Smart Medical Information Technology for Healthcare (SMITH)*

Data Integration based on Interoperability Standards

Alfred Winter[1]; Sebastian Stäubert[1]; Danny Ammon[2]; Stephan Aiche[3]; Oya Beyan[4]; Verena Bischoff[5]; Philipp Daumke[6]; Stefan Decker[4]; Gert Funkat[7]; Jan E. Gewehr[8]; Armin de Greiff[9]; Silke Haferkamp[10]; Udo Hahn[11]; Andreas Henkel[2]; Toralf Kirsten[12]; Thomas Klöss[13]; Jörg Lippert[14]; Matthias Löbe[1]; Volker Lowitsch[10]; Oliver Maassen[15]; Jens Maschmann[16]; Sven Meister[17]; Rafael Mikolajczyk[18]; Matthias Nüchter[12]; Mathias W. Pletz[19]; Erhard Rahm[20]; Morris Riedel[21]; Kutaiba Saleh[2]; Andreas Schuppert[22]; Stefan Smers[7]; André Stollenwerk[23]; Stefan Uhlig[24]; Thomas Wendt[25]; Sven Zenker[26]; Wolfgang Fleig[27,]; Gernot Marx[15,**]; André Scherag[28, 29,**]; Markus Löffler[1,**]**

[1]Leipzig University, Institute of Medical Informatics, Statistics and Epidemiology, Leipzig, Germany;
[2]University Medical Center Jena, Central Service Provider For Information Technology, Jena, Germany;
[3]SAP SE, Potsdam, Germany;
[4]RWTH Aachen University, Chair of Computer Science 5, Aachen, Germany;
[5]University of Leipzig Medical Center, Division Staff and Justice, Leipzig, Germany;
[6]Averbis GmbH, Freiburg, Germany;
[7]University of Leipzig Medical Center, Division Information Management, Leipzig, Germany;
[8]University Medical Center Hamburg-Eppendorf, Business Division for Information Technology, Hamburg, Germany;
[9]Essen University Hospital, Central Information Technology, Essen, Germany;
[10]RWTH Aachen University Hospital, Division Information Technology, Aachen, Germany;
[11]Friedrich-Schiller-Universität Jena, Language & Information Engineering Lab (JULIE Lab), Jena, Germany;
[12]Leipzig University, LIFE Research Centre for Civilization Diseases, Leipzig, Germany;
[13]Martin-Luther-Universität Halle-Wittenberg Medical Center, Medical Director, Halle, Germany;
[14]Bayer AG, Wuppertal, Germany;
[15]RWTH Aachen University Hospital, Department of Intensive Care and Intermediate Care, Aachen, Germany;
[16]University Medical Center Jena, Medical Director, Jena, Germany;
[17]Fraunhofer Institute for Software and Systems Engineering, Dortmund, Germany;
[18]Martin-Luther-Universität Halle-Wittenberg, Institute of Medical Epidemiology, Biometry and Informatics, Halle, Germany;
[19]University Medical Center Jena, Institute of Infectious Diseases and Infection Control, Jena, Germany;
[20]Leipzig University, Department of Computer Science – Database Group, Leipzig, Germany;
[21]Forschungszentrum Jülich, Jülich Supercomputing Centre, Jülich, Germany;
[22]RWTH Aachen University, Institute for Computational Biomedicine II, Aachen, Germany;
[23]RWTH Aachen University, Informatik 11 – Embedded Software, Aachen, Germany;
[24]RWTH Aachen University, Medical Faculty, Dean, Aachen, Germany;
[25]University of Leipzig Medical Center, Data Integration Center, Leipzig, Germany;
[26]University of Bonn Medical Center, Department of Anesthesiology and Intensive Care Medicine, Bonn, Germany;
[27]University of Leipzig Medical Center, Medical Director, Leipzig, Germany;
[28]University Medical Center Jena, Center for Sepsis Control and Care, Jena, Germany;
[29]University Medical Center Jena, Institute of Medical Statistics, Computer and Data Sciences (IMSID), Jena, Germany

Correspondence to:
Prof. Alfred Winter
Leipzig University
Institute of Medical Informatics, Statistics
and Epidemiology
Haertelstr. 16-18
04107 Leipzig
Germany
E-mail: alfred.winter@imise.uni-leipzig.de

Methods Inf Med 2018; 57(Open 1): e92–e105
https://doi.org/10.3414/ME18-02-0004
received: March 5, 2018
accepted: May 7, 2018

Funding
This work has been supported by German Federal Ministry of Education and Research (Grant No's. 01ZZ1609A, 01ZZ1609B, 01ZZ1609C, 01ZZ1803A, 01ZZ1803B, 01ZZ1803C, 01ZZ1803D, 01ZZ1803E, 01ZZ1803F, 01ZZ1803G, 01ZZ1803H, 01ZZ1803I, 01ZZ1803J, 01ZZ1803K, 01ZZ1803L, 01ZZ1803M, 01ZZ1803N).

Keywords
Health information interoperability, integrated advanced information management systems, systems integration, phenotyping, intensive care, antimicrobial stewardship

Summary

Introduction: This article is part of the Focus Theme of Methods of Information in Medicine on the German Medical Informatics Initiative. "Smart Medical Information Technology for Healthcare (SMITH)" is one

* Supplementary material published on our website https://doi.org/10.3414/ME18-02-0004
** Shared senior authorship

of four consortia funded by the German Medical Informatics Initiative (MI-I) to create an alliance of universities, university hospitals, research institutions and IT companies. SMITH's goals are to establish Data Integration Centers (DICs) at each SMITH partner hospital and to implement use cases which demonstrate the usefulness of the approach.

Objectives: To give insight into architectural design issues underlying SMITH data integration and to introduce the use cases to be implemented.

Governance and Policies: SMITH implements a federated approach as well for its governance structure as for its information system architecture. SMITH has designed a generic concept for its data integration centers. They share identical services and functionalities to take best advantage of the interoperability architectures and of the data use and access process planned. The DICs provide access to the local hospitals' Electronic Medical Records (EMR). This is based on data trustee and privacy management services. DIC staff will curate and amend EMR data in the Health Data Storage.

Methodology and Architectural Framework: To share medical and research data, SMITH's information system is based on communication and storage standards. We use the Reference Model of the Open Archival Information System and will consistently implement profiles of Integrating the Health Care Enterprise (IHE) and Health Level Seven (HL7) standards. Standard terminologies will be applied. The SMITH Market Place will be used for devising agreements on data access and distribution. 3LGM^2 for enterprise architecture modeling supports a consistent development process.

The DIC reference architecture determines the services, applications and the standards-based communication links needed for efficiently supporting the ingesting, data nourishing, trustee, privacy management and data transfer tasks of the SMITH DICs. The reference architecture is adopted at the local sites. Data sharing services and the market place enable interoperability.

Use Cases: The methodological use case "Phenotype Pipeline" (PheP) constructs algorithms for annotations and analyses of patient-related phenotypes according to classification rules or statistical models based on structured data. Unstructured textual data will be subject to natural language processing to permit integration into the phenotyping algorithms. The clinical use case "Algorithmic Surveillance of ICU Patients" (ASIC) focusses on patients in Intensive Care Units (ICU) with the acute respiratory distress syndrome (ARDS). A model-based decision-support system will give advice for mechanical ventilation. The clinical use case HELP develops a "hospital-wide electronic medical record-based computerized decision support system to improve outcomes of patients with blood-stream infections" (HELP). ASIC and HELP use the PheP. The clinical benefit of the use cases ASIC and HELP will be demonstrated in a change of care clinical trial based on a step wedge design.

Discussion: SMITH's strength is the modular, reusable IT architecture based on interoperability standards, the integration of the hospitals' information management departments and the public-private partnership. The project aims at sustainability beyond the first 4-year funding period.

1. Introduction and Objectives

"Using the wealth of data – for improved patient care" [1] is the motto motivating the German Federal Ministry of Education and Research for funding the projects in the German Medical Informatics Initiative (MI-I) and introduced in this special issue. Better usage of the wealth of data can contribute to validating the relevance of novel diagnostic and therapeutic methods as well as to directly improving care. Patient care would benefit from the availability of actionable, data-driven decision support standardized across care delivering organizations, e.g. to optimize diagnosis and therapy in intensive care patients with lung failure or antibiotic therapy of patients with blood-stream infections.

Taking up the initiative of the German Federal Ministry of Education and Research, Leipzig University and University Hospital Leipzig, Jena University Hospital and Friedrich-Schiller-University Jena, University Hospital RWTH Aachen and RWTH Aachen University founded the "Smart Medical Information Technology for Healthcare (SMITH)" consortium in order to better use the wealth of data for the sake of patients. In addition, the following research organizations and companies joined the initial SMITH set-up: SAP SE, Fraunhofer Institute for Software and Systems Engineering, Bayer AG, März Internetwork Services AG, Averbis GmbH, ID GmbH & Co. KGaA, Forschungszentrum Jülich – Jülich Supercomputing Centre. In a second consolidation phase, University Hospitals Halle, Bonn, Hamburg and Essen completed the SMITH consortium.

The SMITH consortium has particularly set the following goals to be reached until 2022:

- to implement an overarching concept of data sharing and data ownership;
- to establish synchronized Data Integration Centers (DIC) at each SMITH partner hospital. They will have two major responsibilities: (1) implement the interoperability processes and the architecture of the SMITH transinstitutional information system SMItHIS; (2) provide data curation, data sharing and trustee functions;
- to implement three use cases in order to demonstrate the usefulness of the DICs.

