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

doi:10.3233/978-1-61499-432-9-773

Health Information Research Platform (HIReP) – An Architecture Pattern

Björn SCHREIWEIS[a,1], Gerd SCHNEIDER[a], Theresia EICHNER[a], Björn BERGH[a] and Oliver HEINZE[a]
[a] *Center for Information Technology and Medical Engineering, Heidelberg University Hospital, Germany*

Abstract. Secondary use or single source is still far from routine in healthcare, although lots of data are available either structured or unstructured. As data are stored in multiple systems, using them for biomedical research is difficult. Clinical data warehouses already help overcoming this issue, but currently they are only used for certain parts of biomedical research. A comprehensive research platform based on a generic architecture pattern could increase the benefits of existing data warehouses for both patient care and research by meeting two objectives: serving as a so called single point-of-truth and acting as a mediator between them strengthening interaction and close collaboration. Another effect is to reduce boundaries for the implementation of data warehouses. Taking further settings into account the architecture of a clinical data warehouse supporting patient care and biomedical research needs to be integrated with biomaterial banks and other sources. This work provides a solution conceptualizing a comprehensive architecture pattern of a Health Information Research Platform (HIReP) derived from use cases of the patient care and biomedical research domain. It serves as single IT infrastructure providing solutions for any type of use case.

Keywords. Clinical data warehouse, secondary use, primary use, architecture, single source

Introduction

More and more clinical information is available electronically within modern healthcare environments [1]. However reusing clinical information for research based upon a data warehouse infrastructure is still rare. Only few projects doing that do exist [2, 3]. One issue is data availability as electronic medical/health records (EMRs/EHRs) are designed for clinical documentation and restrict access to data on patient care settings. Another issue is data quality. Data are captured in different settings redundantly and hence it is unclear which data are representing the truth.

A clinical data warehouse might help integrate different data sources and thus enable physicians to get a holistic view of the patient. Concerning patient care it would also enable the treating physician to compare the patient's visit to similar visits and make the right conclusions. A clinical data warehouse could also be the single point-of-truth required to improve data quality and hence be a data source for biomedical research.

---

[1] Corresponding Author.

An architecture pattern of a research platform has to implement patient care and medical research requirements to better integrate both parts. A clinical data warehouse is meant to be one component of a research platform. Further components necessary for integration of patient care and biomedical research are identified in this study. The objective is to define a generic architecture pattern called *Health Information Research Platform (HIReP)* is defined to better integrate both parts.

1. Methods

The Design and Creation strategy as described by Oates [4] was used. Three data sources were combined in a triangulation to minimize weaknesses in the results. Data sources were use cases, laws and guidelines and an expert panel discussion. First, use cases, which the architecture pattern of a research platform is supposed to implement, were defined. Afterwards, an analysis of German data privacy laws and data privacy guidelines by the *Technology, Methods, and Infrastructure for Networked Medical Research* [5, 6] (acknowledged non-profit association) (TMF e.V.) was performed. This literature analysis was conducted to find out about the specific regulations for research in healthcare settings. Finally the resulting architecture pattern was discussed with an expert panel (TMF working group on data privacy). All results were merged and resulted in the architecture pattern described in the following chapter.

2. Results

2.1. Use cases driving the HIReP approach

The HIReP approach is driven by two major use cases derived from enquiries of the National Center for Tumor Diseases Heidelberg (NCT), the Biomaterial Bank Heidelberg (BMBH) and various departments at Heidelberg University Hospital.

One use case is care-oriented and describes the combination of two different information sources to support patient treatment. This primary use of biomedical information includes a clinical data warehouse and a biomaterial bank. The intended purpose of the clinical data warehouse is to provide an integrated view on patient’s health information to support physicians at Heidelberg University Hospital in medical treatment. It integrates structured and unstructured data from multiple data sources and allows for analyzing unstructured documents. Additionally, data can be extracted from unstructured documents and transformed to structured data for further use. All clinical information can be found in the clinical data warehouse. A context integration between EMR/EHR and clinical data warehouse allows for a holistic view of the currently in the EMR/EHR selected patient. The clinical data warehouse also has functionality for comparison of the selected patient’s current visit to similar visits of other patients in an anonymized view. In addition to the health information captured in the context of treatment it is of use for the physicians to query the BMBH for available biomaterial like patient’s tissue or liquor samples. The query’s results should be integrated into the patient’s health information view of the clinical data warehouse. This additional information fosters the physicians’ decision on further analysis of available biomaterial in support of patient treatment.

