the research question of relevant data and its insights, the first interesting point is: Assuming some kind of relevant data can be gathered, what could be done with this data?

The main aspect of upcoming insurance business models seems to be individualized policies based on the personal needs of the insured persons. In case of health insurances, this can be implemented by offering a certain policy model with a cheaper insurance contribution if the customer is proven to pursue a healthy lifestyle. Another policy model could be offered for people with chronic diseases, e.g. including a more intensive customer support. But which data is useful to categorize the customers in such way?

For this endeavor, some physical information is important like: age, gender, blood pressure, pulse, body weight, body height, (previous) diseases and disabilities. But most of these data artifacts are only helpful in conjunction with other information. The body weight can give relevant clues, but only by knowing the body height as well. Age and gender could also provide appropriate statements. Furthermore, not only physical information is important for such a classification. Additional information like possible hereditary diseases, eating habits, sports activities, alcohol and drug consumption, workplace or being a smoker or not can help the health insurance acquire a better picture of the customer. To enable a customer classification, certain criteria have to be derived from the available data, in order to match a customer to the most adequate policy model.

Fitness trackers are useful devices for collecting some of the mentioned data. A comparison of several models from different manufacturers showed that most of them can track the following metrics: number of steps, distance covered, calories burned, pulse and heart rate zone [Fi16]. This data could not only be helpful for categorizing the

customers into classes, the health insurances also get relevant information about the everyday life of the customers. To motivate the customers to share their fitness data, a health insurance could not only give discounts on the insurance premium [Mü16] but also offer a platform on which the customer can get a central overview of all his fitness activities. Furthermore, the insights about the customer given by his fitness data enables the insurance to suggest individual nutrition plans or fitness programs. In case of life-threatening conditions like heart diseases, the permanent usage of fitness trackers allows monitoring for irregularities. Moreover, the amount of collected data from different insured allows pattern-matching leading to improved diagnoses.

3.2 Requirements for a Data Architecture

The foundation of constructing a data architecture is to know which requirements need to be fulfilled. Therefore, requirements engineering is used, which is a systematic and disciplined approach for the specification and management of requirements. Requirements can be functional or non-functional [Ru14]. In the case of health insurance data, the functional requirements are defined by:

Structure of Health Insurance Data: Factors to be considered are the number and kind of stakeholders, record structure and storage location. Regarding the chosen use cases, only the insured and their respective doctors are relevant stakeholders. An enlarged use case would also include pharma industry, research, government, gyms, further healthcare providers, as well as insurance shareholders. Data records can be organized by source, in chronological order, by a predefined protocol structure or by assigning all records to certain problems [Sa12]. For fitness data which has a continuous inflow and may come from different sensor sources a chronological storage structure fits best. The data can be stored decentralized, centralized or in a hybrid structure [Sa12] [Br17]. In the evaluated use case the insurance acts as a central data host for all stakeholders.

Use Cases of Stakeholders: Each use case introduced by the stakeholders or the insurance company consists out of a process of sourcing, collecting, storing, processing and serving data [Gi17]. The individual process defines which data types are needed. Data sources of the observed use case are smart devices such as fitness trackers and smartphones providing different kinds of fitness data.

Data Governance and Data Management: The data needs to be governed and managed [DM17] to ensure quality, uniqueness and security. Especially the question of data control needs to be defined. Considering use cases in scope, aggregated fitness data can be created, updated and deleted by the insurance, while the other stakeholders have read permission only.

Health and Insurance Standards: To enable data exchange within the health community a common terminology, message exchange and object model are essential [Sa12]. Additionally, the different device manufacturers need to be considered.

Further important considerations for non-functional requirements are:

Availability: Critical health and insurance customer data must be at hand 24/7. Normally, fitness data is not critical. However, the case of patient monitoring requires high availability.

Performance: Real time data from health devices and certain analytics need to be processed in a fast pace, but data like archive documents and diagnostic reports do not require such a fast procedure. The architecture has a clear need for a type of two speed architecture.

Extensibility: To be future oriented, flexible and competitive the architecture must be open for new stakeholders, data types, contents, interfaces and applications.