This paper's objective is to give insight into fundamental design issues underlying SMITH. It will especially introduce the governance of SMITH and its basic policies, explain the overall SMItHIS architecture to be implemented and the methods used, and will give more details of the use cases to be implemented in order to show the feasibility of the architecture.

2. Governance and Policies

The SMITH consortium's governance structure as well as the SMItHIS architecture results from a strictly federated approach. Instead of one central DIC, local

DICs are coordinated by appropriate committees. These committees aim at integrating the interests and activities of patient care, i.e. the member hospitals, of research, i.e. the member medical faculties and non-university research institutions, and of the industrial partners. Therefore, leadership as well as means for participation are carefully balanced.

In this governance system, the Supervisory Board bears responsibility for the SMITH consortium and the overall project. This board nominates the delegates for the National Steering Committee of the MI-I, which coordinates the activities of all funded consortia in the MI-I. The Supervisory Board appoints the Executive Board.

The Executive Board bears the scientific responsibility for SMITH and steers the project. It is supported by the External Advisory Board whose members are selected international experts in the field.

The Coordination Board represents delegates from all sub-projects and from all partners and coordinates their work.

Besides local management units at each partner's site, there is a consortium management unit for SMITH located in Leipzig. The consortium management unit takes care of operations of SMITH as a whole and is supported by the local management units. Operations include the entire project management, controlling, contracting, public relations and communication.

Based on the governance structure, functionally identical DICs will be set up in Leipzig, Jena, Aachen and, subsequently, in Halle, Bonn, Hamburg and Essen. DIC staff will have the same job descriptions and will use common standard operating procedures. The DICs will adhere to the following policies:

- Local access to the HIS: The DICs have to provide access to the data of the Electronic Medical Records (EMR) in the local hospital information systems (HIS). This access is subject to German legislation on the use of patient data and established data protection and privacy regulations. Each DIC will analyze and annotate patient data on an individual basis within the hospital. Authorized local DIC staff will have access to the EMR in the locally operational HIS system. This requires the DIC to be organizationally linked to the hospital and technically interoperable with its information system and with the clinical documentation procedures (cf. Data and Metadata Transfer Management (DMT) in section 3.2.1).
- Trustee and Consent for Research: Germany has strict regulations concerning data protection and privacy. Research on patient data (with local amendments, annotations etc.) is only possible if patient consent is provided. We anticipate that multilevel consent will be provided. This will range from consenting to data being shared regarding anonymous or pseudonymous healthcare data to additional consent for specific research projects. The DIC provides the relevant services for the trustee center (cf. Data Trustee and Privacy Management (DTP) in section 3.2.1).
- Health Data Storage (HDS) to provide electronic health records (EHR): The local HISs typically exhibit different IT architectures and consist of numerous specialized application components, which are also currently being refined. However, functionally identical Health Data Storages will be established in the participating institutions to maintain standardized and interoperable EMRs curated and amended by the DIC staff (cf. SMItHIS architecture in section 3.2.2.1).
- Organization: Overall management policies will be identical across the DICs. However, we will have local sub-policies on organizational embedding.

In order to further strengthen Medical Informatics, education in Health and Biomedical Informatics at different levels will be driven by systematic, evidence-based and outcome-oriented curriculum development. This process is concerted by the Joint Expertise Center for Teaching (SMITH-JET) as part of the SMITH governance structure.

3. Architectural Framework and Methodology

3.1 Methodology

3.1.1 Open Archival Information System (OAIS)

Our approach to data integration centers and their organizational structure as well as their tasks is inspired by the Reference Model for an Open Archival Information System (OAIS) [2]. This model provides a framework, including terminology and concepts, for describing and comparing architectures and operations of archives and thus for sharing their content. OAIS is the most common standard for archival organizations (ISO Standard 14721:2012). OAIS conformant archives have to take care of proper ingest of Submission Information Packages ("raw data") from producers and their transformation to Archival Information Packages ("metadata-enriched data"), which are processed by Data Management and Archival Storage. Consumers order or query Dissemination Information Packages ("exported data") which are processed by the archive. OAIS is helpful in structuring tasks within the DIC without prescribing implementation details. It is already used as guideline for data and model sharing at the LIFE Research Center for Civilization Diseases [3, 4, 5] and the Leipzig Health Atlas project [6].

3.1.2 Three Layer Graph Based Meta Model (3LGM²) for Enterprise Architecture Modeling

In order to provide an integrated and consistent view of the design of the entire system of data processing and data exchange within SMITH and beyond, we decided to use an enterprise architecture modeling approach. By using the three layer graph based meta model 3LGM² for modeling health information systems [7], especially transinstitutional information systems [8] and, therefore, the entire information system of SMITH with its local institutional components can be described by concepts on three layers [9]:

- The *domain layer* describes an information system independent of its implementation by the tasks it supports

(e.g. "Ingesting Data and Knowledge in DIC Health Data Storage"). Tasks need certain data and provide data for other processes. In 3LGM² models, these data are represented as entity types.
- The *logical tool layer* focusses on *application components* supporting tasks (e.g. "Health Data Storage (HDS)", "Data Integration Engine"). Application components are responsible for the processing, storage, and transfer of data. Computer-based application components are installed software. *Interfaces* ensure the communication among application components and, therefore, enable their integration.
- The *physical tool layer* consists of physical data processing systems (e.g. personal computers, servers, switches, routers, etc.), which are connected in a network.

This approach for structuring and describing information systems turned out to be an appropriate basis for not only describing health information systems at the interface between patient care and research [10] but also for assessing their quality [11].

3LGM² model data for both the domain and the logical tool layer have been collected at several workshops and interviews from all partners of the SMITH consortium. In the first step, we collected tasks and entity types (i.e. the data) needed or produced by the tasks in workshops with experts in the domain of data sharing in and between health care and medical research. In the second step, the workshops were held with the CIOs and their experts of health information systems architectures and standards-based communication. The application components found and their interfaces and the communication links between them were connected to the tasks and to the entity types determined by the domain experts. Thus, we developed a blueprint for the transinstitutional SMITH information system, which integrates the domain layer view of tasks and entity types used with a tool layer view of applications, services and communication. This was the basis for the SMITH DIC reference architecture design explained in more depth in 3.2.2.1. The continuously updated 3LGM² model will be used during the entire project as an evolving blueprint for the development process.

3.1.3 Integrating the Healthcare Enterprise (IHE)

Our goal is to share medical and research data not only inside the SMITH consortium but also with a stepwise growing number of partners in the near future. Therefore, we decided to design the logical tool layer of SMITH's information system strictly based on communication and storage standards as far as possible. Especially, we decided to implement IHE integration profiles, which describe the precise and "coordinated implementation of communication standards, such as DICOM, HL7 W3C and security standards" [12] (DICOM: Digital Imaging and Communications in Medicine; W3C: World Wide Web Consortium). IHE profiles ensure *processual interoperability*, since they define how application systems in a certain role (described as actor, e.g. "document consumer", "document provider", "document repository") can communicate with other application systems through certain transactions.

In particular, the following IHE integration profiles have been used for designing the SMITH DIC reference architecture [12]:
- *Patient Identifier Cross-referencing* (PIX) profile for pseudonym management across all SMITH sites.
- *Patient Demographics Query* (PDQ) profile for querying demographic data from any patient management system of a SMITH partner hospital.
- *Cross-Enterprise Document Sharing* (XDS) profile for sharing documents within an affinity domain, which usually covers (parts of) a SMITH partner hospital.
- *Cross-Community Access* (XCA) profile for sharing documents across affinity domains, e.g. between SMITH partner hospitals.
- *Basic Patient Privacy Consents* (BPPC) and *Advanced Patient Privacy Consents* (APPC) profiles for recording patients' privacy consents in a way that they can be used for controlling access to documents via XDS and XCA.
- The *Cross-Enterprise User Assertion* (XUA) profile for checking and confirming the identity of persons or systems trying to access data.
- The *Audit Trail and Node Authentication* (ATNA) profile for providing data integrity and confidentiality even in case local interim copies of confidential data have to be allowed for safety reasons.
- The *Cross-Enterprise Document Workflow* (XDW) profile to define and use workflows upon document sharing profiles like XDS.
- The *Cross-Community Patient Discovery* (XCPD) profile to find patient data across the SMITH partner hospitals.
- The *Cross-Community Fetch* (XCF) profile: like XCA, this profile accounts for sharing documents, but in a simplified way. In SMITH, we also use this profile in a modified way to transport FHIR queries and resources.
- The *Personnel White Pages* (PWP) integration profile for managing user contact information.
- The *Healthcare Provider Directory* (HPD) profile for managing healthcare provider information, including roles and access rights.

These profiles have been adapted to the planned architecture at the logical tool layer of the SMITH information system. Using 3LGM², we mapped IHE profile actors to the planned application components and used the transactions as templates for defining the applications' communication interfaces. This endeavor was supported by predefined templates and corresponding workflows which have been described in more detail in [13].

3.1.4 Communication Standards HL7 CDA and HL7 FHIR

IHE profiles, especially XDS und XCA, support the shared use of medical documents. For this purpose, the HL7 Clinical Document Architecture (CDA) provides specifications for various types of documents typically used in clinical care. In this XML-based format, the CDA level describes a degree of structuring in the XML body, ranging from level 1 (containing only textual information) via level 2 (coded sections) to level 3 (structured entries).

