

HiGHmed – An Open Platform Approach to Enhance Care and Research across Institutional Boundaries

Birger Haarbrandt^{1*}; Björn Schreiweis^{2*}; Sabine Rey³; Ulrich Sax³; Simone Scheithauer⁴; Otto Rienhoff³; Petra Knaup-Gregori⁵; Udo Bavendiek⁶; Christoph Dieterich⁷; Benedikt Brors⁸; Inga Kraus³; Caroline Marieken Thoms³; Dirk Jäger⁹; Volker Ellenrieder¹⁰; Björn Bergh¹¹; Ramin Yahyapour¹²; Roland Eils^{13,14}; HiGHmed Consortium; Michael Marschollek¹

¹Peter L. Reichertz Institute for Medical Informatics, TU Braunschweig and Hannover Medical School, Germany;

²Institute for Medical Informatics and Statistics, Kiel University and University Medical Center Schleswig-Holstein, Campus Kiel, Germany;

³Department of Medical Informatics, University Medical Center Goettingen, Goettingen, Germany;

⁴Central Division of Infection Control and Infectious Diseases, University Medical Center Goettingen, Goettingen, Germany;

⁵Institute of Medical Biometry and Informatics, University Hospital Heidelberg, Heidelberg, Germany;

⁶Department of Cardiology and Angiology, Hannover Medical School, Hannover, Germany;

⁷Section of Bioinformatics and Systems Cardiology, Department of Internal Medicine III, Klaus Tschira Institute for Integrative Computational Cardiology, University Hospital Heidelberg, Heidelberg, Germany;

⁸Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany;

⁹Department of Medical Oncology, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Heidelberg, Germany;

¹⁰Department of Gastroenterology and Gastrointestinal Oncology, University Medical Center Goettingen, Goettingen, Germany;

¹¹Institute for Medical Informatics and Statistics, Kiel University and University Medical Center Schleswig-Holstein, Germany;

¹²Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG), University of Goettingen, Goettingen, Germany;

¹³Digital Health Center, Berlin Institute of Health (BIH) and Charité, Berlin, Germany;

¹⁴Health Data Science Unit, University Hospital Heidelberg, Heidelberg, Germany

Keywords

Secondary use, information systems, medical informatics, decision support systems, data security, semantic interoperability, EHR

Summary

Introduction: This article is part of the Focus Theme of Methods of Information in Medicine on the German Medical Informatics Initiative. HiGHmed brings together 24 partners

from academia and industry, aiming at improvements in care provision, biomedical research and epidemiology. By establishing a shared information governance framework, data integration centers and an open platform architecture in cooperation with independent healthcare providers, the meaningful reuse of data will be facilitated. Complementary, HiGHmed integrates a total of seven Medical Informatics curricula to develop col-

laborative structures and processes to train medical informatics professionals, physicians and researchers in new forms of data analytics.

Governance and policies: We describe governance structures and policies that have proven effective during the conceptual phase. These were further adapted to take into account the specific needs of the development and networking phase, such as roll-out, care-related aspects and our focus on curricula development in Medical Informatics.

Architectural framework and methodology: To address the challenges of organizational, technical and semantic interoperability, a concept for a scalable platform architecture, the HiGHmed Platform, was developed. We outline the basic principles and design goals of the open platform approach as well as the roles of standards and specifi-

Correspondence to:

Birger Haarbrandt
Peter L. Reichertz Institute for Medical Informatics
of TU Braunschweig and Hannover Medical School
Muehlenpfordtstr. 23
38106 Braunschweig
Germany
E-mail: Birger.Haarbrandt@ppli.de

Methods Inf Med 2018; 57(Open 1): e66–e81

<https://doi.org/10.3414/ME18-02-0002>

received: February 19, 2018

accepted: May 26, 2018

* These authors contributed equally to this work.

cations such as IHE XDS, openEHR, SNOMED CT and HL7 FHIR. A shared governance framework provides the semantic artifacts which are needed to establish semantic interoperability.

Use cases: Three use cases in the fields of oncology, cardiology and infection control will demonstrate the capabilities of the HiGHmed approach. Each of the use cases entails di-

verse challenges in terms of data protection, privacy and security, including clinical use of genome sequencing data (oncology), continuous longitudinal monitoring of physical activity (cardiology) and cross-site analysis of patient movement data (infection control).

Discussion: Besides the need for a shared governance framework and a technical infrastructure, backing from clinical leaders is a

crucial factor. Moreover, firm and sustainable commitment by participating organizations to collaborate in further development of their information system architectures is needed. Other challenges including topics such as data quality, privacy regulations, and patient consent will be addressed throughout the project.

1. Introduction

The need to use the wealth of data residing in electronic health records for research and improvement of care for the individual is widely acknowledged [1, 2, 3, 4]. Yet, the vast amount of contemporary hospital information systems is shaped by disparate and proprietary application systems and databases. As these systems frequently lack standardized data definitions and openly accessible messaging interfaces, acquiring routine and research data from these “data silos” in a timely fashion and with sufficient data quality is challenging [5, 6]. To overcome these barriers, clinical data warehouses have been established within many organizations [7, 8, 9]. However, the push towards precision medicine, systems medicine and the emerging paradigm of learning healthcare systems, requires a holistic approach that allows to utilize data collections spread across multiple healthcare providers [10, 11].

To give an example, molecular stratification sub-divides cancer entities into increasingly smaller subgroups, making it difficult even for larger cancer centers to recruit enough patients into stratified clinical trials. This applies even more to translational research on cancer subgroups, which also requires that data is linked to powerful biobanks of tumor samples. Thus, combining information on prevalence of stratifying biomarkers and on their predictive power with respect to different therapy options makes cross-center, national or even international data access a bare necessity [12].

As a prerequisite to facilitate the reuse of once collected data effectively for the benefit of patients and the wider population, hospitals need to join forces and transform

their sociotechnical infrastructures from an isolated and function-based approach to a collaborative and data-driven way of

thinking. While a main aspect for the paradigm shift lies in the successful cultural change to establish trustful collaboration and sharing of data between organizations, an underlying technical framework needs to be carefully engineered to support the richness and complexity of clinical data, workflows, analysis methods, data privacy and security considerations, and access policies.

The HiGHmed core consortium brings together three university hospitals – Heidelberg University Hospital (UKL-HD), University Medical Center Goettingen (UMG), and Hannover Medical School (MHH) – along with a further research institution, the German Cancer Research Center (DKFZ). The consortium is complemented by a line-up of 20 partners, including research institutions, industry and other associated partners (► Table 1).

By establishing a shared information governance structure between all participating clinical sites in accordance with the FAIR principles of data reuse [13], and by iteratively introducing an extensible and scalable software architecture, the translation between care provision and biomedical research will be enhanced. Newly founded and dedicated organizational units called *Medical Data Integration Centers* (MeDICs) and the Omics Data Integration Center (Omics DIC) at DKFZ will hold the competencies and resources needed to implement the HiGHmed concepts and to manage, integrate and provide clinical and research data. To demonstrate the capability of the technical design and the organizational structure, the MeDICs take over the responsibility to demonstrate the suitability of the HiGHmed approach by supporting three medical use cases in

Table 1 HiGHmed partners.

Core Partners	Heidelberg University Hospital/ Medical Faculty (UKL-HD)
	University Medical Center Goettingen (UMG)
	Hannover Medical School (MHH)
	German Cancer Research Center (DKFZ)
Project Partners Academia	Hasso Plattner Institute, Potsdam (HPI)
	HAWK Hochschule Hildesheim/ Holzminen/Goettingen (HAWK-HHG)
	Heidelberg University (UHD)
	Heilbronn University (HHN)
	Helmholtz Center for Infection Research (HZI)
	Hochschule Hannover (HSH)
	Robert-Koch-Institut (RKI)
	TU Braunschweig (TU-BS)
	TU Darmstadt (TUD)
	Private Project and Networking Partners
InterComponentWare AG (ICW)	
NEC	
SAP	
Associate Partners	Siemens Healthcare GmbH
	German Heart Foundation
	Heidelberger Selbsthilfebüro (self-help organization)
	openEHR Foundation
	HL7 Deutschland e. V.
	DellEMC

the domains of oncology, cardiology and infection control.

Apart from the technical, methodological and organizational infrastructure for sharing, re-using and analyzing data, a strong focus will be put on educational advancement. Clinical personnel, researchers and Medical Informatics professionals need continuous education on decision support, data-driven research and health data management. Hence, seven educational and two industry partners have combined forces, and have committed themselves to strengthen existing Medical Informatics curricula, and to develop and implement new forms of education for computer scientists, researchers, physicians, nurses and other health professionals.

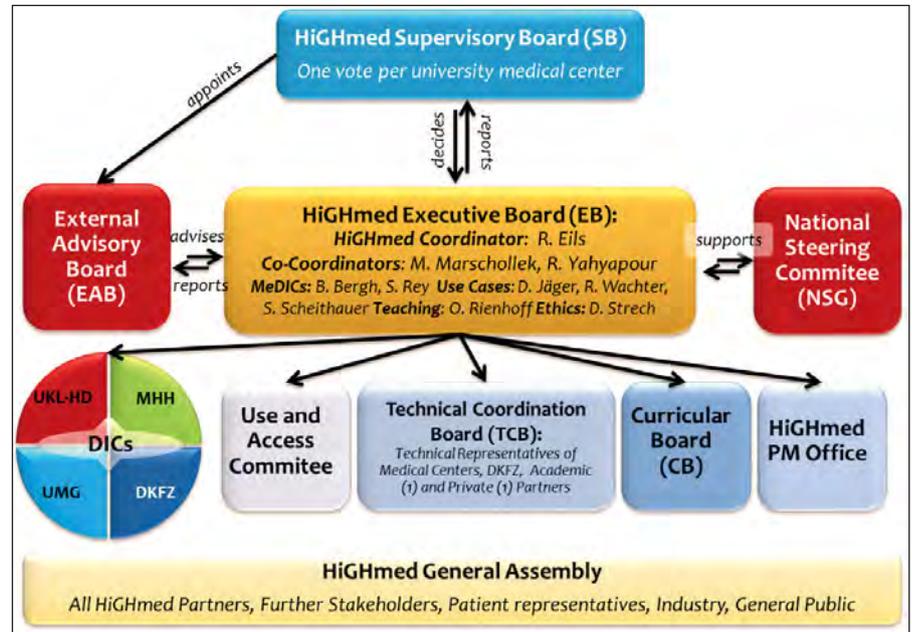


Figure 1 Overview of the HiGHmed governance structure.

