

Viewpoint

Ten Topics to Get Started in Medical Informatics Research

Markus Wolfien^{1,2}, Dr-Ing; Najia Ahmadi¹, MSc; Kai Fitzer³, MSc; Sophia Grummt¹, MSc; Kilian-Ludwig Heine¹, MA; Ian-C Jung¹, MSc; Dagmar Krefting⁴, Prof Dr; Andreas Kühn¹, Dr med; Yuan Peng¹, MSc; Ines Reinecke¹, MSc; Julia Scheel⁵, MSc; Tobias Schmidt⁶, MA; Paul Schmücker⁶, Prof Dr; Christina Schüttler⁷, PhD; Dagmar Waltemath^{3,8}, Prof Dr; Michele Zoch¹, Dipl.-Wi.-Inf.; Martin Sedlmayr^{1,2}, Prof Dr

¹Institute for Medical Informatics and Biometry, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

²Center for Scalable Data Analytics and Artificial Intelligence, Dresden, Germany

³Core Unit Data Integration Center, University Medicine Greifswald, Greifswald, Germany

⁴Department of Medical Informatics, University Medical Center, Goettingen, Germany

⁵Department of Systems Biology and Bioinformatics, University of Rostock, Rostock, Germany

⁶Institute for Medical Informatics, University of Applied Sciences Mannheim, Mannheim, Germany

⁷Central Biobank Erlangen, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

⁸Department of Medical Informatics, University Medicine Greifswald, Greifswald, Germany

Corresponding Author:

Markus Wolfien, Dr-Ing

Institute for Medical Informatics and Biometry

Faculty of Medicine Carl Gustav Carus

Technische Universität Dresden

Fetscherstraße 74

Dresden, 01307

Germany

Phone: 49 3514587723

Email: markus.wolfien@tu-dresden.de

Abstract

The vast and heterogeneous data being constantly generated in clinics can provide great wealth for patients and research alike. The quickly evolving field of medical informatics research has contributed numerous concepts, algorithms, and standards to facilitate this development. However, these difficult relationships, complex terminologies, and multiple implementations can present obstacles for people who want to get active in the field. With a particular focus on medical informatics research conducted in Germany, we present in our Viewpoint a set of 10 important topics to improve the overall interdisciplinary communication between different stakeholders (eg, physicians, computational experts, experimentalists, students, patient representatives). This may lower the barriers to entry and offer a starting point for collaborations at different levels. The suggested topics are briefly introduced, then general best practice guidance is given, and further resources for in-depth reading or hands-on tutorials are recommended. In addition, the topics are set to cover current aspects and open research gaps of the medical informatics domain, including data regulations and concepts; data harmonization and processing; and data evaluation, visualization, and dissemination. In addition, we give an example on how these topics can be integrated in a medical informatics curriculum for higher education. By recognizing these topics, readers will be able to (1) set clinical and research data into the context of medical informatics, understanding what is possible to achieve with data or how data should be handled in terms of data privacy and storage; (2) distinguish current interoperability standards and obtain first insights into the processes leading to effective data transfer and analysis; and (3) value the use of newly developed technical approaches to utilize the full potential of clinical data.

(*J Med Internet Res* 2023;25:e45948) doi: [10.2196/45948](https://doi.org/10.2196/45948)

KEYWORDS

medical informatics; health informatics; interdisciplinary communication; research data; clinical data; digital health

Introduction

Digital health care information, as opposed to analog information, empowers clinicians, researchers, and patients with a wealth of information aiming to improve diagnosis, therapy outcome, and clinical care in general. According to Wyatt and Liu [1], medical informatics is the study and application of methods to improve the management of patient data, clinical knowledge, population data, and other information relevant to patient care and community health. Medical informatics can be seen as the subset of health informatics that is focused on clinical care, while the latter encompasses a wider range of applications. However, knowing, integrating, and using current computational technologies bears numerous pitfalls, limitations, and questions [2]. To shed light on current standards, applications, and underlying technologies, we present 10 topics to get started in the field of medical informatics research. Our key objective here was to improve interdisciplinary communication among stakeholders (eg, clinicians, experimental researchers, computer scientists, students, patient representatives), thereby bringing everyone on the same page of state-of-the-art medical informatics practices. In particular, improved interdisciplinary communication is essential in real-world problems and can be motivated by the following aspects:

- Advancing open research: Open collaboration between parties from different disciplines can lead to new research questions, innovative approaches, and novel discoveries [3].
- Bridging knowledge domains: Interdisciplinary communication can stimulate novel solutions, allowing

researchers to gain a more comprehensive understanding of a specific problem or phenomenon [4], or can improve clinical decision-making [5].

- Addressing complex problems: Complex problems, such as the latest disease outbreak, require input from multiple domains to be comprehensively understood. Here, interdisciplinary communication is one key aspect to pinpoint the root causes and develop effective solutions [6].
- Promoting scientific inclusivity and diversity: Interdisciplinary communication was recently shown to foster diversity and inclusivity in science, by bringing together researchers from different backgrounds, cultures, and perspectives [7,8].