Since we do not only want to share documents but discrete individual data as well, we took advantage of recent developments of IHE profiles using Fast Healthcare Interoperability Resources (FHIR) [14]. FHIR defined "resources" can be used to arrange patient data, e.g. about allergies and intolerances, family member histories, medication requests and provide them for access by remote application systems. Modern web-based application programming interface (API) technology, especially the RESTful protocol, can be used to access this data using certain operations, which are defined as "interactions" by FHIR.

Since entries contained in CDA documents can be linked to FHIR resources and several FHIR resources together can be combined in a document, both formats are suitable to facilitate *syntactic interoperability* for retrieval and exchange of clinical care data.

3.1.5 Medical Terminologies

For a transinstitutional use of medical and research data, these data must include or refer to machine-processable descriptions of their content. International medical terminologies like SNOMED CT offer code systems and define relations between subject entities for an unambiguous specification of certain medical information. We will set up terminology services to provide detailed representations of conceptual entities. This includes browsing terminologies, displaying concepts, facetted searching in terminologies and exporting. We will import standard terminologies for coded medical data like the International Classification of Diseases (ICD) and the Logical Observation Identifiers Names and Codes (LOINC); terminologies and ontologies used for text mining and phenotyping, like Medical Subject Headings (MeSH), ICD, LOINC or Human Phenotype Ontology (HPO); furthermore, local vocabularies will be used in a DIC (e.g. for medication). In addition, new vocabularies and concepts can be created and mapped to standardized terminologies to facilitate the use of core data sets. These terminology services will be based on the Common Terminology Services 2 (CTS2) data model.

Both HL7 CDA and HL7 FHIR refer to the use of medical terminologies for the encoding of clinical concepts, thus enabling *semantic interoperability* of shared data.

A link is created from process-related, technical and semantic interoperability with IHE, HL7 (procedure description and definition), with HL7 CDA, FHIR, Clinical Quality Language (CQL), etc. (definition of protocols and file formats) and LOINC, SNOMED-CT, etc. (Information Models).

3.1.6 Industrial Data Space (IDS)

SMITH wants to go beyond mere data aggregation and therefore will develop the SMITH Market Place (SMP) for devising agreements on data access and distribution (see 4.2.2). The architecture of the SMP is based on the Industrial Data Space (IDS) project, led by Fraunhofer ISST. IDS is a virtual data space using standards and common governance models to facilitate the secure exchange and easy linkage of data in business ecosystems. It thereby provides the basis for creating and using smart services and innovative business processes, while at the same time ensuring digital sovereignty of data owners, decentral data management, data economy, value creation, easy linkage of data, trust and secure data supply chains and, finally, data governance [15].

To facilitate the above-mentioned characteristics, three key architectural components were defined within the IDS. The *connector* allows the secure interconnection with data sources and execution of apps. To mediate between different connectors, the *broker* allows the semantical description of data sources as well as apps. The last component is the *app store,* which manages certified apps. With respect to the SMP, the already defined concepts should be used to enable an all-encompassing data market place for medical informatics, paying attention to specific requirements like regulatory affairs (e.g. consenting, data protection and privacy) and existing data exchange and integration standards (see 3.1.3 and 3.1.4).

The concepts of the connector as well as the broker, will be reflected in the conceptualization process of the SMP. The connector will be a secure and trusted software-based endpoint for external users to enable secure handshaking and contracting between external participants and the SMP. Furthermore, the connector has to be used to initiate the onboarding process. The broker will serve as an interface to the metadata repository for data provisioning and services. Like a directory, the broker will allow external as well as internal users to query a metadata repository in order to fulfill a customer-specific demand.

3.2 Architectural Framework of SMItHIS

SMITH intends to build a medical network of its partner hospitals, medical faculties and research institutions. This network is based on SMItHIS, the network's transinstitutional Health Information System (tHIS) [8, 16]. SMItHIS connects the information systems of the partner institutions, including their Data Integration Centers (DIC) and local clinical and research application components. SMItHIS integrates them by added components. In this section, we introduce the architectural framework for SMItHIS, which will guide the entire development process.

In order to provide a holistic and consistent view of the design of the entire system of data processing and data exchange within SMITH and beyond, we decided to use the Enterprise Architecture Modeling approach 3LGM². Using 3LGM², the entire transinstitutional information system SMItHIS with its local institutional components can be described by concepts on three layers. Since there are no special solutions on the physical tool layer, we here focus on the domain and logical tool layers.

SMITH has designed a generic concept for its data integration centers. They share identical services and functionalities to take best advantage of the interoperability architectures and of the data use and access process planned. In the following section 3.2.1, we describe the main tasks of a DIC in SMITH; they are common for the DICs of all partners. In section 3.2.2 we describe the application components and their interfaces and communication links. We will do this first by a generic reference model

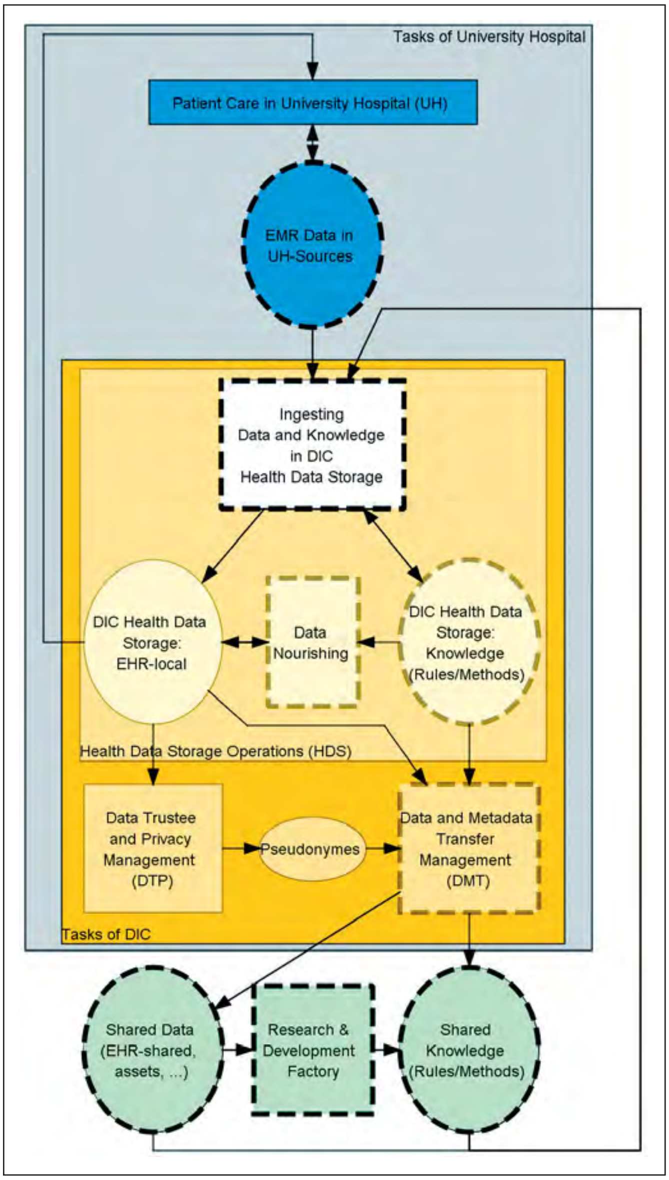
Tasks of University Hospital
Patient Care in University Hospital (UH)
EMR Data in UH-Sources
Ingesting Data and Knowledge in DIC Health Data Storage
DIC Health Data Storage: EHR-local
Data Nourishing
DIC Health Data Storage: Knowledge (Rules/Methods)
Health Data Storage Operations (HDS)
Data Trustee and Privacy Management (DTP)
Pseudonymes
Data and Metadata Transfer Management (DMT)
Tasks of DIC
Shared Data (EHR-shared, assets, ...)
Research & Development Factory
Shared Knowledge (Rules/Methods)

Figure 1 High-level tasks (rectangles) and entity types (ovals) of SMItHIS. Arrows pointing from a task T to an entity type (ET): ET is updated by T, other direction: ET is used by T. Dotted lines indicate availability of more detailed refinements. Rectangles inside other rectangles indicate part-of relations [1].

and show later how we will assemble and integrate the local instances of the reference model.

3.2.1 SMItHIS Domain Layer: Tasks of DIC

The domain layer describes which tasks have to be performed in SMITH and what data are used for these tasks. ►Figure 1 gives a high-level overview of the tasks and entity types (data) relevant for SMItHIS from the DIC perspective (dotted lines indicate availability of more detailed refinements that can be found in the figure of the ►Online Appendix).

Each DIC operates a Health Data Storage (HDS) containing a local EHR (electronic health record) [17, 18] covering both data in the local EMR (electronic medical record) [18] ingested from a local partner University Hospital (UH) and data ingested from other sources. In addition, the HDS contains knowledge, i.e. rules and methods on how to nourish the ingested data. *Health Data Storage Operations* in DIC cover ingesting data and knowledge, as well as data nourishing.