2. Governance and Policies

2.1 HiGHmed Governance Structure

The consortium is directed by a HiGHmed coordinator and two appointed deputies, and is based on the following bodies that govern the consortium (see also ► Figure 1):

2.1.1 Executive Board

HiGHmed's major management body is the Executive Board (EB), which bears fiduciary responsibility for overall operations within HiGHmed as well as budgetary responsibility. The EB is comprised of the HiGHmed coordinator and its deputies, speakers of the local MeDICs, speakers of HiGHmed's clinical use cases oncology, cardiology and infection control, the teaching program coordinator, and an ethics and stakeholder representative. EB's major tasks are progress monitoring of the project, quality assurance and risk management, conflict resolution and interaction with all HiGHmed bodies and stakeholders involved, expansion and strengthening of partnerships, collaboration with the Medical Informatics Initiative coordination office and the National Steering Committee, and planning and implementing measures for sustainability. The EB establishes and

dissolves additional working groups and positions based on organizational needs.

2.1.2 Supervisory Board

HiGHmed's Supervisory Board supervises the policy pursued by the EB and its managerial duties. The Supervisory Board consists of representatives of the management boards of HiGHmed's participating university medical centers, including scientific and administrative executives. In the case of non-unanimous decisions, each core university medical center will have one vote. All fundamental decisions regarding the progress of HiGHmed, the role of partners and the acceptance of new partner hospitals, will be prepared in the EB and decided in the Supervisory Board. Both, the Supervisory Board and External Advisory Board provide important strategic counsel to HiGHmed's EB, taking into account the interests of all HiGHmed stakeholders.

2.1.3 Technical Coordination Board

The Technical Coordination Board supports the EB in all technical aspects and coordinates the technical part of the HiGHmed project. The Technical Coordination Board consists of representatives of

technical executives of HiGHmed's participating university medical centers as well as representatives from academic and industrial partners. One member of the project management team will support the work of the Technical Coordination Board. The Technical Coordination Board reports to the EB.

2.1.4 Project Management

To support the HiGHmed management structure, a project management office under the supervision of the HiGHmed coordinator, interfaces with local project management in the MeDICs. Additionally, the project management office will coordinate support in communication, marketing and reporting activities.

2.1.5 Curricular Board

For the design and implementation of a HiGHmed Educational Concept and its alignment with HiGHmed's other activities, a curricular board will develop teaching formats usable at different sites and for different target groups addressing major issues in precision medicine. The Curricular Board reports to the EB.

2.1.6 Use and Access Committee

To guarantee timely, legitimate and secure provision of data to researchers in accordance to all HiGHmed and national policies, each university medical center operates a *Use and Access Committee* (UAC). Supported by service-oriented *Data Transfer Units*, selected representatives from multiple stakeholder groups, including physicians, researchers and data protection commissioners, decide transparently and objectively on data access applications from internal and external requestors. The *Use and Access Committee* will seek regular consultancy by patient representatives to safeguard patients' interests in its governance.

2.1.7 Ethics Working Group

HiGHmed's leading experts for practice-oriented ethics and governance research with manifold experiences in empirical stakeholder research (social, political, cultural sciences), philosophy, and meta-research undertake interdisciplinary research on ethical, policy and social issues with regard to the collection, linking, and use of patient data in medical research and health care. A central aim is the evidence-based development of practice-oriented frameworks and recommendations for governance solutions, with immediate relevance for the consortium as well as for the ongoing national and international debates in the field. The Ethics Working Group reports to the EB.

2.1.8 External Advisory Board

The External Advisory Board (EAB) ensures both alignment of the HiGHmed concept with international activities and compliance with international standards. It advises the EB regarding HiGHmed's project progress and continuation. The External Advisory Board consists of international experts in the field of Medical Informatics, who will be supported by representatives from patient organizations to receive feedback about patient participation in research and the quality of service the patients have received. Therefore, the board will also provide a basis to moni-

tor the progress of HiGHmed's core principle 'patients first'.

2.1.9 General Assembly

In order to maximize the impact of the broad expertise provided by HiGHmed's project and associated partners, and to include all stakeholders in the development of HiGHmed, representatives of all HiGHmed partners meet annually in the General Assembly. During the General Assembly, the partners are informed about the activities within the HiGHmed consortium and the progress of the HiGHmed project and jointly discuss the options for the future of the HiGHmed partnership. The General Assembly contains a public part, to which external academic and industrial experts and stakeholders are invited.

2.2 Integration of Further Clinical Sites and Partners

The HiGHmed concept is designed as an open platform and HiGHmed will actively recruit new partners for the implementation of HiGHmed's technology platform. Following the integration of further university medical centers to the HiGHmed consortium, representatives of these sites will join the respective governing bodies, namely the Executive Board and Supervisory Board.

HiGHmed's teaching program is specifically designed to integrate with other curricula in the context of the national digital agenda of the Federal Ministry of Education and Research (BMBF)¹, so that a swift integration of other university medical centers into our curricula development program is envisioned.

2.3 Interaction with the National Steering Committee

Technical and management representatives of HiGHmed's governance bodies closely interact with the national Medical Informatics coordination office, the National Steering Committee (NSG), to support

joint collaboration of the consortia within the Medical Informatics Initiative. The National Steering Committee plays a key role regarding aspects of semantic and syntactic interoperability. Besides a strong contribution of HiGHmed's representatives to all National Steering Committee working groups, HiGHmed will assist in the establishment of a strategy and implementation process to enable consortia partners to use SNOMED CT in Germany.

3. Architectural Framework and Methodology

To support data management and data sharing in the context of patient care, research and knowledge management on an institutional, regional, national and international scale alike, a diverse set of technical and non-technical requirements needs to be addressed by the MeDICs. This holistic view on care and research demands a multi-layered architecture to build an integrated platform that supports current and future use cases including recent developments within omics technologies.

Based on a common understanding of the pressing challenges and needs for a collaborative information and knowledge sharing, we have defined core principles to guide our considerations for the technical and organizational design of a shared socio-technical process, hereinafter referred to as the HiGHmed Platform:

1. **Patients first:** All activities in HiGHmed are centered on the question what benefits patients most. Therefore, in addition to fine-grained use-and-access control, patients will be able to view and obtain their health data through user-friendly tools (e.g. mobile apps, patient portals) similar to the *Personal Electronic Health Record* (PEHR) approach [14, 15]. By unlocking health data for patients, potential barriers in the patient-caregiver relationship may be reduced. Gaining better insights into their own health situation results in better health literacy, which might have positive effects on health care quality and outcome [16]. In close cooperation with patient organizations, a human-centered design approach and investi-

¹ <https://www.bmbf.de/en/index.html>

- gations into how health data could be best visualized and presented to empower patients, will be implemented.
2. **Data Safety and Privacy:** Each step in planning the HiGHmed Platform has been undertaken by considering the highest level of data safety and data privacy including the patients' right of informational self-determination. Medical data may only be processed for research purposes after a patient has given his or her informed consent. Based on a fine-grained use-and-access control mechanism, usage of medical data by individuals and organizations is strictly controlled, documented and requires authorization through the patient's consent. Patients have the right to revoke their consent at any time.
 3. **Clinical Relevance:** All components of the HiGHmed Platform are defined along the needs of healthcare professionals and clinical researchers. Clinical relevance, not mere technical feasibility, is the driver for our clinical use cases and design choices. Data management and semantic interoperability have been given the priority.
 4. **Clinically-led Data Modeling:** It is widely accepted that interoperability is not only a technical challenge. Interoperability must be negotiated between stakeholders. For this reason, healthcare professionals and researchers alike need to be engaged actively to establish semantic interoperability. This creates a new clinico-technical role, the so-called data stewards. Data stewardship implies the professional long-term care of data from design to processing and sharing data for the different purposes (e.g. research, training). Consequently, semantic interoperability needs to be tackled from a domain specialists' view by separating technical implementation aspects from the activity of domain modeling. In particular, we focus on practice-proven, technology-independent formalisms based on open international standards with established tool-support, elaborated artifact versioning, and model lifecycle management.
 5. **Semantic Traceability:** Semantic models need to be computable and should not be created for specific messaging or document standards, but serve as definitive models of semantics that can be used to generate downstream artifacts for different purposes such as the definition of user entry forms, XML schemas, data validation schemas, REST APIs, database queries etc. [17]. We consider data steward-supported single-source modeling as a mandatory precondition in a distributed environment to enforce semantic traceability through all system layers at all sites, thus avoiding errors in the local implementation of semantic artifacts.
 6. **Proven Technology:** To meet the ambitious goals of HiGHmed, we prefer robust and practice-approved technology supported by industry and solidly funded open source projects over research prototypes and proof-of-concepts. Practice-approved to us means wide adaption and incorporation of the technology by regional and national eHealth programs, hospitals and health trusts.
 7. **Scalability:** While a technical solution needs to be optimized to handle high volumes of complex as well as constantly changing information and clinical workflows, a good balance between technical standardization on the one hand, and flexibility regarding the organization and configuration of each particular MeDIC on the other hand, is needed. For example, automated provision of data to researchers in terms of export formats should be highly flexible in each MeDIC: from export to well-known analytics tools to contemporary data persistence technologies such as graph databases, semantic triple stores, column-based databases, map-reduce frameworks etc. At the same time, we seek to minimize the requirements needed to connect to the HiGHmed Platform.
 8. **Sustainability:** It is certain that the underlying technology of the platform will be affected by progress and changes in information technologies. While, for example, REST interfaces are the state of the art for many application scenarios in web-based environments, it is not possible to foresee the development of technology within the next ten years. In addition, new developments in database technologies need to be incorporated seamlessly without introducing new technological dependencies. Therefore, we seek to externalize knowledge artifacts as clinical content models, terminologies, guidelines, algorithms, queries etc. by expressing them in technology-neutral and open formats, allowing the data and their definitions to migrate from one technology stack to another efficiently.
 9. **Decentralization:** Patient data will reside at each site. Until access is needed by a network participant and patient consent has been given, no data is transferred to another (partner's) data repository of any kind. Furthermore, we see vast potential in privacy-preserving methodologies and approaches transferring analysis algorithms towards data instead of vice versa.
 10. **Enable Innovation:** Finally, a technical infrastructure that shall serve a great variety of use cases at the intersection of care provision and research needs to be highly extensible. Therefore, we only consider technologies as appropriate that facilitates the development of new clinical applications, mobile apps and analytics. Omics technologies and sensor data of patient medical devices are already part of new clinical processes. Within HiGHmed, we will incorporate ongoing progress of leading biomedical research in the omics field and health-enabling technologies.