3.3 Classification of Data Model and Architecture Approaches

The challenge within finding the right data architecture to model all the different data types for a digital health insurance is to get an abstract view. Therefore, we shortly introduce different data models and architectures and assess their suitability regarding the identified requirements from the previous chapter.

With a “data warehouse model” is it possible to model the process from gathering and transforming the data to storing it. The focus of this approach is not the data, but rather the transforming and loading of data into a data warehouse. With lots of different sources and data types, it gets inefficient and inflexible with new data sets [GIW97].

The “data structure diagram” is a predecessor of the entity-relationship model (ER model) which will be described subsequently. While the data structure diagram sets its focus on the relationship of the elements within an entity, the ER model is more about the relationship between entities. As in the contemplated use cases data is linked to a large extent, the relationships between entities are considered to be much more important.

A “data lifecycle model” provides a high-level overview over the stages in which the data is involved. It complements a data model to express an exact data process, but is no data model itself and will therefore not be used for the intended data architecture.

The “ER model” describes an abstract data structure which consists of entities, attributes and relationships. The entities represent objects in the real world. Entities are connected via relationships which represent the relations and dependencies between them. Each entity or relationship can have a number of attributes assigned, denoting their relevant characteristics [ES07]. To model all the different data types, data sources and their context, the ER model is the best option in this case.

In order to describe a holistic data architecture including not only the data itself, but also the data gathering, processing and storing, another modeling approach is necessary. For

this purpose ArchiMate 3.0, a modeling language for Enterprise Architectures specified by the Open Group, is used. ArchiMate is distinguished into a business, application, technology and a strategy layer [OG12a]. In this paper, the application layer, which models information system architectures, is used. It shows a set of elements, ranging from data objects via application services, processes and events to application interfaces and components, as well as the relationships between those elements [OG12b]. Hence, we can model a complete view of the data, the relationships and the process from gathering data to handling user requests.

4 Data Model and Architecture

In this chapter, the identified relevant data for each considered use case plus the requirements are aggregated into two models, functioning as data architecture.

The ER model in Fig. 1 shows the detailed composition of the required data artifacts and participants.

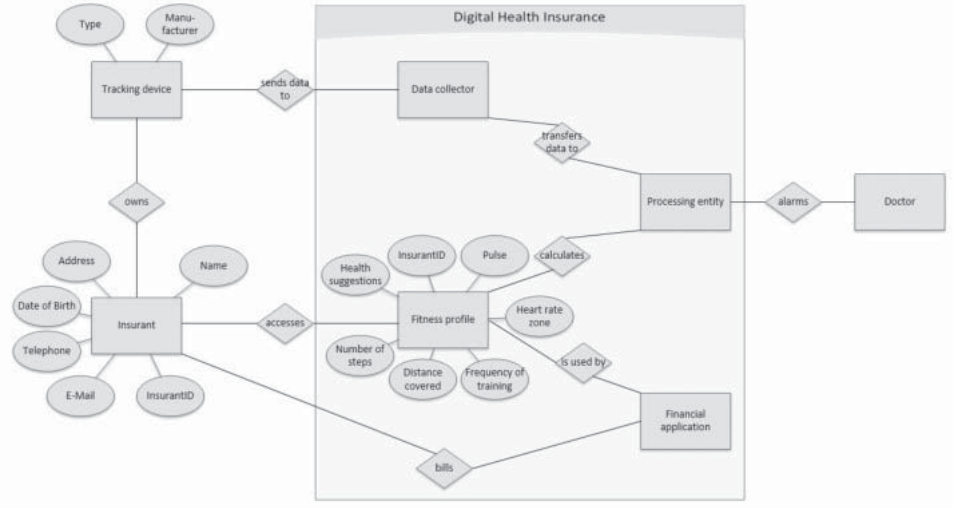

Fig. 4: ER model for the identified use cases

For each insured, personal data such as name, address, telephone and a unique InsurantID is stored. The insured may own a tracking device, e.g. a smartphone or a fitness tracker which acts as data source. The tracked data is sent to the central data collector provided by the insurance, which then transforms the data into a standardized form and transfers it to the processing entity. Here the analysis takes place, including the monitoring for irregularities as required in use case 3. If critical irregularities are detected, a doctor will be informed to contact the insured. After processing, the data is stored in the form of a fitness profile. This profile includes the correspondent InsurantID,

the tracked raw data and an analysis on the frequency of the training to determine the insured person's fitness status as well as individualized health suggestions. These suggestions may be recommendations of sport courses and other health-improving activities based on the current preferences of the insured, but also warnings or reminders if a medical check-up appointment seems to be necessary. The insured can access his fitness profile via a web platform. In the back end, the fitness profiles of the insured are also used by the financial application to calculate discounts on the insurance premiums based on the respective fitness status.