- *Ingesting Data and Knowledge*: Considerable parts of local EMR data are still unstructured data in reports, findings or discharge letters. However, structured, i.e. discrete individual data is needed. Taking unstructured patient data under DIC supervision calls for incorporating natural language processing (NLP) tools. Text-mining algorithms extract information from narrative texts and render it as structured data [19]. The needed algorithms will be developed within the so-called Research & Development Factory. By this term, we summarize the research and development projects using the shared data and deriving rules and methods, i.e. knowledge, out of the data. Within the factory, the methodological use case "Phenotype pipeline" (PheP) will e.g. develop phenotyping rules (see 4.1). A DIC will as well be able to ingest data from insurance companies (record linkage) and from patients themselves (e.g. patient reported outcomes) by providing specific services in a Patient Portal. Ingested patient data from EMR in a certain university hospital and from shared resources are stored as a local EHR in the HDS. Knowledge will be ingested to the local HDS as well. Note that ingesting data into the HDS does not necessarily mean that data has to be copied. Rather, links to the data in their sources may be used.
- *Data Nourishing* deals with adding value to the ingested patient data. Ingesting comprises metadata management, data curation and phenotyping.
  Metadata management provides and maintains a local metadata repository, which describes data elements and their semantics. It will conform to ISO/IEC norm 11179, which is widely used in the health care domain. As part of data and metadata transfer management, subsets of this catalog may be shared. Additionally, semantic information out of the local catalog of metadata enrich patient data in the local EHR.
  Curation and applying the phenotyping rules to data in the local EHR is supported by a Rules-Engine executing rules, which are taken from the previously ingested knowledge. The DIC is responsible for proper operation of the Rules-Engine and thus of automatically executing the rules in daily routine. Based on the automatic execution of

rules for data curation, data of patients are checked for plausibility, consistency, redundancy, etc. Curated and semantically annotated EHR data is amended with computed phenotype tags in the course of the phenotyping pipeline (see ▶Figure 4). The computed phenotype tags are used, e.g. by computerized decision support systems during patient care in the hospitals (cf. 4.2 and 4.3). Nourished EHR-local data will be used for patient care in the university hospital. Thus, the innovation cycle from bedside to bench and back from bench to bedside [20] is closed.

Tagging rules derived from phenotyping are developed and provided especially by the use cases ASIC and HELP in the Research & Development Factory (cf. 4.2 and 4.3). They use shared EHR data as input, which is provided by the *Data and Metadata Transfer Management* units (DMT) in the DICs of the partner university hospitals. The DMT acts under control of the Data Use and Access committee (DUA). Besides distribution of assets like data and knowledge, consulting for research groups is an important task of DMT. Depending on privacy regulations and patients' consents, the *Data Trustee and Privacy Management* unit (DTP) will ensure pseudonymization prior to data transfer.

3.2.2 SMItHIS Logical Tool Layer: Application Components and Services Supporting the DIC Tasks

The SMItHIS logical tool layer consists of application components and their interfaces and communication links.

At each partner site, application components are needed to support the local DIC, i.e. supporting the execution of tasks as well as storing and communicating data (entity types) as discussed in the previous section. Data and knowledge sharing between sites and between patient care and Research & Development Factory have to be enabled by appropriate communication links and dedicated components. Communication for data and knowledge sharing at the SMItHIS logical tool layer uses the IDS and is standards-based, if possible, using especially IHE profiles, CDA, FHIR and medical terminologies to ensure processual, syntactic and semantic interoperability, respectively (cf. section 3).

The architecture of the local system of application components and their communication links at each site, i.e. each of the local sub-information systems of SMItHIS, follows the DIC reference architecture. Using IHE profiles and the IDS based SMITH Market Place, the local sub-information systems are integrated in order to build SMItHIS as a whole. Still, SMItHIS allows for local peculiarities by applying the DIC reference architecture locally.

3.2.2.1 DIC Reference Architecture

As explained before, a DIC has to ingest data from various *Data Sources*, i.e. different application components of a local HIS. We classify communications between application components into three categories, A, B, and C, according to their interface type "if-type" (see legend in ▶Figure 2). Sources of type A are already designed according to common coding schemes and nomenclatures and they transfer data according to processual standards, using e.g. IHE profiles. Type B sources export data in standard formats, such as HL7, DICOM etc., but overarching standardization, preventing variations in a technical standard or semantically coded metadata for a unified data exchange is missing, and these have to be added in a transformation step. Finally, type C sources are proprietary, such as data provided by comma-separated (CSV) files. While data transformations usually are not necessary for type A sources, they need to be specified for type B and C sources. Such data transformations consist of data type conversions, transformations/calculations into a standard unit of measurement (e.g. weight in kg), and additions or replacements of the codes and labels of categorical data in accordance with a standard terminology or value set (coding scheme).

The *Data Integration Engine* will execute all kinds of data transformation and load processes from sources into the *Health Data Storage* (HDS). The HDS contains both a component for storing HL7 FHIR resources (*Health Data Repository*), providing data by RESTful interfaces [21], and an IHE *XDS Document Repository*, comprising clinical data in HL7 CDA documents [22]. Thus, the HDS is the central and harmonized base for all user-specified queries, data exchanges, reports, etc. Using the interface-type scheme (A, B, and C), we integrate data beyond department borders within a single hospital, i.e., laboratory, data and pre-analytic data are stored in the same way as treatment data from medical documentation systems. Note, at this generic stage we do not specify whether data of certain sources are virtually referred by or materially copied into the HDS.

The data catalogue contained in the HDS is a superset of the National Core Data Set for Health Care [23], a community-developed set of data from different health care domains ("modules"). It consists of basic modules for representing demographic data, encounters, diagnoses, procedures and medications in a standardized way and extended modules, e.g. for diagnostics, imaging, biobanking, Omics and ICU data. It will be utilized in inter-consortia use cases for evaluating the level of interoperability between the consortia described in this issue.

Metadata curation and harmonization services will define a process to harmonize common metadata elements from each site by creating descriptive metadata at both document and data element level, including semantic, technical and provenance metadata. Varying coding schemes and vocabularies will be semantically annotated, mapped and harmonized by alignment. A quality management process will be set up to ensure better metadata quality at the metadata entry stage by applying terminology management and semantic data validity checks. Metadata management will strictly follow the FAIR principles [24]. Having captured data that are semantically enriched by metadata (cf. "nourishing" in previous section) is a prerequisite for later analyses. Such metadata are typically provided by type A sources. Type B and C sources necessitate the capture of meaningful names and descriptions on the data element level, i.e., there is a name and a description for each column of a data table or class of data files, if they contain data of lowest granularity, such as images. Such metadata are centrally managed at each site by a *Metadata Repository* as one part of the

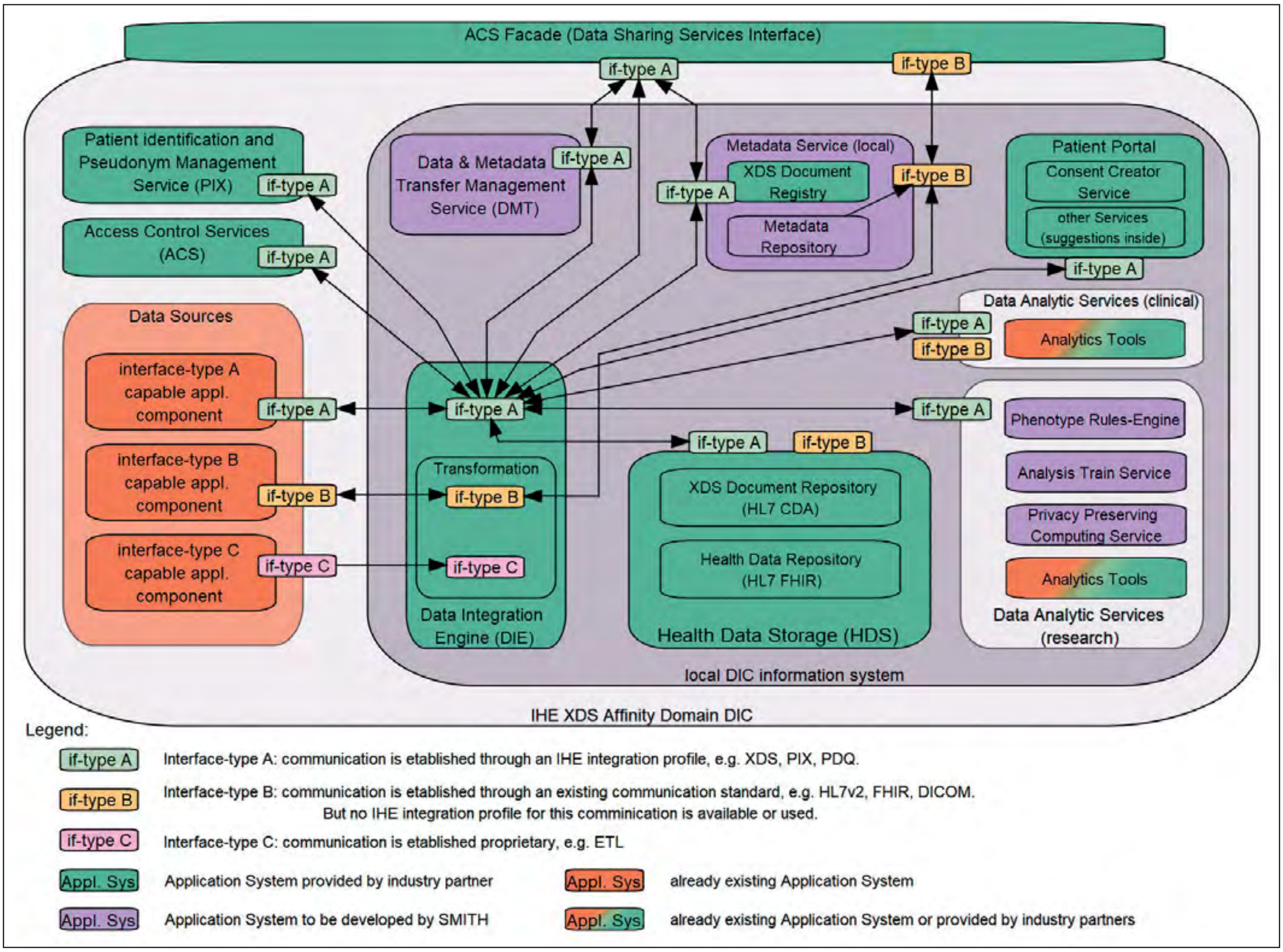
ACS Facade (Data Sharing Services Interface)
if-type A
if-type B
Patient identification and Pseudonym Management Service (PIX)
if-type A
Access Control Services (ACS)
if-type A
Data & Metadata Transfer Management Service (DMT)
if-type A
Metadata Service (local)
XDS Document Registry
Metadata Repository
if-type A
if-type B
Patient Portal
Consent Creator Service
other Services (suggestions inside)
if-type A
Data Analytic Services (clinical)
if-type A
if-type B
Analytics Tools
Data Sources
interface-type A capable appl. component
if-type A
interface-type B capable appl. component
if-type B
interface-type C capable appl. component
if-type C
if-type A
Transformation
if-type B
if-type C
Data Integration Engine (DIE)
if-type A
if-type B
XDS Document Repository (HL7 CDA)
Health Data Repository (HL7 FHIR)
Health Data Storage (HDS)
if-type A
Phenotype Rules-Engine
Analysis Train Service
Privacy Preserving Computing Service
Analytics Tools
Data Analytic Services (research)
local DIC information system
IHE XDS Affinity Domain DIC
Legend:
if-type A
Interface-type A: communication is etablished through an IHE integration profile, e.g. XDS, PIX, PDQ.
if-type B
Interface-type B: communication is etablished through an existing communication standard, e.g. HL7v2, FHIR, DICOM.
But no IHE integration profile for this comminication is available or used.
if-type C
Interface-type C: communication is etablished proprietary, e.g. ETL
Appl. Sys
Application System provided by industry partner
Appl. Sys
already existing Application System
Appl. Sys
Application System to be developed by SMITH
Appl. Sys
already existing Application System or provided by industry partners