3.1 Open Health Platform Architecture

Based on these core principles, we defined an interoperable, open health data platform, the *HiGHmed Platform*. In accordance with the definition provided by Beale [18], the HiGHmed Platform is designed as a coordinated process that steers the iterative definition of open service interfaces, open system specifications, operating procedures and open clinical information models. Consequently, the following characteristics are essential:

1. **Open Service Models:** all specifications of the provided application programming interfaces (APIs) are openly

accessible to everybody. Specifications include data security and privacy, electronic health record (EHR) management and database queries.

2. **Open Information Models:** All clinical models are well defined based on established open standards. Data based on these models can be reliably processed and computed in local and distributed environments. In addition, all models defined and reused in HiGHmed are made openly available to the community.
3. **Open System Specifications:** All system components and protocols are openly specified using licenses feasible for commercial and non-commercial use, so that every component in the system can be replaced by software from multiple vendors or open source communities.

By providing an open platform architecture, HiGHmed aims to avoid any mandatory procurement of proprietary solutions that would cause vendor lock-in. Instead, participants in HiGHmed are able to acquire relevant components from different vendors, open source initiatives or by self-development. This architecture aims to foster an ecosystem based on open service interfaces and clinical models. To a large extent, this approach follows the “Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap” by the US healthit.gov². Developers will be readily engaged in the development of software applications based on clinical and research data, fulfilling the requirements of patients, caregivers and researchers alike.

► Figure 2 provides a high-level view of the platform components and the local information systems on the examples of the MeDICs of Heidelberg and Goettingen. As a basic principle, the HiGHmed approach aims for a careful balance between site-specific configuration of the MeDICs and necessary standardization.

While the hospital information system environment will be the main source for

data, relevant information from different healthcare sectors is integrated and made available via the respective regional PEHR based on IHE Cross-Enterprise Document Sharing (Imaging) (XDS-I) infrastructures. Local integration layers are built autonomously, due to heterogeneous individual hospital information system landscapes, to optimally address site-specific needs and to ensure provision of data to the HiGHmed Platform in time and with high data quality. Most commonly, this is achieved by establishing extract, transform, load (ETL) processes and by utilizing a communication server. Furthermore, other cross-sectional and longitudinal data sets from registries, clinical trials and networked research projects will be of high relevance. To meet local data analysis requirements, extensions of the HiGHmed Platform, such as adding a *data lake* to the ETL processes, are possible.

The HiGHmed Platform consists of an IHE XDS Affinity Domain combined with an openEHR clinical data repository. IHE XDS-I and Cross-Enterprise Document Reliable Interchange (XDR) profiles form a coarse-grained and scalable distribution service. Fine-grained APIs for data provision and access will be established by openEHR through the Archetype Query Language (AQL) and a REST API based on SMART on FHIR. By using this combination, imaging data as well as (semi-) structured and unstructured data can be managed, distributed and provided within an organization based on shared information models. This way, the HiGHmed Platform provides means to develop a wide range of applications, ranging from small decision support apps, clinical registries and research databases to comprehensive electronic health record components. To support the HiGHmed use cases, new clinical application systems and data analysis apps will be directly developed against the HiGHmed Platform and deployed within the clinical care environment.

To complement the information management capabilities of the platform, knowledge services are connected to the MeDICs to provide external information sources, such as literature and biomedical databases. These are essential to improve clinical care and research activities. Gather-

ing medical information from different databases usually is a time-consuming manual task and needs extensive domain knowledge. A central knowledge connector component for the consortium will act as a “one-stop-shop” for information for users (via graphical user interface) and machines (via API), granting access to commercial and open databases by own API-based services.

To provide data to partners in HiGHmed, data needs to be de-identified and made accessible through a data federation service. On request for a certain research project and based on a patient's consent, his or her data is automatically pseudonymized by a trusted third party (GECKO Institute at Heilbronn University) and afterwards either transferred to the requesting partner or made available in a XDS-based research data mart. Patients will have access to their healthcare information and control access to this by healthcare providers. Exchange of record-level data is implemented based on IHE Cross-Community Document Reliable Interchange (XCDDR). IHE Cross-Community Access (XCA) is implemented to access data in XDS-based research data marts.

While in general, de-centralization of components is a leading principle, some services need to be implemented inter-organizationally. These services cover information model governance, terminology servers, de-identification, privacy-preserving record linkage, services to involve external parties and more. Since automated anonymization of free-text currently does not achieve 100% sensitivity [19], natural language processing and information retrieval have to be performed within the MeDICs. Thus, free-text data is never processed outside the MeDICs or even exchanged between the MeDICs. If possible, algorithms instead of data are to be exchanged, bringing algorithms to data not vice versa, to prevent massive data transfers for imaging and omics data.

In the long term, we consider the need of data mappings to cause lasting expenses. Therefore, based on the introduction of a collaborative information and knowledge governance process, participants in HiGHmed will have the capability and

2 <https://www.healthit.gov/sites/default/files/hie-in-teroperability/nationwide-interoperability-roadmap-final-version-1.0.pdf>

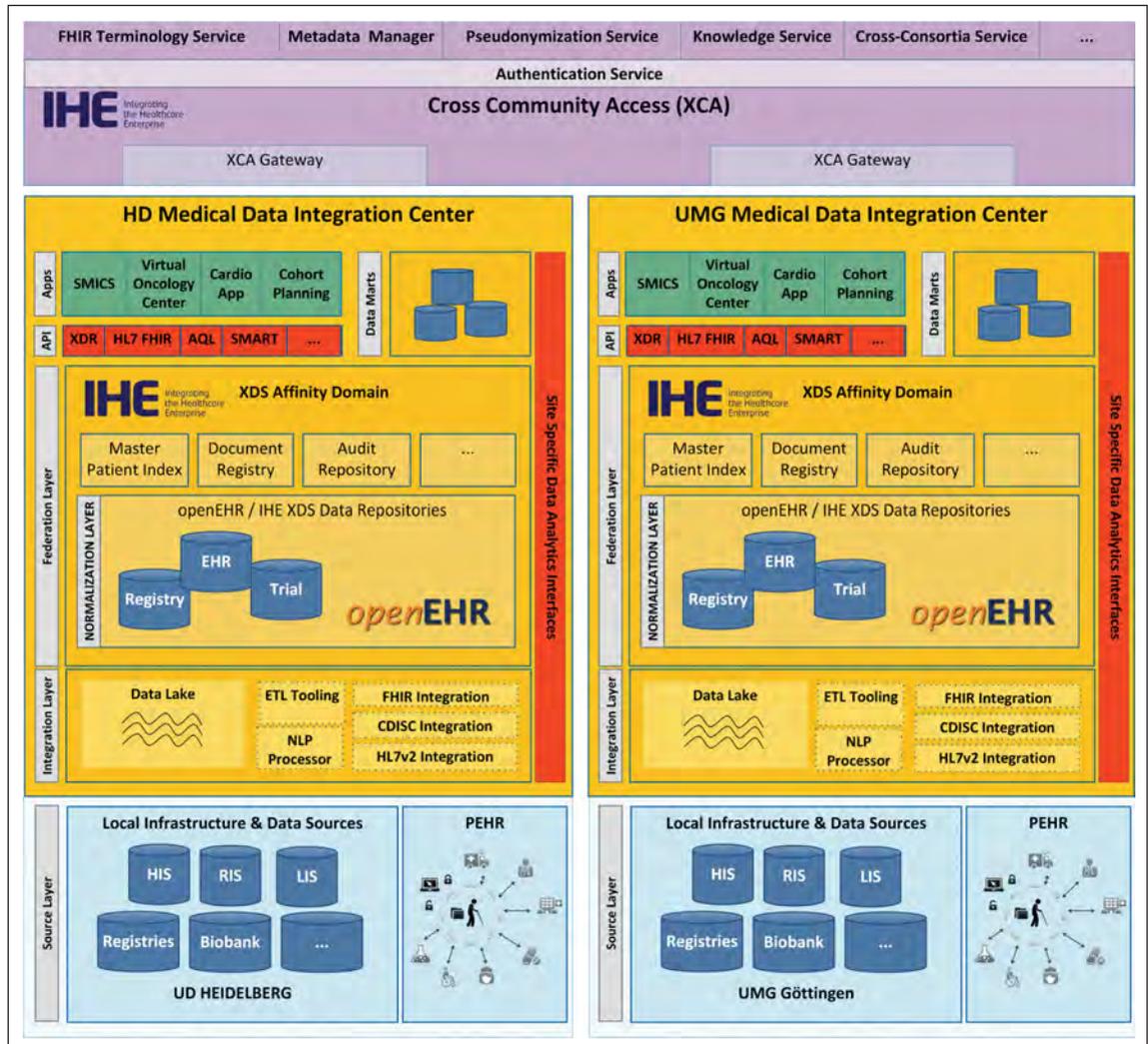


Figure 2

High level view of the HiGHmed technical infrastructure on the example of the medical universities of Heidelberg and Goettingen. Blue depicts the existing infrastructure of the partners; yellow depicts the MeDICs; purple depicts inter-organizational components.

tooling to iteratively harmonize their data semantics within clinical application systems and research databases. In doing so, newly developed and existing applications can increasingly produce data that can be directly loaded into the clinical data repositories of the MeDIC without laborious ETL-processes and extensive mappings that are costly and semantically lossy. This approach promises to greatly facilitate and streamline the provision of data within HiGHmed and to ensure sustainability.