As data collection and processing are important elements for dealing with the described use cases, the application model in Fig. 2 shows them in a more detailed manner.

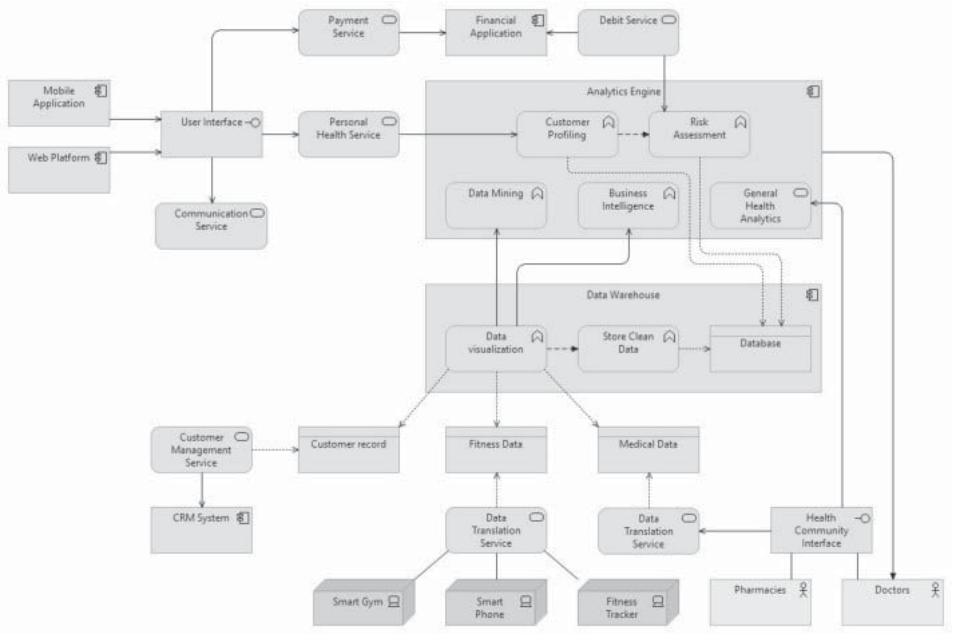

Fig. 5: Model of the ArchiMate 3.0 Application Layer

The central component of the architecture is data warehouse, but instead of falling back to the traditional ETL (Extract, Transform, Load) process, a more flexible approach is chosen. As ETL does not deal well with many data types from different sources [Ki16], it cannot fulfill the requirement of *Extensibility*. A solution is the usage of data virtualization as an abstracted layer across conventional customer records, specific medical data and fitness data. The latter is aggregated by a data translation service, which does not only collect the data from various devices, but also translates it according to a common standard. On top of the data warehouse an analytics engine is placed which combines different techniques from business intelligence and data mining. An exploratory data analysis evaluates (hidden) relationships among different variables

[Se09]. Thereby, undetected diseases may be revealed and prognoses on the world health may be derived. A detailed customer profile is generated which enables the personal health service to pass personal fitness statistics and health suggestions to the user interface. Based on the customer profile a risk assessment of the insured is conducted which determines the individual insurance premium. The requirement of *data control* is taken into account by exposing only small partitions of the data via services, so that no external party has direct access to it.

5 Discussion

In first validations of the proposed architecture with a health professional and a big data specialist, the architecture was perceived as reasonable for the chosen scope. However, several aspects were revealed that have not yet been taken into account. Both interviewees brought up ethical concerns, as an improved information flow about the health status of the insured may lead to disadvantages for chronical ill patients. Furthermore, fitness trackers are still quite easy to manipulate, so fraud may become a bigger problem. For the designed architecture, only a small number of data sources was regarded, but many more may be interesting for analyses. Insurances may be confronted with huge amounts of unstructured data, e.g. contracts of the insured with sports clubs or gyms, records of respiration, sleep screenings or data from medical devices like cardiac event recorders. Information security and privacy has to be ensured for all channels. The current architecture has to be augmented to match these additional requirements.