Figure 2 SMITH-DIC Reference Architecture.

Rounded rectangles represent application systems and services. Small rounded rectangles represent interfaces. Lines between interfaces represent communication links. Arrows indicate initiation of communication.

Metadata Services. Conversely, the *XDS Document Registry* comprises another type of metadata. It describes each CDA document of the XDS Document Repository according to the Dublin Core standard [25], i.e., when and by whom the document has been imported (provenance metadata). Furthermore, just like in the Metadata Repository, subsets of metadata stored in the XDS Document Registry need to be shared or unified, e.g. metadata on how documents are categorized into classes and types. To support an overarching semantic interoperability, unified metadata at the conceptual level will be linked to internationally shared terminologies. National and international initiatives, such as Clinical Information Modeling Initiative (CIMI), the National Metadata Repository and the German value sets for XDS are carefully observed.

There are different *Data Analytic Services* within each DIC. Various analysis routines will be made available for data in the HDS. These routines are used, for example, for filtering patients (or cases) according to specific disease entities, such as pre-diabetes, taking data of different sources into account (examination data, laboratory data, and genetic data, if available). In this way, essential tasks for clinical research (hypothesis generation, patient recruitment for clinical trials) as well as for health care (quality indicators, hospital controlling) are seamlessly supported at the architectural level. Similarly, data will be prepared for sharing by the *Data & Metadata Transfer Management Service*. Between all consortia, an agreement towards the development of a common data model has been reached, where data transfer between DICs should be based on. The model will be based on published information models and best practice examples [26, 27, 28]. The national core data set will be mapped to an exchange model based on the reference model. In this way, metadata and data can be analyzed as to whether relevant data and sufficient cases/patients are available to set up an analysis project for an intended medical hypothesis. Analysis routines running on patient data will be executed in a *privacy-preserving computing environment* [29, 30] without interaction beyond the DIC; the results will then be shared with specific applicants of analysis projects. Large scale analysis routines utilizing clinical data from multiple sites will be executed in a distributed manner by bringing algorithms to the data, e.g. by providing docker containers [31].

In addition to the Data Analytic Services, there are also other application systems that interact with the HDS, such as the Patient Portal to support patient empowerment, which is shown in ▶Figure 2. In addition to care-related services, patients shall be able to get detailed information on conducted clinical research and data sharing and to “donate” their own data to selected projects. The first step is to provide a module showing which types of consent a patient has given. This *Consent Creator Service* also provides patients (or by proxy medical personnel) with the opportunity to consent electronically. A second step will provide information on the specific usage of their data in data analysis projects. A third step will be to provide a link to the EMR documents available. Further web portals based on DIC services are planned to provide additional support for *Data Trustee and Privacy Management* and *Data and Metadata Transfer Management* functions.

Among others, the reference architecture contains a *Patient Identification and Pseudonym Management Service*, which relies on IHE PIX and PDQ profiles to manage patient-related identifiers independent from clinical care or personal data.

All components are connected by the *Data Integration Engine*. In this way, analysis services can transparently access data of source systems via the Health Data Storage. The same holds for authenticated users who have been granted data access. Both – named users and automatically running analysis routines – extensively use available metadata, which are managed by the *Metadata Repository* and the *XDS Document Registry*.

All cross-community, i.e. transinstitutional communication requires authentication, authorization and auditing. Access control interoperability is crucial for a successful and sustainable health information exchange. The *Access Control Services (ACS)* use several international standards and frameworks (e.g. IHE profiles A/BPPC, XUA and ATNA, c.f. section 3.1.3) to fulfill access control workflows and to ensure proper auditing. The *ACS Facade (Data Sharing Services Interface)* handles all transinstitutional communication and encapsulates the associated security aspects based on the facade pattern of software design. It provides all IHE-based interfaces to access structured data and unstructured documents and a FHIR-based interface to access discrete data objects, and enforces the appropriate consents and authorizations.

3.2.2.2 SMItHIS High-Level Architecture

Based on the DIC reference architecture, the SMItHIS High-Level Architecture (see ▶Figure 3) describes the services for a connection of all DICs, other Medical Informatics Initiative consortia and external partners. The SMITH consortium introduces two additional architectural concepts: Data Sharing Services and the SMITH Market Place (SMP) (see ▶Figure 2). The high-level architecture clearly distinguishes between access and contracting services provided for researchers and DICs through the SMP and Data Sharing Services. While the SMP provides services for contracting and granting access to data by incorporating concepts from the Fraunhofer Industrial Data Space (IDS), The Data Sharing Services provide functions for started projects and connect all DIC ACS Facades for access control and standardized data sharing based on IHE profiles.

The SMP will be a central contracting endpoint for internal as well as external data consumers to provide data and knowledge to all interested researchers. As such, it will enable researchers to identify relevant data sets by providing a GUI to create and execute feasibility queries based on the HL7 CQL standard for queries to clini-

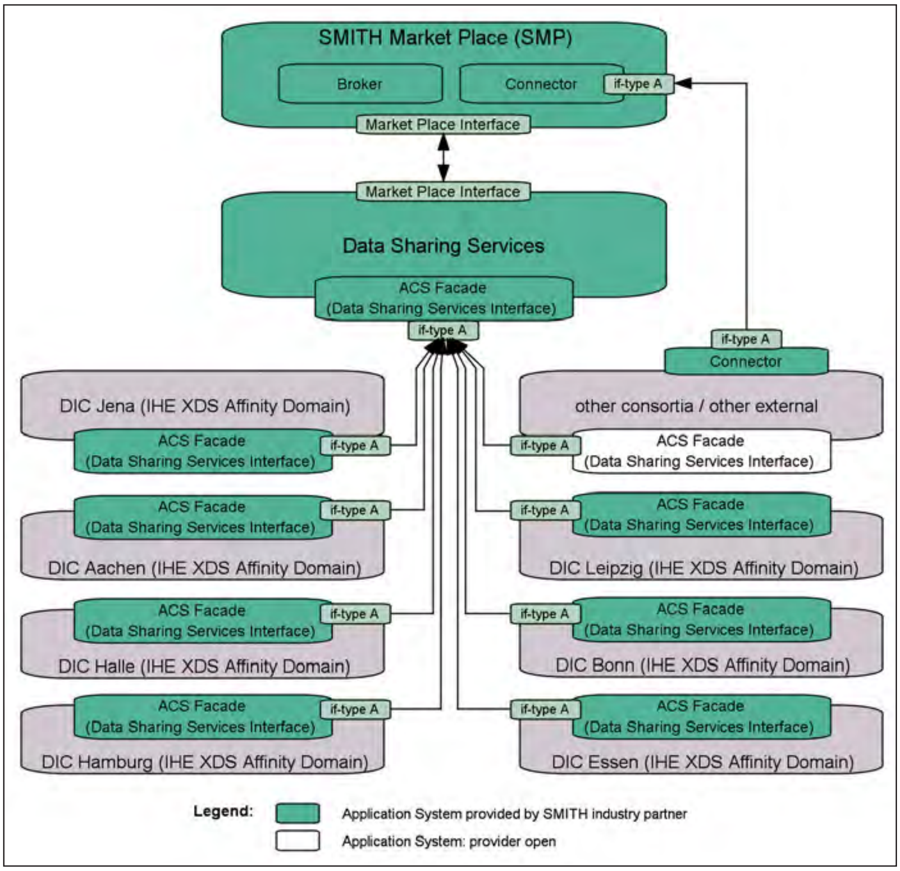
SMITH Market Place (SMP)
Broker
Connector
if-type A
Market Place Interface
Market Place Interface
Data Sharing Services
ACS Facade
(Data Sharing Services Interface)
if-type A
if-type A
Connector
DIC Jena (IHE XDS Affinity Domain)
ACS Facade
(Data Sharing Services Interface)
if-type A
other consortia / other external
if-type A
ACS Facade
(Data Sharing Services Interface)
ACS Facade
(Data Sharing Services Interface)
if-type A
DIC Aachen (IHE XDS Affinity Domain)
if-type A
ACS Facade
(Data Sharing Services Interface)
DIC Leipzig (IHE XDS Affinity Domain)
ACS Facade
(Data Sharing Services Interface)
if-type A
DIC Halle (IHE XDS Affinity Domain)
if-type A
ACS Facade
(Data Sharing Services Interface)
DIC Bonn (IHE XDS Affinity Domain)
ACS Facade
(Data Sharing Services Interface)
if-type A
DIC Hamburg (IHE XDS Affinity Domain)
if-type A
ACS Facade
(Data Sharing Services Interface)
DIC Essen (IHE XDS Affinity Domain)
Legend:
🟩 Application System provided by SMITH industry partner
□ Application System: provider open