3.2 Semantic Modeling

To achieve the ambitious goals of HiGHmed, we consider semantic modeling as one of the most critical parts as this activity will generate the necessary knowledge to safely access data in a cross-enter-

prise environment with high precision and recall. As eHealth standards and systems will continue to evolve over time, semantic domain models should be built as completely independent entities, separated from specific software products, solutions, or technologies, run by and for domain experts. Thus, we incorporate a multi-level modeling methodology that clearly separates domain knowledge from aspects of technical implementation. As introduced by the openEHR specifications, multi-level modeling uses a concept (called *archetypes*) that allows defining rich and computable metadata models of clinical information by applying constraints on a reference model [20]. An archetype defines the set of information and data structures related to a particular clinical concept. Importantly, the underlying formalism, the *Archetype Defi-*

nition Language (ADL) provides comprehensive modeling capabilities such as specialization/inheritance, nesting, terminology bindings, version tracking, lifecycle states etc. [21]. Archetypes follow a formal syntax to support the exchange and semantic interoperability of these concepts, but at the same time, they are very close to the mindset of clinicians, being easily edited and interpreted without deep knowledge of technical details. This allows to drive the development of standards by the real-world needs of physicians and to capture their exact clinical needs for complex use cases, which can be represented by arranging and further constraining archetypes within so called *templates* [22].

Archetypes have been adopted by the CEN/ISO 13606 standard and by the HL7 Clinical Information Modeling Initiative

(CIMI) [23, 24, 25]. Implying a separation between the software and database representation of data on the one side and the computable and sharable definition of information models and metadata models on the other side, the problem of representing complex and ever changing clinical data in software systems is simplified as changes in the domain layer can be more easily adapted in the software layer [26]. Once archetypes are defined, they represent definite models of semantics that can be used for multiple purposes, including the generation of data capture forms in EHRs, database schemas, transformations, messages, data validation algorithms, data querying, etc. In HiGHmed, we directly deploy archetypes and templates to openEHR repositories to guarantee semantic traceability throughout all sites. Above-mentioned software artifacts are thus automatically and directly created from the experts' domain understanding without the need for manual transformations. As a consequence, all agreed-upon semantics are enforced and guaranteed through all software layers at the vendor-independent data repositories of all network participants.

3.3 Information Governance Model

We adapted and adjusted the approach that is used by the national archetype governance model of Norway (Nasjonal IKT) to serve as a starting point for the information governance processes within HiGHmed [27, 28]. ► Figure 3 shows an overview of the HiGHmed governance process for identifying or defining clinical information models.

The first step is the identification of a need to standardize data items. In the early phase of HiGHmed, these needs will be determined by the selected use cases and by the input of stakeholders like researchers, clinicians and industry partners. At each site, the data stewards of the particular clinical domain take responsibility to gather local needs and to communicate with the superordinate *HiGHmed Modeling Group*, which is responsible for the organization of the cross-enterprise data modeling activities. Upon approval by the Modeling Group, the local data stewards investigate if an archetype is readily available from the international Clinical Knowledge Manager of the openEHR Foundation³ If not, a first draft archetype is created from scratch. HiGHmed intends to reuse as many archetypes as possible. This will allow to map data on models that have been internationally agreed upon and that are already in practice in several countries. Named reuse will save resources, as many archetypes have been created considering the input of several hundred clinical experts from around the world [29]. Before an archetype is incorporated, data stewards and other domain experts like physicians or researchers will discuss and contribute their local requirements. Additionally, after a technical and functional review, archetypes are approved by all participants and are ready for use in production. By contributing newly developed archetypes to the international community, domain experts and those from different disciplines can be involved in the development of glo-

bal standards. As openEHR archetypes are routinely inspected by other standardization development organizations like HL7 to define message formats, this will also help to align our data model with newly emerging messaging standards as HL7 FHIR [30].

3.4 Terminologies

Besides the use of a reference model and domain models, terminologies (including ontologies) are the third pillar to achieve semantic interoperability [31]. They are useful to express unambiguous and computable statements about patients and health-related objects, to query clinical data repositories, to combine EHR data with external sources, and to connect structured EHR data with concepts obtained from free-text using natural language processing (NLP) methods [32]. They also help to express local differences in the levels of data granularity by defining site-specific value sets, which are incorporated within templates.

In HiGHmed, we strive to encode data using terminology standards, such as ICD-10, LOINC, and SNOMED CT, where applicable. Due to its comprehensiveness and its progressing development towards an ontology based on formal description logics, we regard SNOMED CT as the core reference terminology for HiGHmed. However, since SNOMED CT cannot cover all aspects of healthcare and already established terminologies have to be considered as well, a coordinated approach on a national and international level is required. Regarding the integration of terminologies within clinical information models, the

³ <https://www.openehr.org/ckm/>

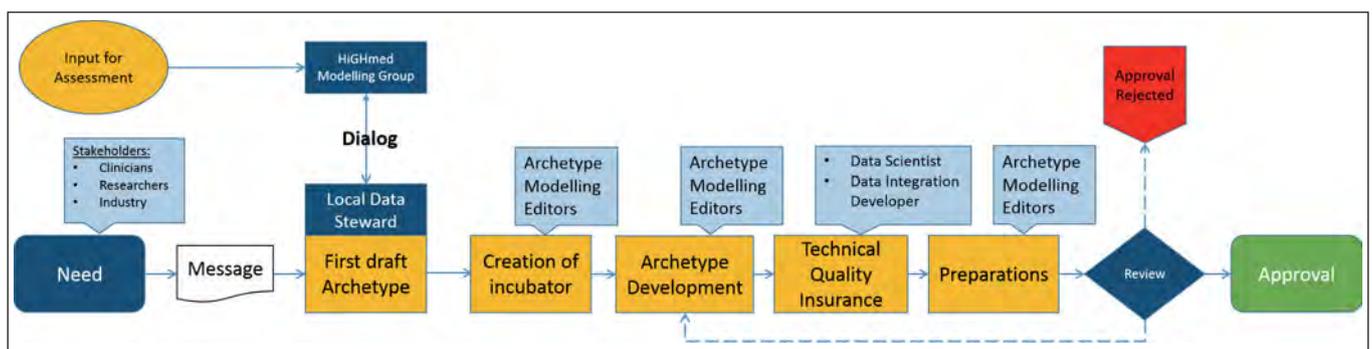


Figure 3 Governance process for archetype creation and approval.

HiGHmed Platform uses formally expressible binding mechanisms called *Terminology Bindings*, which are defined by the Archetype Definition Language. Thus, clinical information models and terminologies can complement each other to support data retrieval and querying, data capture and semantic interoperability between disparate sites [21, 33].

Regarding the development, administration, maintenance and distribution of terminologies between different sites, we will start with a setup that relies on the availability of a FHIR Terminology Service interface [34]. While this interface has been defined considering the Common Terminology Services 2 (CTS2) standard of the Object Management Group, it also offers flexibility to interface to other technologies. It provides a simple, yet powerful REST interface to search/query code systems and defined value sets as well as to conduct some basic management functions. To support the use of SNOMED CT, it provides functions to enable subsumption testing, testing for subset membership and subset expansions. Thereby, queries on concepts and value sets will be defined and combined with queries on the data level. If further features or harmonization with a national terminology service is needed at a later stage, the FHIR terminology service could be seamlessly integrated with CTS2 standard (as e.g. incorporated within the Austrian ELGA project [35]). In addition to standard terminologies, there will be a real-world need for HiGHmed-specific terminologies, code lists and value sets. These aspects will need to be coordinated within the Technical Coordination Board of HiGHmed in accordance with national standards.

3.5 Data Provision and Distribution Services

To optimally support health information exchange scenarios with a broad range of participants, it is required that each MeDIC implements an IHE XDS infrastructure including IHE Patient Identifier Cross-referencing (PIX) by providing an individual Master Patient Index (MPI) [36]. Access rights and patient consents require each MeDIC to provide an IHE Healthcare Pro-

vider Directory (HPD) and consent management based on IHE Advanced Patient Privacy Consents (APPC). Directed communication will be possible via IHE XDR, a profile for supporting peer-to-peer communication of medical content [37]. Since IHE XD* defines documents as content-agnostic containers, arbitrary content can be sent using IHE XDR. HiGHmed will use IHE XDR for responding to queries by providing one document per patient. The querying site implements a Document Consumer Actor, and the responding site implements a Document Source Actor. Large amounts of data are not supposed to be sent to the requestor, but instead the algorithms are sent to the data (e.g. omics pipelines in each MeDIC). This approach reduces the amount of data sent, and the requestor will only receive the analysis results.