The health professional moreover suggested an integration of special software used in medical practices and hospitals. This could improve the user acceptance of the system, as it reduces time and effort to become acquainted with the new possibilities, compared to introducing an additional application.

6 Conclusion

The question of relevant data is not trivial as relevance is strongly dependent on the considered use case. The more data is available for an insurance, the more relations between behavior patterns and insurance claims can be found and predicted. Functional requirements are defined by the structure of the collected data, the use cases and the processes defined by the stakeholders. Data governance and management as well as national and international health and insurance standards, determine further requirements. Especially important non-functional requirements for the analyzed use cases are availability, performance and extensibility.

Two models were chosen to display the proposed architecture. An ER model defines the composition of the data whereas the ArchiMate model describes the interaction between the components and actors. It is difficult to examine the data architecture independent

from applications and technologies. Therefore, we do not intend to provide a generally valid model, but rather give an understanding of the steps necessary to derive requirements and a corresponding architecture from given use cases.

In future work, the data architecture has to be adapted as indicated in chapter 5 and further validations will be required. Additionally, an investigation of the state of practice in several health insurances could provide further insights on correctness and completeness of the proposed data architecture.

References

[BCS91] Batini, C.; Ceri, S.; Shamkant, B.: Conceptual database design: an Entity-relationship approach, The Benjamin/Cummings Publishing Company, 1991.

[Br17] Bresnick, J.: How Health Information Exchange Models Impact Data Analytics. Available at: http://healthitanalytics.com/news/how-health-information-exchange-models-impact-data-analytics.

[DM17] The Data Management Association: DAMA Guide to the Data Management DMBOK. Available at: https://www.dama.org/content/body-knowledge.

[ES07] Esakkirajan, S.; Sumathi, S.: Fundamentals of Relational Database Management Systems, 2007. Springer, Berlin, pp. 31-33.

[Fi16] Fitnessarmband.eu: Fitness-Tracker Test: Wir haben die Testsieger 2016!, available at: http://fitnessarmband.eu/krankenkassen-bezuschussen-fitness-armbaender/, accessed: 2017-01-03.

[Gi17] Gillis, C.: Customer Presentation, Big Data Alliance Hewlett Packard Enterprise -Enterprise Services and Software.

[GIW97] Graziano, K.; Inmon, W.; Silverston, L.: The Data Model Resource Book: A Library of Logical Data Models and Data Warehouse Designs, John Wiley & Sons, 1997.

[Ki16] King, T.: What is Data Virtualization? Solutions Review, 2016, Available at: http://solutionsreview.com/data-integration/what-is-data-virtualization/, accessed: 2017-01-17.

[Mü16] Müller, J.: PKV - 2017 kommt der erste digitale Krankenversicherer, available at: http://www.versicherungsbote.de/id/4842298/PKV-2017-digitaler-Krankenversicherer/, accessed: 2017-01-03.

[OG12a] Open Group: ArchiMate 3.0 Specification: Introduction, 2012. Available at: http://pubs.opengroup.org/architecture/archimate3-doc/chap01.html#_Toc451757908, accessed: 2017-01-12.

[OG12b] Open Group: ArchiMate 3.0 Specification: Application Layer, 2012. Available at: http://pubs.opengroup.org/architecture/archimate3-doc/chap09.html#_Toc451758026, accessed: 2017-01-12.

[Ru14] Rupp, C. & die SOPHISTen: Requirements-Engineering und –Management, Hanser Verlag, München, 2014, p.13.

[Sa12] El-Sappagh, S. et al.: Electronic Health Record Data Model Optimized for Knowledge Discovery. In: International Journal of Computer Science, Vol. 9, Issue 5, No 1, September 2012, pp. 329-338.

[Se09] Seltman, H. J.: Experimental Design and Analysis, Carnegie Mellon University, Pittsburgh, 2009.