Figure 3 SMItHIS high-level architecture. Lines and rectangles have the same meaning as in ▶Figure 2. Each grey shaded application system “DIC ...” is an adapted instance of the generic DIC reference model in ▶Figure 2.

cal knowledge. Once a suitable dataset has been identified, the SMP will support researchers in creating data use and access proposals. Furthermore, the SMP will support the Data Use and Access Committee in the review and approval process. Finally, when the proposal is approved, the SMP will activate the required access policies, enabling data transfer to the applicant. A central aspect is to enforce a contract between the data user and the data provider. Foreseeing a potential extension of the SMP, the contracting mechanisms will be implemented in a generic way, enabling the reuse of the same principles when sharing not only data, but also analytical services, for example. Similarly, the architecture of the SMP will be designed to enable the integration of third-party products. The SMP does not store any data on its own, i.e. it has no data storage functionality besides the one required for processing the actual contracts and access requests. The SMP will define and provide interfaces and facades to mediate between data use and access requests and consortia-specific data storage and access policies (interconnection to the Data Sharing Services).

The Data Sharing Services include an overarching identity provider (IHE HPD) for managing all participating identities and their roles for clinical studies and analysis projects. An Enterprise Master Patient Index (EMPI, IHE PIX/PDQ) manages the linkage of patient pseudonyms across the DICs whereby no personal patient data is stored. Access to the EMPI is protected by the Access Control System (SAML / XACML). It utilizes security tokens and policies (IHE XUA) to secure communication with the central components. The consent registry (IHE XDS.b, APPC) for providing information about consent policy documents is also included. This registry is connected to the Consent Creator Services, i.e. local Patient Portals for the creation of such consents based on patients' informed permissions to use their data. It can be queried by a Consent Consumer, e.g. the Data and Metadata Transfer Management Service, via IHE XCA Query and then further processed to check, for example, whether certain data can be shared and used for a specified purpose. Terminology services manage and provide codes and concepts on a consortium level. These can also be used to connect to external terminology services, e.g. at the national level, or to connect to other third-party services, such as existing public key infrastructures or external identity providers. Thus, the Data Sharing Services comprise functions for a federation of standardized data sharing and access control which can be cascaded from the DICs up to a national level.

In summary, the SMITH Market Place and the Data Sharing Services connect the individual sites and provide the functionalities and services to initiate projects, to share medical knowledge, data and algorithms and to support existing and new use cases for all possible partners (further university hospitals during the development and networking phase as well as additional healthcare institutions and network partners in an elaborated roll-out process) in the Medical Informatics Initiative.

4. Use Cases

In SMITH, we will implement one methodological use case (PheP) and two clinical use cases (ASIC and HELP). The use cases shall gather evidence for the usefulness and benefit of the DICs and their services implemented on top of the SMItHIS architecture. The clinical and patient-relevant impact of both clinical use cases will be evaluated by stepped wedge study designs.

4.1 Methodological Use Case PheP: Phenotype Pipeline

The SMITH consortium plans to develop and implement a set of tools and algorithms to systematically annotate and analyze patient-related phenotypes according to classification rules and statistical or dynamic models. Using DIC services (c.f. section 3.2.1, ▶Figure 1) EHR data in the HDS is annotated with computed phenotype tags. The annotations and derivatives will be made available for triggering alerts and actions, e.g. by study nurses in order to acquire additional findings from certain patients for data completion. The tags are subject to sharing and will be used for analyses of patient care and outcomes. This set of tools and algorithms constitutes the "Phenotype pipeline" (PheP). PheP will utilize data from the participating hospital's information systems. Besides structured data, unstructured textual data from clinical reports and the EMR will be incorporated and will be subject to natural language processing (NLP). The technology will be implemented in the DICs and first used to support the clinical use cases HELP and ASIC and in other research driven specific data use projects.

▶Figure 4 illustrates the basic tasks of and entity types used in PheP. This figure refines the generic DIC tasks as illustrated in ▶Figure 1:

- Within the Research & Development Factory, research groups may be established to develop knowledge to be used in the DIC. Especially phenotyping rules for phenotype classifications and NLP algorithms for information extraction from unstructured documents are in focus. Following a strict division between research and patient care, the research groups in the Research & Development Factory will use pseudonymized patient data provided for sharing by the DMT unit of one or more DICs.
- We use the term "phenotype" in a very general sense, referring to a set of attributes that can be attached to an individual. We distinguish between observable and computable phenotypes, defined as "a clinical condition, characteristic, or set of clinical features that can be determined solely from the data in EMRs and ancillary data sources and does not require chart review or interpretation by a clinician" [32]. Phenotype classification rules are based on classification trees, statistical models or simulation models. This will result in annotations (called tags) of a broad spectrum of attributes linked to a patient and his/her pattern of care and outcome.
- The main prerequisite for performing phenotyping is the availability of structured data. Phenotype information will be automatically extracted from unstructured EMR entries and clinical documents using NLP. For this, we plan two building blocks: a clinical document corpus (ClinDoC; for a preliminary version, cf. [33]) and a collection of NLP

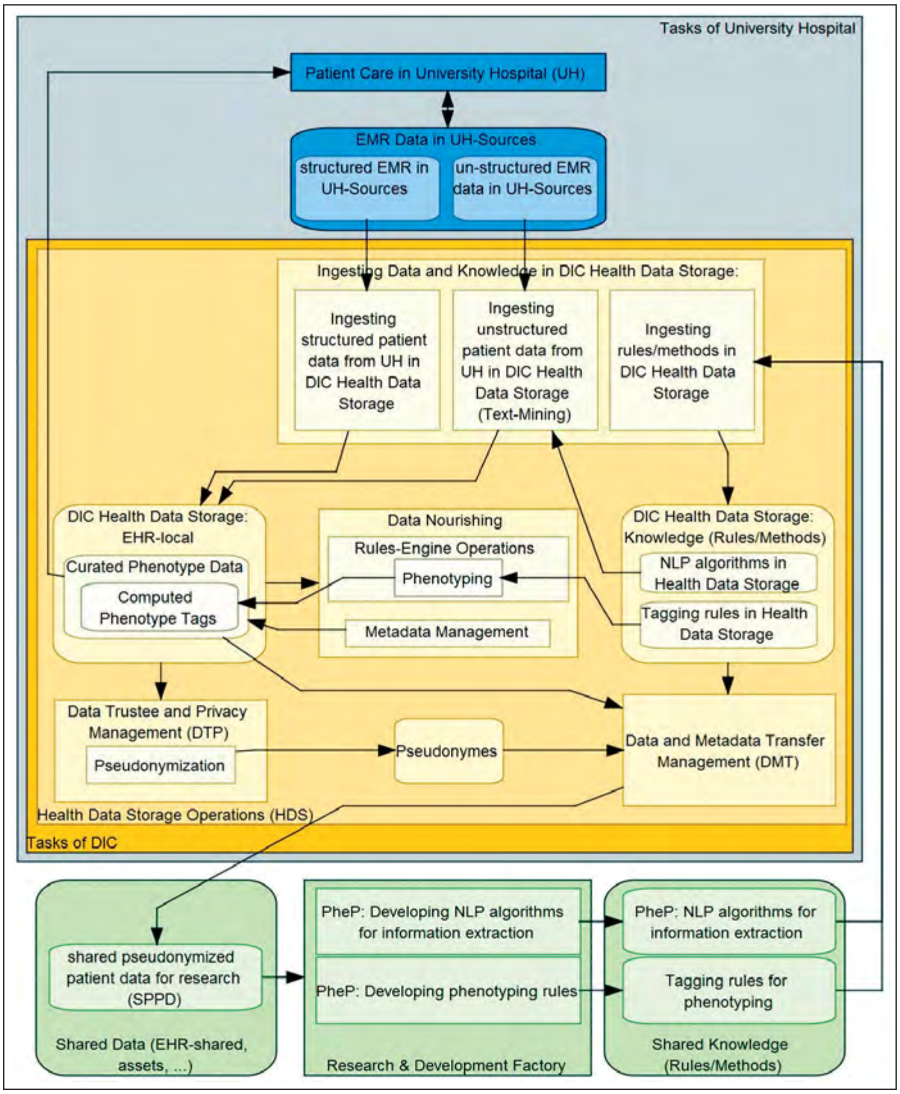
Tasks of University Hospital
Patient Care in University Hospital (UH)
EMR Data in UH-Sources
structured EMR in UH-Sources
un-structured EMR data in UH-Sources
Ingesting Data and Knowledge in DIC Health Data Storage:
Ingesting structured patient data from UH in DIC Health Data Storage
Ingesting unstructured patient data from UH in DIC Health Data Storage (Text-Mining)
Ingesting rules/methods in DIC Health Data Storage
DIC Health Data Storage: EHR-local
Curated Phenotype Data
Computed Phenotype Tags
Data Nourishing
Rules-Engine Operations
Phenotyping
Metadata Management
DIC Health Data Storage: Knowledge (Rules/Methods)
NLP algorithms in Health Data Storage
Tagging rules in Health Data Storage
Data Trustee and Privacy Management (DTP)
Pseudonymization
Pseudonymes
Data and Metadata Transfer Management (DMT)
Health Data Storage Operations (HDS)
Tasks of DIC
shared pseudonymized patient data for research (SPPD)
Shared Data (EHR-shared, assets, ...)
PheP: Developing NLP algorithms for information extraction
PheP: Developing phenotyping rules
Research & Development Factory
PheP: NLP algorithms for information extraction
Tagging rules for phenotyping
Shared Knowledge (Rules/Methods)