To realize the workflows of cross-enterprise data federation within the consortium for the technical IHE-based solution, processes like anonymization, pseudonymization, consent management, use and access management (i.e. storing the user request, the integrated data and their analysis results) will be implemented. Not all processes are currently supported by IHE profiles. Hence, new profiles will be developed and ideally agreed upon nationally before proposing them to IHE International (either the IT Infrastructure (ITI) or Quality, Research and Public Health (QRPH) Domain).

3.6 Data Validation and Persistence

To load data into the XDS repository and, thereby, make them available to the HiGHmed Platform, each document is submitted using IHE ITI-41 (IT Infrastructure) messages. As the IHE XDS Registry Actor only accepts metadata for known patients, only data of patients that are already registered at the MPI can be loaded successfully. Each IHE ITI-41 message contains a metadata set that is necessary to create an entry in the XDS Registry. Each openEHR composition submitted to the openEHR repository is formally checked for compliance with the constraints defined in its associated template. This mechanism

is completely automated, meaning that all constraints set in the template will be checked for correctness on each data item. In case of a constraint violation, a report is created and sent back to the providing system, giving detailed information about the path of the data element and the type of failure. By using this validation approach, constraint checking does not have to be defined in the ETL-process, but is done automatically from the constraints that domain experts, researchers and clinicians have agreed upon beforehand. This approach facilitates the standardization of data validation in cross-enterprise environments. The Guideline Definition Language (GDL) [38] will complement these constraints by allowing the expression and processing of complex business rules directly referencing archetypes and templates. After successful validation, the data is finally stored in the clinical data repository. As the use of the openEHR Reference Model and templates allows for a fully generic persistence of data, there is no need to make changes to the underlying database. Consequently, newly added data can be queried immediately using the Archetype Query Language or be accessed using the REST interface.

3.7 Pipelining Data Analytics

The representation of data in a canonical and standard-based format not only allows for distributed querying and access via APIs, but facilitates transformations and provision of data in the format that is most suitable for particular data analytics and data mining applications. From the local data repository, all structured data can be (semi-)automatically transformed to a variety of formats. For example, the automated creation of tranSMART and i2b2 data marts is feasible [39]. This is a benefit of using archetypes, which allow to flexibly change clinical models while the transformation algorithms rely on the stable openEHR Reference Model. Thereby, access to a clearly defined and limited set of detailed patient data can be provided timely and without rewriting ETL-processes whenever new data definitions are added to the HiGHmed Platform. Analogously, arbitrary structured EHR data can be automatically transformed to the Web

Ontology Language (OWL). This allows for easy linkage of data with other sources that are readily available in OWL/RDF formats [40]. This also permits to use the full capability of ontologies of SNOMED CT for reasoning and semantic queries. Apart from these advanced approaches, automated data exports to relational database tables and interfaces to a range of analytics tools including R, Apache Spark, Apache Drill and Power BI will be supported eventually. The HiGHmed Platform will combine these transformations along with virtual images for the on-demand creation and provision of dedicated data marts for researchers at the local sites, HiGHmed and external partners alike. These data marts will also be stored in an archive.

3.8 Querying

By using openEHR and IHE XD*, we are able to establish data provision and access at different levels of detail. While IHE XD* offers a very high-level access to document metadata, openEHR enhances the provision services of IHE XDS by offering the possibility to create fine-grained data queries on structured data sets via the Archetype Query Language (AQL) [33].
 ▶ Figure 4 shows an exemplary AQL query. AQL directly references openEHR archetypes and terminologies to establish a semantically enabled, model-based approach to data querying. As AQL statements are defined at the clinical information model level, the execution of the queries is independent of the physical representation of the clinical data in the database layer. Thereby, queries and analytics scripts can be reliably shared between software systems from different vendors and dissimilar MeDICs, which nevertheless incorporate the same set of archetypes. For the HiGHmed Platform, this introduces a way to query data, and deploy algorithms and clinical decision support systems in a highly-distributed environment. By decoupling queries and algorithms from the physical representation of the data, it appears possible to migrate to high-performance databases (incorporating ‘big data’ technology such as in-memory databases, map reduce frameworks) without the need

to rewrite any queries or adjust connected decision support systems.

3.9 Application Programming Interfaces

The HiGHmed Platform offers an open and standardized REST API based on openEHR that is flexible enough to implement small apps as well as full-blown EHR systems [41]. Applications in HiGHmed can directly query the MeDIC data repositories or dedicated data marts for analytical queries along with an operational system to collect and manage data. As all underlying implementations of the openEHR standard share the same information models, a seamless exchange between all systems will be possible. To complement openEHR, we will use HL7 FHIR to provide further interfaces for application developers and data exchange [42]. FHIR will be implemented in combination with SMART interface, which has already led to a multitude of useful health applications that aim to access data in a standardized way [43].

3.10 Use and Access Management

▶ Figure 5 depicts a basic workflow for use and access management. Based on the shared metadata model, AQL queries are directly defined by the requestor via a drag and drop user interface. These queries are distributed to all relevant sites to provide counts of matching patients. Eventually, the requestor receives aggregated counts grouped by sites. Next, to access the record-level data for statistical analysis, the research project is proposed to all relevant institutional review boards (IRB). After approval, the requestor sends a data access

request via the local Data Transfer Unit (DTU) to the local Use & Access Committee (UAC). After a positive vote, the DTU uses the information from the data access request form to obtain a list of eligible patients. Afterwards, an IHE Document Consumer Actor uses this list along with optional constraints on document types and temporal constraints to query the IHE XDS Registry Actor. Based on the particular requestor, all access rules defined by the patients’ HiGHmed consent documents are checked based on IHE APPC. Then, using an IHE ITI-43 transaction (Retrieve Document Set), all applicable documents and data sets are retrieved by the Data Mart Manager and automatically loaded into an instance of the openEHR repository data mart which can be accessed by the requestor.

4. Use Cases

HiGHmed’s three prototypical clinical use cases were chosen deliberately, reflecting typical, frequently occurring, yet quite diverse research questions and care scenarios that require seamless data sharing and integration. The three use cases entail diverse challenges in terms of data protection and privacy, including clinical use of genome sequencing data (oncology), continuous longitudinal monitoring of physical activity (cardiology), and cross-site analysis of patient movement data (infection control). While the use cases will demonstrate the capabilities of the HiGHmed Platform to serve as foundation for diverse and complex clinical and research application systems, they are also designed to deliver immediate added value for clinical care and

```
SELECT
  e/ehr_status/subject/external_ref/id/value,
  diagnosis/data/items[at0002.1]/value FROM
  EHR e
  CONTAINS Composition c[openEHR-EHR-COMPOSITION.problem_list.v1]
  CONTAINS Evaluation diagnosis[openEHR-EHR-EVALUATION.problem-
  diagnosis.v1] WHERE
  c/name/value='Current Problems'
  AND diagnosis/data/items[at0002.1]/value/defining_code matches {
  terminology://SnomedCT/hierarchy?rootConceptId=50043002 }
```

Figure 4 This sample AQL query shows the capability of AQL to incorporate terminologies within queries: an URI directs to the SNOMED-CT terminology and additionally states an instruction to apply a “hierarchy” function call to retrieve all descendants of the given concept.

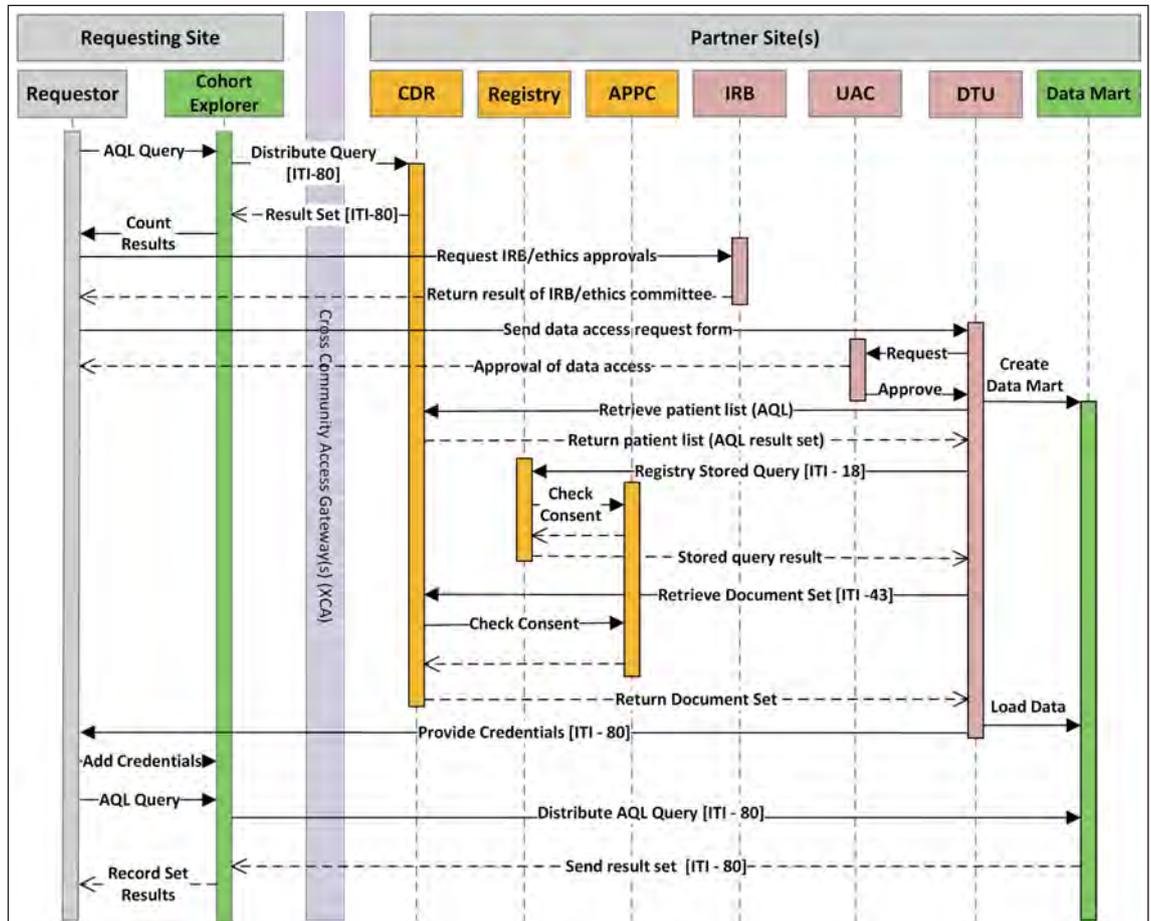


Figure 5
Basic technology-supported workflow of the use and access management.

all involved stakeholders. Patient participation is given highest priority, and, thus, each of the medical use cases is tightly intertwined with ethical research and considerations. Our use cases will provide practical case studies and training ground for the educational activities.