The second use case driving the HIReP approach is research-oriented and based on needs of the NCT which is interested in patients' health information captured in the context of treatment as well as in information on available biomaterial. This information is used for research purposes, therefore it has to be either anonymized or - in case there is an informed consent of the patient - pseudonymized medical information. This is contrary to the first use case which requires personalized medical information for patient care. Within the research-oriented use case the information providers are the clinical data warehouse at Heidelberg University Hospital and the BMBH, whose interaction with the information consumer (NCT) could be either separated or combined. The separated mode means that the consumer queries each provider individually whereas the combined mode allows querying various sources with a single request. This request gets forwarded to the addressed information providers and the results of all providers get fused and returned to the consumer. E.g. as a biomaterial bank does not include additional health information of the patient a sample belongs to, the combined approach allows querying for biomaterial of patients with specific health criteria. The identification of patients matching to the given health criteria takes place in the clinical data warehouse.

The HIReP approach provides an integrated and extensible solution for both mentioned use cases.

2.2. High-level architecture of HIReP

Based upon the described use cases and with the prospective extension by additional systems in mind, a generic architecture for HIReP not limited to the initial use cases' scope was designed. Furthermore, it was decided to comply with the restrictive German and European data protection and security guidelines [5-7] for networked medical research. The combination of these guidelines allows for the integration of patient care and research in one platform and results in conformance to the required informational separation of powers. As the second use case matches to the research category and the designed architecture also supports the care-oriented first use case, the HIReP slogan *"primary use enabled, secondary use compliant"* was composed.

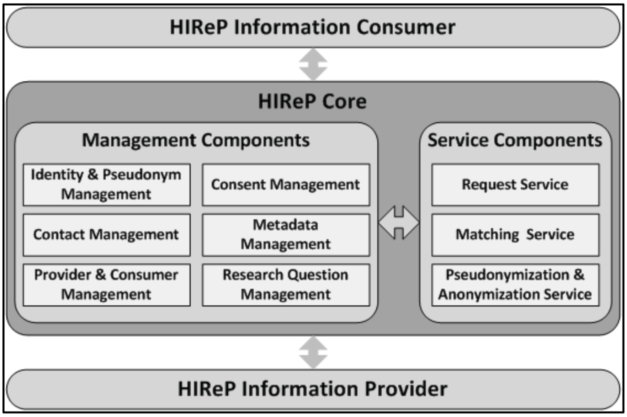

Figure 1. High-level architecture of HIReP

The HIReP architecture shown in Figure 1 is divided into three parts: *HIReP Information Consumer*, *HIReP Information Provider* and the joining *HIReP Core*. The *HIReP Core* is divided into the interconnected sections *Management Components* and *Service Components*. The *Management Components* [8] are characterized by low data variability and low data throughput. Patient's identifying data and related pseudonyms resulting from the dissemination of patient's health information for research purposes are administered by the component *Identity & Pseudonym Management*. This allows for de-pseudonymization in justified cases, e.g. major research breakthroughs supporting further treatment of the patient. The component *Consent Management* keeps track of patients' informed consents related to the further dissemination of their health information for medical research. Based upon an informed consent the health information can be disseminated pseudonymized otherwise anonymized. The *Contact Management* [9] component allows addressing and informing patients in case of justified de-pseudonymization. As HIReP enables flexible information consumer and information provider extension the architecture includes on the one hand a *Provider & Consumer Management* and on the other hand a *Metadata Management*. The *Provider & Consumer Management* is a service registry with an extension to define and administer permitted connections between consumers and providers. This is necessary due to data security reasons as for example a research-oriented information consumer shall only be allowed to connect to information providers relevant to the consumer's research question. The component *Metadata Management* is a dictionary enabling an integrated view on the data that could be provided via HIReP. All information providers have to add information about their data to this dictionary which can be browsed by information consumers to determine if relevant information will be available. The final component within the section *Management Components* is the *Research Question Management*. As a research-oriented information consumer has to apply for access to health information relevant for research it has to define the research question and has to justify the necessity to access the information. Such an application can be prepared via the component *Research Question Management*. Subsequently, this application gets evaluated by an expert panel and in case of acceptance all necessary HIReP configurations to setup the information flow will be carried out.