Figure 4 Detailed model of tasks and entity types of a DIC focusing on the phenotyping pipeline PheP. Here, entity types are depicted as rounded rectangles. Rectangles and arrows have the same meaning as in ►Figure 1.

components for processing German-language clinical documents, the SMITH Clinical Text Analytics Processor (ClinTAP; according to software engineering standards outlined in [34], cf. [19] for a preliminary version). ClinDoC will serve as the backbone for system performance evaluation and as a training and development environment for single NLP modules, which will be integrated, after quality control, in the ClinTAP information extraction (IE) engine.

- The Phenotyping Rules Engine (see ►Figure 2) automatically applies the tagging rules for phenotyping and computes the phenotype tags. As tools mature, rules and tags will be handed over for routine use to the DICs.
- Any DIC interested in using this knowledge, i.e. the rules for NLP and phenotyping, will ingest it into its Health Data Storage. Whereas structured data may easily be ingested in the HDS, unstructured documents from a hospital's EMR have to be processed by information extraction and text mining tools using the NLP algorithms.
- The Metadata Repository (MDR) will provide access to medical terminologies and code systems and thus supports semantic interoperability. In the PheP use case, we will enhance the MDR with metadata from a large variety of healthcare datasets, epidemiological cohorts, clinical trials and use cases. The MDR will also provide a directory of data items available projects in the Research & Development Factory (catalog of items).

4.2 Clinical Use Case ASIC: Algorithmic Surveillance of ICU Patients

Due to the epidemiological challenges in Germany, demand for intensive care medicine will increase over the next 10-15 years. At the same time, there will be a shortage of staff resources and so there is an urgent need for interoperable and intelligent solutions for using data to improve outcomes and processes. The need for outcome improvement is especially urgent in patients suffering from the acute respiratory distress syndrome (ARDS). Incidence of ARDS worldwide remains high, with 10.4% of total ICU admissions and 23.4% of all patients requiring mechanical ventilation [35]. The use case Algorithmic Surveillance of ICU Patients (ASIC) will therefore initially focus on ICU patients with ARDS.

By means of continuous analyses of data from the Patient Data Management System (PDMS), a model-based "algorithmic surveillance" of the state of critically ill patients will be established. To predict individual disease progression, ASIC will utilize pattern recognition technologies as well as established mechanistic systems medicine models, complemented by machine learning, both integrated in a hybrid virtual patient model [36]. Ultimately, the virtual patient model will enable individual prognoses to support therapy decisions, clinical trials and training of future clinicians. Training the virtual patient model requires high-performance computing.

The resulting ASIC system will be an on-line rule-based computerized decision-support system (CDSS). As such, it will extensively use the DIC services and especially the Phenotype Rules-Engine, which applies the phenotyping rules developed in PheP. Its aim is to accelerate correct and sound diagnosis and increase guideline compliance regarding ventilation. The rules may be realized by explicit decision trees, complex models, mechanistic or machine learning-based, or combinations of both. Its main component is the Diagnostic Expert Advisor (DEA). The development utilizes the PheP efforts in the following way (c.f. ▶Figure 4 and ▶Online Appendix):

- Usually PDMS data is well structured and will be ingested to the DIC Health Data Storage without NLP. Specific NLP-processed data mining from other sources (e.g. radiology and microbiology reports) will complement the data. The Data and Metadata Transfer Management (DMT) will provide pseudonymized patient data taken from the different partner HDS for training the DEA.
- The ASIC team contributes to the Research & Development factory. It will provide phenotyping rules for computing phenotype alert tags. Established machine learning approaches (with preference given to hierarchical clustering, random forest, Support Vector Machines, neural networks and Hidden Markov Models) will be applied to the training data collections to identify new patient classification schemes. We will also use the large data sets from all consortium's DICs to evaluate and utilize the predictivity of deep learning for time series.
- The identified data patterns and models will be expressed in a system of rules in the Phenotype Rules-Engine, which will be provided for ingesting into the knowledge bases of all partnering DICs. In the initial phase, this will be used for research purposes, in the later phase the ASIC-apps will be certified for care according to the medical products regulations.
- The rule system will be used for phenotyping of new patients using the EHR data in the HDS. The computed tags will then be used as input for the ASIC decision support system.

The utility of the ASIC approach will be demonstrated in the clinical care setting using a step wedge design. In this setting, the ASIC app will provide the interface between the DIC, the surveillance algorithms and the medical professionals at the bedside. Physicians will be automatically informed if *a priori* specified limits are exceeded. This alert function will be enhanced by smart latest action-based algorithms, preventing medical professionals from moot alerts (and, thus, alert fatigue).

4.3 Clinical Use Case HELP: A Hospital-wide Electronic Medical Record-based Computerized Decision Support System to Improve Outcomes of Patients with Blood-stream Infections

Antimicrobial stewardship programs (ABS programs) use a variety of methods to improve patient care and outcomes. One of the core strategies of ABS programs is prospective audit and feedback, which is labor-intensive and costly. When an infectious diseases specialist is involved in a patient's care and the physician in charge follows his/her recommendations, patients are more often correctly diagnosed, have shorter lengths of stay, receive more appropriate therapies, have fewer complications, and may use fewer antibiotics, overall. However, physicians with infectious disease expertise are rare in Germany and usually limited to a few university hospitals.

As an alternative, CDSS have been recommended by a recent German S3 Guideline on ABS programs [37]. CDSS may help to establish the correct diagnosis, to choose appropriate antimicrobial treatment, and to balance optimal patient care with undesirable aspects such as the development of antibiotic resistance, adverse events, and costs. However, the availability of CDSS for ABS program implementation is currently underdeveloped. HELP aims to develop and implement a CDSS. We will first focus on Staphylococcal bacteremia, i.e. staphylococcal bloodstream infections, since Staphylococci are the most frequent pathogens detected in blood cultures (BCs). Furthermore, a recent meta-analysis has shown that the particularly high mortality rate of staphylococcus aureus bacteremia can be reduced by 47% when an infectious diseases specialist is involved. This reduction is achieved by increased adherence to an evidence-based bundle of care which will be implemented into the HELP CDSS [38].

Development and use of the HELP-CDSS will use SMITH's Phenotype Pipeline PheP similarly like ASIC (see ▶Figure 4, and ▶Online Appendix):

- The HELP-CDSS requires the integration of structured clinical data and unstructured clinical reports from the partnering hospitals' EHRs. Previous work (e.g. [39]) has indicated a potential benefit of such an integration in research fields related to our HELP objective. Again, structured data will be ingested based on technical standards, whereas NLP algorithms will be used to ingest the reports into HDS, and the structured information derived therefrom.
- The DMT will provide and share pseudonymized patient data for developing the HELP-CDSS.
- As above, the HELP team is part of the Research & Development Factory and will develop phenotyping rules. In HELP, these rules are to classify patients according to their need of antibiotic stewardship. The proposed CDSS algorithm consists of several decision levels for which i) no human support is required, ii) all required information are or will become digitally available and iii) an automated directive/management proposal can be reported immediately to the treating physician.
- These rules will be used for phenotyping of new patients using the EHR data in the HDS. The Rules-Engine, therefore, monitors patient data in real-time, phenotypes the patients applying the rules, which represent the generalized and derived guidelines and models. The HELP CDSS uses the computed phenotype tags and issues an alert if a potential candidate for action is identified.

Feedback to and support of the health care professionals will be based on the HELP App, which will be based on a generic SMITH app – like the ASIC app. The goal of the HELP app is to provide immediate alerts/directives to the treating physician along the outlined algorithms.

5. Discussion

This paper outlines an ambitious program. It is ambitious not only with regard to the standards-based SMItHIS architecture, but with regard to the use cases as well. We are confident of being successful in the end because

- SMITH is not only an academic consortium. Moreover, the CIOs of the partnering hospitals and their information management departments are actively involved. Hence, we will be able to connect bench and bedside in both directions [20].
- A unique public-private partnership of complementary partners and the incorporation of experienced companies in the field of interoperable transinstitutional information systems and in the field of analyzing medical data (structured as well as unstructured) efficiently will effectively help implementing the standards-based architecture.

However, we are aware of the challenges and risks of such an endeavor, e.g.

- The DICs in general and the intended use cases require high quality clinical documentation. Although we will implement NLP methods to analyze unstructured documents, more and more structured documentation will be needed. Therefore, the sustainability of DICs depends on the ongoing support by the hospital's management and, most importantly, by their health care professionals. This support can only be obtained when healthcare and patients, in particular, will benefit from SMItHIS and thus promote the use of clinical data for its services.
- The rights of citizens and patients for informational self-determination and the legal regulations for data protection and privacy may come into conflict with important medical research goals in SMITH. The German Ethics Council recognized this potential for conflict and recommends legal regulations based on the principle of data sovereignty [40]. In SMITH, we will act closely to this recommendation especially by integrating a patient portal. The patient portal shall give individuals information on the usage of their data and shall provide the opportunity for data donation.