4.1 Oncology

Cancer diseases of the liver, pancreas and the bile duct system are collectively classified as carcinomas of the hepato-pancreato-biliary (HPB) tract. HPB cancers are among the most challenging malignancies that patients and oncologists face. The personalization of diagnosis and treatment of HPB cancer diseases is an emerging approach that requires highly complex methods such as multi-modal imaging, individualized radiotherapy planning, germline and cancer genome sequencing. Moreover, many recent scientific reports emphasize the power of molecular stratification and

imaging biomarkers, indicating that specific (germline and somatic) genomic, radiomic, radiogenomic, epigenetic, or gene expression alterations and pathway dysregulations not only define cancer development and progression, but also mediate therapeutic resistance [44]. However, to what extent individual molecular signatures correlate with certain patient characteristics (e.g. with clinical presentation at onset or with metastatic pattern), or with specific phenotypic features on radiological imaging or histology, is largely unknown [45].

For this reason, the overall aim of the use case oncology is to establish a functional infrastructure based on the HiGHmed Platform, which supports patients and clinical research as well as collaborative decision making by providing historical patient and study data from participating sites as well as external background knowledge to retrieve and provide

the most relevant information for each individual case at the decision point.

To meet these complex requirements, we will iteratively establish an integrated application system called virtual oncology center (VOC). The VOC will be built upon the APIs of the HiGHmed Platform and provide a technical and organizational infrastructure to be used for consultation, second opinion, tumor boards, informed treatment decisions, decision support, and knowledge discovery not only between the participating centers, but also with referring doctors and hospitals, and with patients. Besides features to support typical workflows to establish cross-enterprise tumor boards (request management, scheduling, preparation, meeting, finalization), the VOC will enable the iterative development and deployment of applications that will allow sharing and visualization of detailed information about a patient's course of treatment and that will help to identify similar cases and provide knowledge in

order to support collaborative decision making for selected patients. For data sharing and visualization, the VOC will use the capability of the HiGHmed Platform to store structured as well as unstructured data of patients in a coherent, longitudinal clinical data repository based on IHE XDS(-I) and openEHR. This includes clinical, pathologic, genetic and genomic, radiologic and radiotherapy data. While IHE XDS will provide a consistent description of document types and well-defined data access mechanisms, openEHR (combined with terminologies and nomenclatures including LOINC, UCUM, SNOMED CT, ICD, ICD-O and OPS) will enable meaningful processing and visualization of particular clinical observations and preconditions. This way, the course of treatment can be provided to all participants to support the preparation of a cross-organizational tumor board. Based on structured and harmonized data, similar patients will be searched for and identified across all sites. The data of similar patients can then be used to discuss treatment options and anticipated outcome. While such an approach will firstly need to rely on pre-defined and parameterized AQL queries providing an unsorted result set, semantic similarity measures based on machine learning and ontologies will be introduced in the mid-term to allow “querying by example” and to provide a meaningful ranking of similar patient profiles.

4.2 Cardiology

Heart failure still represents one of the most common reasons for hospitalization worldwide. Morbidity and mortality remain high: 5-year survival for patients with decompensated heart failure is 40 % [46] despite advances in drug and device therapy. Importantly, prognosis and quality of life of heart failure patients markedly worsens with every hospitalization due to deterioration of heart failure. Therefore, it is crucial to identify patients at high risk for heart failure hospitalizations to intensify or innovate current treatment strategies, hence avoid heart failure hospitalizations and improve prognosis as well as quality of life of these patients.

The use case aims to identify disease/data patterns in heart failure patients indicating an increased risk for future heart failure hospitalizations by integration of different data sources to tailor personalized treatment strategies improving prognosis, morbidity and quality of life. High quality and harmonization of data will be crucial to ensure valid results gained by data integration and development of algorithms for high-risk patient identification. As non-parametric, clinical patient data obtained from hospital information systems (e.g., patient history and clinical examination documented in patient records) regularly are of low or non-sufficient quality, first, a standardized set of validated and high quality clinical data is collected. This is an important requirement to ensure proper development of algorithms identifying high-risk patients, which have to be successfully applied also on non-validated, lower quality health care and patient data in the long term. Also, standardized and harmonized parametric analytical data (e.g. clinical chemistry, ECG, echocardiography) will be integrated. In addition, new sensor technologies will be applied to integrate longitudinal data such as heart rate, physical activity as well as subjective patient-related outcomes (mobile patient app). Harmonized and integrated data will be correlated to events of heart failure hospitalizations longitudinally collected by telephone questionnaires conducted by a heart failure nurse. Finally, the therapeutic value of disease/data patterns obtained by data integration indicating high-risk for heart failure hospitalizations will be prospectively validated in a clinical intervention trial.

The HiGHmed Platform will allow to analyze and later treat many more patients than each center could realize alone. With a standardized infrastructure, this may easily be extended to other centers. Thereby, it provides a solid foundation to move forward into personalized medicine. In addition, the quality and reliability of data capture will be substantially improved because of binding rules for all centers for the data sets. Furthermore, the cardiology departments in Heidelberg, Goettingen, and Hannover can link their own high quality cross-sectional and longitudinal datasets

(registries, cohorts, HIS, clinical trials) with the MeDIC data.

4.3 Infection Control

This use case was chosen to demonstrate the HiGHmed Platform’s capabilities for the detection of complex, multi-case interrelations in a data-rich, highly relevant environment with heterogeneous data sources. While infection surveillance is already in place in hospitals as well as on a national scale, it frequently suffers from a lack of data standardization and, consequently, from limited data integration, and limited and timely availability of relevant data. Actual infection outbreaks are rare events, stressing the need for hospitals to join forces and their data to uncover transmission mechanisms and evaluate prevention measures.

Multidrug-resistant organisms (MDRO) and their outbreaks are a major challenge nowadays for healthcare systems all over the world. Hospitals represent settings in which MDRO often occur in both a single patient and in a group of patients representing clusters and outbreaks. Healthcare-associated infections (HAI) are a major cause of excess patient morbidity and mortality, and cause significant added expenses for hospitals. The cumulative burden of the six most frequent HAIs was higher than the total burden of all other 32 communicable diseases in a recent study in the EU [47].

Failures in early recognition of transmissions and clusters result in large outbreaks that are difficult to contain. Clusters can remain undetected since transfers of patients between wards or institutions occur frequently, and microbiological detection of phenotypically similar pathogens at different spatiotemporal scales might not be considered as epidemiologically connected. Current human-based integration of information is error-prone and time consuming. The primary aim of this use case is to integrate all necessary pathogen-related data to establish a *smart infection control system* (SmICS). This will support interactive visualization of aggregated patient and pathogen movement data for the identification of crossing points, the exploration of transmissions, clusters and outbreaks and for deriving hypotheses about

new transmission pathways. In a second step, algorithmic detection of clusters and genuine outbreaks will be introduced. Our long-term aims include (1) setting up evidence for the adequate time point and mode of infection control measures, (2) supporting differentiation between colonization and infection, and (3) finally harmonizing infection control policies based on evidence-based standards.

SmiCS will be implemented as a HiGHmed Platform application on the basis of generic APIs, using shared semantic models for multiple clinical and organizational data sources, and therefore will be transferable across sites. With regard to new, non-standardized data sources, e.g., patient movement data or organizational data, which exist on different levels of detail at each of the partner sites, we expect particular benefits from HiGHmed's flexible, multi-level cross-enterprise modeling approach as well as from its scalability to facilitate roll-out. As a starting point, a minimal data set (MDS) has been defined, comprising of the following essential elements: basic patient and case data including demographics, diagnoses and procedures along with specific data relating to infectious diseases; microbiological, clinical chemistry and hematology lab data, e.g. microbiological culture results, including resistance patterns and molecular test results; movement data, including the allocation to a ward or a single room, bed occupancy.

5. Educational Concept

HiGHmed addresses the urgent needs to adapt the teaching systems in health and Medical Informatics to new didactic options based on digital learning and teaching formats. The teaching and training module of HiGHmed tries to address three major goals, all of which need new technical solutions to become operational in Germany:

1. Install interoperable teaching platforms for digital media in various curricula,
2. Develop a technical toolbox for the production of media for the platform, and
3. Demonstrate functional operation and evaluate acceptance.

To achieve these goals, a kernel of seven different curricula collaborates, reflecting the results of a two-year analysis regarding utilization of digital teaching formats in German curricula and how to speed up their use by the Federal Ministry of Education and Research. The group is open for additional partners.