In contrast to the *Management Components* the *Service Components* are characterized by high data throughput. The component *Request Service* is responsible for handling unique information requests or information subscriptions from information consumers and involves the additional components *Matching Service* and *Pseudonymization & Anonymization Service*. The *Matching Service* supports the *Request Service* at handling complex information requests comprising various information providers and returns a merged result to the *Request Service*. Finally, the *Pseudonymization & Anonymization Service* gets involved in case the requested health information is used for research purposes. Therefore the service must work closely with the components *Consent Management* and *Identity & Pseudonym Management*. If an informed consent of the patient is available via *Consent Management* the health information can be delivered pseudonymized and the service removes the patient's identifying data from the health information and adds a pseudonym. To keep track of all pseudonyms generated for a patient the service adds each new pseudonym to the *Identity & Pseudonym Management*. In case there is no informed consent available the service removes all identifying data from the health information and delivers it anonymized.

2.3. Initial HIReP setup

According to the driving use cases the initial setup of HIReP includes the clinical data warehouse at Heidelberg University Hospital and the BMBH as information providers. Information consumers are the NCT's research database and the clinical data warehouse at Heidelberg University Hospital for care purposes. As the clinical data warehouse acts as provider and consumer the term *prosumer* is used to emphasize its combined role. Furthermore the initial *HIReP Core* includes only the four components *Identity and Pseudonym Management*, *Request Service*, *Matching Service*, and *Pseudonymization and Anonymization Service*. Based upon this setup the feasibility of the HIReP approach can be verified, especially the operability of the designed interfaces between information providers/consumers and the realized components of the *HIReP Core* can be tested and the initial HIReP use cases can be configured and executed.

3. Discussion

In contrast to other system architectures supporting secondary use of biomedical data [2, 3], the *HIReP approach* replicates data only once prior to making it available for research. The *Provider & Consumer Management* as part of the *HIReP Core* allows for new services to be registered and - in combination with adding metadata to the *Metadata Management* - can be queried for data. This functionality helps to easily add/remove or change modules in the architecture. Thus each data store can be consumer, provider or prosumer of data.

References

[1] Köpcke F, Trinczek B, Majeed RW, Schreiweis B, Wenk J, Leusch T, Ganslandt T, Ohmann C, Bergh B, Röhrig R, Dugas M, Prokosch HU. Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. *BMC Med Inform Decis Mak.* 2013; **13**(2013):37-44.

[2] Prokosch HU, Beck A, Ganslandt T, Hummel M, Kiehntopf M, Sax U, Uckert F, Semler S. IT Infrastructure Components for Biobanking. *Appl Clin Inform,* 2010, **1**(4), 419-429.

[3] Gerlich MG, Moller A, Schafer C, Strohbach AK, Krafczyk-Korth J, Hoffmann W. Epidemiology of multi-resistant organisms - challenges to a regional data management system. *GMS Krankenhhyg Interdiszip,* 2011, **6**(1), Doc26.

[4] Oates BJ. Researching Information Systems and Computing. London: Sage; 2006.

[5] Simon JW, Paslack R, Robienski J, Goebel JW, Krawczak M. *Biobanking – legal framework.* Berlin: MWV; 2006.

[6] Reng CM, Debold P, Specker C, Pommerening K. *Generic data protection solutions for medical research networks*. Berlin: MWV; 2006.

[7] European Commission, Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data. Brussels; 2012.

[8] Schwanke J, Löhnhardt B, Bahls T, Borg A, Drepper J, Geidel L, Heinze O, Lablans M, Langner D, Muscholl M, Quade M, Schreiweis B, Pommerening K. Standpunktpapier für die Notwendigkeit der Harmonisierung von Lösungsansätzen für das Probandenmanagement in medizinischen Forschungsnetzen. 2013.

[9] Schwanke J, Rienhoff O, Schulze TG, Nussbeck SY. Suitability of customer relationship management systems for the management of study participants in biomedical research. *Methods Inf Med,* 2013; **52**(4); 340-350.