References

1. Federal Ministry of Education and Research, Germany. Medical Informatics Funding Scheme: Networking data – improving health care. Berlin, Germany; 2015. Available from: https://www.bmbf.de/pub/Medical_Informatics_Funding_Scheme.pdf.
2. CCSDS Secretariat. Reference Model for an Open Archival Information System (OAIS): Magenta Book: NASA Headquarters; 2012. Available from: http://public.ccsds.org/publications/archive/650x0m2.pdf.
3. Kirsten T, Kiel A, Wagner J, Rühle M, Löffler M. Selecting, Packaging, and Granting Access for Sharing Study Data. In: Eibl M, Gaedke M, editors. Informatik 2017 – Bände I-III: Tagung vom 25.-29. September 2017 in Chemnitz. Bonn: Gesellschaft für Informatik; 2017. p. 1381–1392 (GI-Edition Proceedings; vol. 275).
4. Kirsten T, Kiel A, Rühle M, Wagner J. Metadata Management for Data Integration in Medical Sciences – Experiences from the LIFE Study. In: Mitschang B, Ritter N, Schwarz H, Klettke M, Thor A, Kopp O, et al., editors. BTW 2017: Datenbanksysteme für Business, Technologie und Web (Workshopband); Tagung vom 6. – 7.März 2017 in Stuttgart. Bonn: Gesellschaft für Informatik; 2017. p. 175–194 (GI-Edition – lecture notes in informatics (LNI) Proceedings; volume P-266).
5. Loeffler M, Engel C, Ahnert P, Alfermann D, Arelin K, Baber R, et al. The LIFE-Adult-Study: Objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. BMC Public Health 2015; 15: 691.
6. Löffler M, Binder H, Kirsten T. Leipzig Health Atlas; 2018 [cited 2018 Jan 30]. Available from: https://www.health-atlas.de/en.
7. Winter A, Brigl B, Wendt T. Modeling Hospital Information Systems (Part 1): The Revised Three-Layer Graph-Based Meta Model 3LGM2. Methods Inf Med 2003; 42(5): 544–551.
8. Winter A, Haux R, Ammenwerth E, Brigl B, Hellrung N, Jahn F. Health Information Systems – Architectures and Strategies. London: Springer; 2011.
9. Staemmler M. Towards sustainable e-health networks: does modeling support efficient management and operation? In: Kuhn KA, Warren JR, Leong T-Y, editors. Proceedings of Medinfo 2007 (Part 1). Amsterdam: IOS Press; 2007. p. 53–57 (Stud Health Technol Inform; vol. 129).
10. Stäubert S, Winter A, Speer R, Loffler M. Designing a Concept for an IT-Infrastructure for an Integrated Research and Treatment Center. In: Safran C, Marin H, Reti S, editors. MEDINFO 2010 Partnerships for Effective eHealth Solutions. Amsterdam: IOS Press; 2010. p. 1319–1323 (Stud Health Technol Inform; vol. 160).
11. Winter A, Takabayashi K, Jahn F, Kimura E, Engelbrecht R, Haux R, et al. Quality Requirements for Electronic Health Record Systems: A Japanese-German Information Management Perspective. Methods Inf Med 2017; 56(Open): e92–e104. Available from: https://www.schattauer.de/index.php?id=5236&mid=27796&L=1.
12. IHE International Inc. IHE Profiles; 2017 [cited 2018 Jan 4]. Available from: http://wiki.ihe.net/index.php/Profiles#IHE_IT_Infrastructure_Profiles.
13. Stäubert S, Schaaf M, Jahn F, Brandner R, Winter A. Modeling Interoperable Information Systems with $3LGM^2$ and IHE. Methods Inf Med 2015; 54(5): 398–405.
14. HL7.org. Introducing HL7 FHIR; 2014 Jan 1 [cited 2018 Jan 4]. Available from: https://www.hl7.org/fhir/summary.html.
15. Otto B, Jürjens J, Schon J, Auer S, Menz N, Wenzel S et al. INDUSTRIAL DATA SPACE: Digitale Souveränität über Daten; 2016. Available from: www.industrialdataspace.org.
16. Juhr M, Haux R, Suzuki T, Takabayashi K. Overview of recent trans-institutional health network projects in Japan and Germany. J Med Syst 2015; 39(5): 234.
17. International Organization for Standardization (ISO). ISO 18308: 2011 Requirements for an electronic health record architecture; 2011 [cited 2014 Jan 23].
18. Garets D, Davis M. Electronic Medical Records vs. Electronic Health Records: Yes, There Is a Difference. Chicago, IL: HIMSS Analytics; 2006. Available from: http://s3.amazonaws.com/rdcms-himss/files/production/public/HIMSSorg/Content/files/WP_EMR_EHR.pdf.
19. Hellrich J, Matthies F, Faessler E, Hahn U. Sharing models and tools for processing German clinical texts. Stud Health Technol Inform 2015; 210: 734–738.
20. Marincola FM. Translational Medicine: A two-way road. J Transl Med 2003; 1(1): 1.
21. FHIR Developer Introduction; 2015 [cited 2016 Apr 13]. Available from: http://hl7.org/fhir/overview-dev.html#1.8.1.6.
22. Health Level Seven International. CDA® Release 2: Health Level Seven International; 2016 [cited 2016 Feb 12]. Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7.
23. Redaktionsgruppe Kerndatensatz der AG Interoperabilität des Nationalen Steuerungsgremiums der Medizininformatik-Initiative; 2017 Mar 10. Available from: http://www.medizininformatik-initiative.de/sites/default/files/inline-files/MII_04_Kerndatensatz_1–0.pdf.
24. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016; 3: 160018.
25. Darmoni SJ, Thirion B, Leroy JP, Douyère M. The use of Dublin Core metadata in a structured health resource guide on the internet. Bull Med Libr Assoc 2001; 89(3): 297–301.
26. Meineke, Frank, Stäubert S, Löbe M, Winter A. A Comprehensive Clinical Research Database based on CDISC ODM and i2b2. Stud Health Technol Inform 2014; 205: 1115–1119.
27. Murphy SN, Mendis M, Hackett K, Kuttan R, Pan W, Phillips LC, et al. Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside. AMIA Annu Symp Proc 2007; 2007: 548–552.

28. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform 2015; 216: 574–578.
29. Kho AN, Cashy JP, Jackson KL, Pah AR, Goel S, Boehnke J, et al. Design and implementation of a privacy preserving electronic health record linkage tool in Chicago. J Am Med Inform Assoc 2015; 22(5): 1072–1080.
30. Vatsalan D, Sehili Z, Christen P, Rahm E. Privacy-Preserving Record Linkage for Big Data: Current Approaches and Research Challenges. In: Zomaya A, Sakr S, editors. Privacy-Preserving Record Linkage for Big Data: Handbook of Big Data Technologies. Springer; 2017.
31. Löbe M, Ganslandt T, Lotzmann L, Mate S, Christoph J, Baum B et al. Simplified Deployment of Health Informatics Applications by Providing Docker Images. Stud Health Technol Inform 2016; 228: 643–647.
32. Richesson RL, Smerek MM, Blake Cameron C. A Framework to Support the Sharing and Reuse of Computable Phenotype Definitions Across Health Care Delivery and Clinical Research Applications. EGEMS (Wash DC) 2016; 4(3): 1232.
33. Hahn U, Matthies F, Lohr C, Löffler M. 3000PA—Backbone for a national clinical reference corpus of German. MIE 2018. In: Ugon A, Karlsson D, Klein GO, Moen A, editors. Building Continents of Knowledge in Oceans of Data: The future of co-created ehealth. Amsterdam: IOS Press; 2018. p. 26–30.
34. Hahn U, Matthies F, Faessler E, Hellrich J. UIMA-based JCoRe 2.0 goes GitHub and Maven Central: State-of-the-art software resource engineering and distribution of NLP pipelines. In: Calzolari N, editor. LREC 2016: [proceedings]. [S. l.: s. n.]; 2016. p. 2502–2509.
35. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med 2015; 7(299): 299ra122.
36. Wolkenhauer O, Auffray C, Brass O, Clairambault J, Deutsch A, Drasdo D, et al. Enabling multiscale modeling in systems medicine. Genome Med 2014; 6(3): 21.
37. With K de, Allerberger F, Amann S, Apfalter P, Brodt H-R, Eckmanns T, et al. Strategies to enhance rational use of antibiotics in hospital: A guideline by the German Society for Infectious Diseases. Infection 2016; 44(3): 395–439.
38. Vogel M, Schmitz RPH, Hagel S, Pletz MW, Gagelmann N, Scherag A, et al. Infectious disease consultation for Staphylococcus aureus bacteremia – A systematic review and meta-analysis. J Infect 2016; 72(1): 19–28.
39. DeLisle S, South B, Anthony JA, Kalp E, Gundlapallli A, Curriero FC, et al. Combining free text and structured electronic medical record entries to detect acute respiratory infections. PLoS One 2010; 5(10): e13377.
40. Deutscher Ethikrat. Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung: Stellungnahme. Berlin; 2017 Nov 30. Available from: http://www.ethikrat.org/dateien/pdf/stellungnahme-big-data-und-gesundheit.pdf.