Germany has a highly developed teaching system, which is strongly regulated. Therefore, it is a great challenge to adapt to the digital transformation of educational processes in science and economy despite many problems already published. One example is the curricular field, which brings together informatics, engineering, and health. Not only have the existing BSc and MSc curricula to be adapted – also new formats for lifelong learning and advanced training have to be set up for all levels of professional performance in medical informatics as well as other fields of health care or research. This process of change has three major dimensions: (1) develop new digital curricular formats, which can be used in different teaching and training scenarios, (2) update training methods, environments, and teachers based on current concepts of digitalization in higher education, (3) adapt the examination schemes and certification procedures accordingly.

The HiGHmed curricular project will build on existing e-learning platforms like the open source learning management systems ILIAS⁴ or Moodle⁵, which are operated in many universities and universities of applied sciences and arts in Germany. However, the local platforms are typically integrated into the operational teaching processes according to the local set ups. Apart from mapping modular learning content, external learning modules can also be integrated into the HiGHmed e-learning platform. This makes it possible to share learning modules across locations as well as to integrate innovative learning applications. Furthermore, e-learning platforms enable the implementation of various didactic learning concepts such as blended learning, instructional as well as self-directed or collaborative learning.

⁴ <https://www.ilias.de/>

⁵ <https://moodle.org/>

Following a competency-based approach, education and training offerings are designed to provide students in Medical Informatics and medicine as well as MI-professionals, physicians and other health professionals and clinical researchers with key competencies in data-management, analysis, and decision-support procedures in clinical care and clinical research.

Therefore, each participating curriculum prepares one teaching subject to be usable and evaluable at different other universities, addressing different aspects of the overall competency-based approach. To update the group on international developments and experience, annual workshops with leading methodologists will be arranged. To utilize industrial developments, a close collaboration has been started with industrial partners like Siemens Healthineers and Ada Health.

The evaluation process will be set up in a standardized way for all participating universities to allow transparency of project progress to international evaluators.

The educational project was prepared in several national meetings in 2017. It was decided to set up the following thematic framework as it addresses those aspects of health, where the digital transformation will push major changes and causes major needs for educational updates (responsible teaching groups named in brackets):

1. Medical Imaging Technologies and Data (HAWK-HHG)
2. Health-Enabling Technologies and Data (TU-BS)
3. Image- and Signal-Based Assistance Systems (HHN)
4. Advanced Concepts of Data Analytics and Curation (HSH)
5. Reliable Use of Data in Research and Care (MHH)
6. Citizen Centered Medical Information Management (UKL-HD)
7. Decision Support in Medical Care (UMG)

At the beginning of the main HiGHmed funding phase in 2018 six separate working streams have been started: (1) an internal skill update of international experiences in new digital teaching concepts, (2) setting up of a team of experts representing each site and working together, (3) developing

the technical basis for nationwide usable teaching units, (4) building up the specialized content in relation to targeted groups, (5) designing a set of workshops for decision makers of how they can make use of the new modules, and (6) adapting the regulatory system to the new formats developed.

It is expected that the project will raise major interest from many sides in the German teaching system. The years until 2020 will be used to become operational, so that evaluators can check innovation and readiness for everyday use and whether the above-mentioned aims were reached. The years after 2020 will focus on daily operations, new content, and quality management by user evaluation.

The ambitious project can only be successful if it is possible to map international theoretical and practical experience into the highly regulated German environment and to provide and maintain a digital education infrastructure, which allows to run digital Medical Informatics teaching units continuously for many years. If successful, the project will be a milestone in German educational development under the digital change [48, 49, 50].

6. Discussion

During the planning phase and, more recently, during the start of the implementation phase of HiGHmed, the anticipated need for a shared information governance between all sites has already been noticeable. Bringing together clinical stakeholders from several university hospitals revealed that data integration approaches can only be the starting point for long-term collaboration. For example, no technological or human intervention can realistically compensate missing or incomplete data that can occur due to variances in the treatment processes and the associated medical documentation. Another challenge is the tension between the need for structured and highly standardized data for reliable computability, analytics and decision support and the freedom of expression and efficiency of free-text documentation. While within HiGHmed, we will incorporate contemporary natural language processing

technologies, we also need to investigate advanced methods to better combine these two documentation approaches to equally address the needs of caregivers and researchers.

Moreover, while we seek to address the vast technological challenges of secondary use in the healthcare domain by specifying a scalable and open platform architecture, we consider the need to drive a cultural change as critical to achieve the ambitious goals of HiGHmed. For this reason, being able to build a strong coalition with selected clinical experts from all participating sites that are well-known as innovators and opinion leaders in their field, has already been regarded a key accomplishment. Senior clinical leaders have been involved early on in the definition of the use cases and the design of the HiGHmed Platform. Being able to provide real-world success stories will help to establish needed trust towards the MeDICs and the HiGHmed consortium. Meanwhile, these clinicians and researchers will also function as hubs to motivate peers to participate in forthcoming projects. This way, we intend to bridge potential chasms between early adopters and the majority group of users, demonstrating mutual benefits. The above measures will be accompanied by an introduction of HiGHmed concepts into curricula for medical students and medical information specialists alike, and into the continuing education of physicians and researchers in biomedical science.

By introducing an open platform approach to all participating hospitals and the DKFZ, HiGHmed pushes for a paradigm shift that provides means to center future clinical application systems around patients' data, and not vice versa. In so doing, the vast amount of patient data could finally be able to outlive the ephemeral software systems they were recorded with. Only the use of harmonized information models, syntax and terminologies right from the beginning of the data lifecycle will ensure semantic scalability and cost effectiveness in the long run. By introducing a shared information governance between disparate organizations and by introducing the capability to build comprehensive EHR applications grounded on a shared semantic layer, reuse of data will be highly fa-

cilitated. By supporting a broad set of emerging standards, including IHE XDS, openEHR and HL7 FHIR, and through the provision of an open system specification of the platform architecture, we seek to foster the establishment of an application ecosystem. In this regard, a clear commitment of the mandating hospitals will be needed to establish confidence on the industry side to develop applications based on the platform.

However, before a sufficient level of interoperability can be achieved 'by design', the HiGHmed Platform will need to deal with challenges of today's hospital information systems in terms of heterogeneity, data quality issues, consistency and more. Hence, the incorporation of data stewards is a bare necessity to ensure the enforcement of the FAIR principles and a sufficient level of data quality.

Furthermore, to reach its full potential, several legal and ethical challenges need to be addressed within HiGHmed and on a national level. For example, the definition of a national patient consent model is a highly complex task, which is beyond the scope of the HiGHmed project. Therefore, working with interim solutions will likely be needed to be able to include patients and, thereby, demonstrate the additional value of the HiGHmed Platform within the selected use cases. Another example for external dependencies is the establishment of SNOMED CT within HiGHmed and on a national level. Although the HiGHmed Platform is technically ready to leverage this terminology, a nationally coordinated effort is needed to provide localization and maintenance.

Within HiGHmed, we contemplated lessons learned from comparable projects. Collaborative infrastructures for increasing case numbers for certain diseases have been developed at National Institutes of Health (NIH) and within National Library of Medicine (NLM) funding activities for many years. As a result, tools like SHRINE [51, 52], i2b2 [53, 54] and tranSMART [55] have been implemented. The Shared Health Research Information Network (SHRINE) relies on a proprietary data model for mapping locally collected data to a head node, enabling all data to be kept locally (according to US HIPAA rules),

while at the same time making metadata accessible to external partners. However, data sharing on patient record-level is not yet commonly done. SHRINE participants compose their queries in i2b2 and receive aggregated results. tranSMART – a “cousin” of i2b2 sharing its data model – represents a modern data integration platform with the ability to flexibly include new analytics features for both high and low dimensional data via a plugin concept for R scripts. i2b2 and tranSMART fall short to address challenges regarding semantic interoperability and standardized interfaces to decision support systems. European initiatives include Electronic Health Records for Clinical Research (EHR4CR) [56], using i2b2 and standardized terminologies to leverage de-identified data from multiple sites to streamline feasibility assessment and recruitment of clinical trials. Finally, the SHARP consortium, which involved institutions such as Mayo Clinic, Intermountain Healthcare and MIT [57] investigated the establishment of large-scale health record data sharing including normalization using clinical information models (Clinical Element Models), natural language processing, and high-throughput phenotyping. This approach did not introduce the iterative development of an open platform that could be used to develop applications within the care provision based on a standard-based EHR architecture and shared semantics.

During the iterative development of the platform, we expect further challenges to occur. Questions regarding intellectual property, new (European) data privacy regulations (i.e. EU General Data Protection Regulation [58]), data sharing policies, the optimal inclusion of the patients and more will require an agile and explorative approach to develop a technical and organizational framework that is scalable on a national or even international level.

Acknowledgment

For the educational concept section, in addition to Otto Rienhoff, Petra Knaup-Gregori, Inga Kraus, and Michael Marschollek, the following persons have contributed: Marianne Behrends, Rolf Bendl, Oliver J. Bott, Thomas Deserno, Mark Hastenteufel,

Martin Hirsch, Volker Lang, Christoph Rußmann, Bernd Stock, Thomas Wetter, and Peter Wuebbel.

References

1. PricewaterhouseCoopers. Transforming Healthcare through Secondary Use of Health Data. 2009.
2. Rasmussen LV. The electronic health record for translational research. *J Cardiovasc Transl Res* 2014; 7(6): 607–614.
3. Sutherland SM, Kaelber DC, Downing NL, Goel VV, Longhurst CA. Electronic Health Record-Enabled Research in Children Using the Electronic Health Record for Clinical Discovery. *Pediatr Clin North Am* 2016; 63(2): 251–268.
4. Liao KP, Cai T, Gainer V, Goryachev S, Zengtreitler Q, Raychaudhuri S, et al. Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2010; 62(8): 1120–1127.
5. Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. *Am J Manag Care* 2007; 13(6 Part 1): 277–278.
6. Dentler K, ten Teije A, de Keizer N, Cornet R. Barriers to the reuse of routinely recorded clinical data: a field report. *Stud Health Technol Inform* 2013; 192: 313–317.
7. Chute CG, Beck SA, Fisk TB, Mohr DN. The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. *J Am Med Inform Assoc* 2010; 17(2): 131–135.
8. Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE – an integrated standards-based translational research informatics platform. *AMIA Annu Symp Proc* 2009; 2009: 391–395.
9. Danciu I, Cowan JD, Basford M, Wang X, Saip A, Osgood S, et al. Secondary use of clinical data: the Vanderbilt approach. *J Biomed Inform* 2014; 52: 28–35.
10. Tian Q, Price ND, Hood L. Systems cancer medicine: towards realization of predictive, preventive, personalized and participatory (P4) medicine. *J Intern Med* 2012; 271(2): 111–121.
11. Tanaka H. Omics-based medicine and systems pathology. A new perspective for personalized and predictive medicine. *Methods Inf Med* 2010; 49(2): 173–185.
12. Mirnezami R, Nicholson J, Darzi A. Preparing for precision medicine. *N Engl J Med* 2012; 366(6): 489–491.
13. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016; 3: 160018.
14. Schreiweis B, Bronsch T, Stein KE, Nost S, Aguduri LS, Brandner A, et al. An architecture for the integration of clinical data from a PEHR in a regional research platform. *Stud Health Technol Inform* 2016; 228: 272–276.
15. Brandner A, Schreiweis B, Aguduri LS, Bronsch T, Kunz A, Pensold P, et al. The patient portal of the personal cross-enterprise electronic health record (PEHR) in the Rhine-Neckar-Region. *Stud Health Technol Inform* 2016; 228: 157–161.
16. Koh HK, Brach C, Harris LM, Parchman ML. A proposed ‘health literate care model’ would constitute a systems approach to improving patients’ engagement in care. *Health Aff (Millwood)* 2013; 32(2): 357–367.
17. Beale T. Semantic scalability – the core challenge in e-health. 2014 [cited 2018 Jan 27]. Available from: <https://wolandscat.net/2014/12/15/semantic-scalability-the-core-challenge-in-e-health/>.
18. Beale T. What is an ‘open platform’. 2014 [cited 2018 Jan 27]. Available from: <https://wolandscat.net/2014/05/07/what-is-an-open-platform/>.
19. Stubbs A, Kotfila C, Uzuner Ö. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. *J Biomed Inform* 2015; 58 Suppl: S11–S19.
20. Baclawski K, Kilov H, Beale T. Archetypes: constraint-based domain models for future-proof information systems. In: Baclawski K, Kilov H, editors. Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer. Northeastern University, Evanston, IL, USA, 2002. p. 16–32.
21. openEHR Foundation. Archetype Definition Language ADL2. 2018 [cited 2018 May 9]. Available from: <http://www.openehr.org/releases/AM/latest/docs/ADL2/ADL2.html>.
22. Garde S, Knaup P, Hovenga E, Heard S. Towards semantic interoperability for electronic health records. *Methods Inf Med* 2007; 46(3): 332–343.
23. Kalra D, Lloyd D. ISO 13606 electronic health record communication part 1: reference model. ISO 13606–1. Geneva: ISO; 2008.
24. Archetype Modeling Language (AML). 2018 [cited 2018 May 9]. Available from: <http://www.omg.org/spec/AML/>.
25. Clinical Information Modeling Initiative (CIMI) [cited 2018 Jan 27]. Available from: <http://www.opencimi.org>.
26. Atalag K, Yang HY, Tempero E, Warren J. Model driven development of clinical information systems using openEHR. *Stud Health Technol Inform* 2011; 169: 849–853.
27. Ljosland Bakke S. National governance of archetypes in Norway. *Stud Health Technol Inform* 2015; 216: 1091.
28. Wulff A, Haarbrandt B, Marschollek M. Clinical Knowledge Governance Framework for Nationwide Data Infrastructure Projects. *Stud Health Technol Inform* 2018; 248: 196–203.
29. Leslie H, Ljosland Bakke S. Peer Review of Clinical Information Models: A Web 2.0 Crowdsourced Approach. *Stud Health Technol Inform* 2017; 245: 905–909.
30. Grieve G. FHIR and openEHR. (video file) [cited 2018 May 9]. Available from: <https://vimeo.com/191939835>.
31. Rector AL. The interface between information, terminology, and inference models. *Stud Health Technol Inform* 2001; 84(Pt 1): 246–250.
32. Lee D, de Keizer N, Lau F, Cornet R. Literature review of SNOMED CT use. *J Am Med Inform Assoc* 2014; 21(e1): e11–e19.
33. openEHR Foundation. Archetype Query Language (AQL). 2017 [cited 2018 May 9]. Available from: <https://www.openehr.org/releases/QUERY/latest/docs/AQL/AQL.html>.

34. HL7 International. FHIR Terminology Service [cited 2018 Jan 27]. Available from: <https://www.hl7.org/fhir/terminology-service.html>.
35. Seerainer C, Sabutsch SW. eHealth Terminology Management in Austria. *Stud Health Technol Inform* 2016; 228: 426–430.
36. Integrating the Healthcare Enterprise. IHE Cross-Enterprise Document Sharing [cited 2018 Jan 27]. Available from: https://wiki.ihe.net/index.php/Cross-Enterprise_Document_Sharing.
37. Integrating the Healthcare Enterprise. IHE Cross-enterprise Document Reliable Interchange [cited 2018 Jan 27]. Available from: https://wiki.ihe.net/index.php/Cross-enterprise_Document_Reliable_Interchange.
38. openEHR Foundation. Guideline Definition Language (GDL). 2015 [cited 2018 Jan 27]. Available from: <http://www.openehr.org/releases/CDS/la-test/docs/GDL/GDL.html>.
39. Haarbrandt B, Tute E, Marscholke M. Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository. *J Biomed Inform* 2016; 63: 277–294.
40. Haarbrandt B, Jack T, Marscholke M. Automated Transformation of openEHR Data Instances to OWL. *Stud Health Technol Inform* 2016; 223: 63–70.
41. HIMSS Analytics. Children's Hospital Ljubljana Achieves HIMSS Analytics EMRAM Stage 6. 2017 [cited 2018 May 9]. Available from: <http://www.himssinsights.eu/childrens-hospital-ljubljana-achieves-himss-analytics-emram-stage-6>.
42. HL7 International. Fast Healthcare Interoperability Resources (FHIR) Specification [cited 2018 Jan 27]. Available from: <http://hl7.org/implement/standards/fhir/>.
43. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inform Assoc* 2016; 23(5): 899–908.
44. Marquardt JU, Galle PR, Teufel A. Molecular diagnosis and therapy of hepatocellular carcinoma (HCC): an emerging field for advanced technologies. *J Hepatol* 2012; 56(1): 267–275.
45. Acharya UR, Hagiwara Y, Sudarshan VK, Chan WY, Ng KH. Towards precision medicine: from quantitative imaging to radiomics. *J Zhejiang Univ Sci B* 2018; 19(1): 6–24.
46. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. *N Engl J Med* 2006; 355(3): 251–259.
47. Cassini A, Plachouras D, Eckmanns T, Abu Sin M, Blank HP, Ducomble T, et al. Burden of Six Healthcare-Associated Infections on European Population Health: Estimating Incidence-Based Disability-Adjusted Life Years through a Population Prevalence-Based Modelling Study. *PLoS Med* 2016; 13(10): e1002150.
48. Bundesministerium für Bildung und Forschung. Digitale Innovationen – Neue Dimensionen von Bildung und Wissenschaft erschließen. 2017 [cited 2018 May 7]. Available from: https://www.bmbf.de/pub/Digitale_Innovationen.pdf. German.
49. Hochschulforum Digitalisierung. The Digital Turn – Hochschulbildung im digitalen Zeitalter. 2016 [cited 2018 May 7]. Available from: <https://hochschulforumdigitalisierung.de/sites/default/files/dateien/Abschlussbericht.pdf>. German.
50. Scheer A-W, Wachter C. Digitale Bildungslandschaften. Saarbrücken: imc information multimedia communication AG; 2016.
51. Weber GM, Murphy SN, McMurry AJ, Macfadden D, Nigrin DJ, Churchill S, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. *J Am Med Inform Assoc* 2009; 16(5): 624–630.
52. McMurry AJ, Murphy SN, MacFadden D, Weber G, Simons WW, Orecchia J, et al. SHRINE: enabling nationally scalable multi-site disease studies. *PLoS One* 2013; 8(3): e55811.
53. Natter MD, Quan J, Ortiz DM, Bousvaros A, Ilowitz NT, Inman CJ, et al. An i2b2-based, generalizable, open source, self-scaling chronic disease registry. *J Am Med Inform Assoc* 2013; 20(1): 172–179.
54. Segagni D, Tibollo V, Dagliati A, Malovini A, Zambelli A, Napolitano C, et al. Clinical and research data integration: the i2b2-FSM experience. *AMIA Jt Summits Transl Sci Proc* 2013; 2013: 239–240. eCollection 2013.
55. Athey BD, Braxenthaler M, Haas M, Guo Y. transSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research. *AMIA Jt Summits Transl Sci Proc* 2013; 2013: 6–8. eCollection 2013.
56. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: the case of the EHR4CR project. *J Biomed Inform* 2015; 53: 162–173.
57. Chute CG, Pathak J, Savova GK, Bailey KR, Schor MI, Hart LA, et al. The SHARPN project on secondary use of Electronic Medical Record data: progress, plans, and possibilities. *AMIA Annu Symp Proc* 2011; 2011: 248–256.
58. The European Parliament, The European Council. General Data Protection Regulation. *Off J Eur Union* 2016; 2014(March 2014): 20–30.