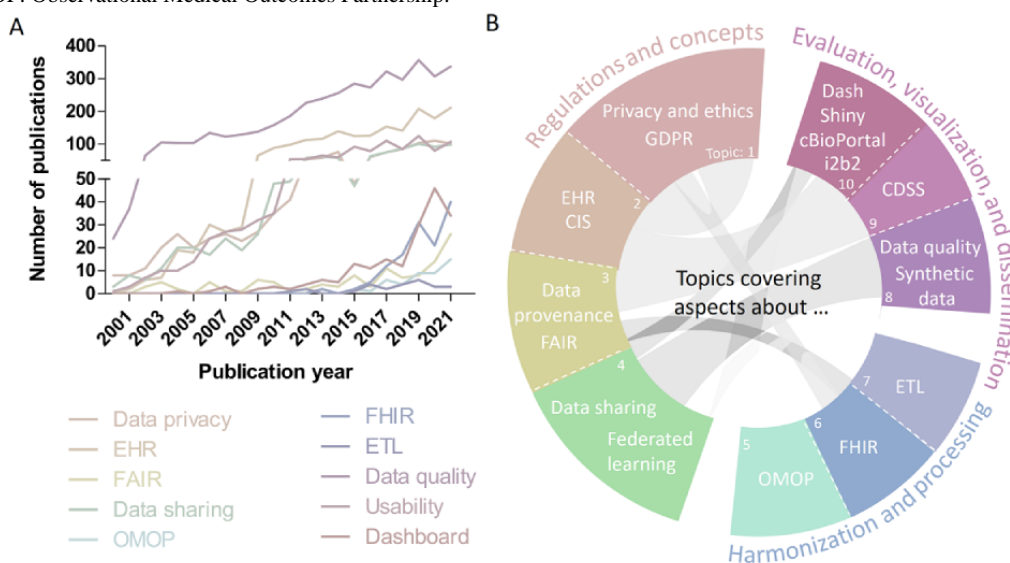
Methods

Here, we describe in detail how the initial topics have been selected from the literature and what design principles and structure each topic follows. A brief outline of the utilized methods for topic dissemination and an exemplary embedding into an educational training program are also presented.

Topic Selection

The initial topics were defined based on current developments in the health informatics field and an increasing number of published manuscripts between 2000 and 2021 (based on title-abstract-keyword screening in Scopus using the keywords “Health” AND “Informatics” AND “domain”) in the respective subdomains (Figure 1A). After a first definition of the specific topics, these were critically revised by internal and external domain experts, as well as scientists previously not familiar with medical informatics research.

Figure 1. Schematic summary and representation of the presented topics: (A) brief literature screening (title-abstract-keywords) for published manuscripts between 2000 and 2021, and the y-axis gap provides improved visibility of the less-occurring keywords; (B) most common topic terminologies, keywords (color-coded sections), and potential connections (grey) among topics in the medical informatics research domain. CDSS: clinical decision support system; CIS: clinical information system; EHR: electronic health record; ETL: extract, transform, and load; FAIR: findable, accessible, interoperable, reusable; FHIR: Fast Healthcare Interoperability Resources; GDPR: General Data Protection Regulation; i2b2: Informatics for Integrating Biology and the Bedside; OMOP: Observational Medical Outcomes Partnership.



Topic Design

The initial number of important topics and keywords exceeded the anticipated number of 10 topics, which found inspiration from the “Ten Simple Rules” collection in PLOS Computational Biology [9]. This is why the authors merged the most matching terms topic wise into groups. These groups finally produced topics that represent the broad range of the medical informatics domain in 3 main concepts, namely “Regulations and concepts,” “Harmonization and processing,” and “Evaluation, visualization, and dissemination” (Figure 1B). Figure 1B also shows the initial keywords for each individual topic, as well as potential cross references between topics, which are highlighted in grey. The following sections provide important “do's and don'ts,” practical hints, and best practice guidelines. Further in-depth resources and practical tutorials will provide basic introductions to the referred domains. Kohane et al [10] already showed the importance of such clarifying introductions. This work extends the initial study and, in addition, provides detailed examples from the German national Medical Informatics Initiative (MII) [11].

All topics were divided into 3 parts to improve comprehension by the readers:

1. Introduction: Background definitions for the specific context that motivated the topic
2. Insight: Practical context to get started, including how to avoid pitfalls, state current limitations, and address current challenges
3. Impact: Take home message and useful resources and best practices to deepen knowledge about the topic

Topic Utilization, Extension, and Embedding

Since it is of the utmost importance to keep the content current and as versatile as possible, we initiated an online resource at GitHub, in which contributions are highly emphasized [12]. Here, keywords and the corresponding literature are collected to allow for swift extension of the currently presented literature body in this article. In addition, the introduction of novel important topics that are not covered in this article might be included. To additionally demonstrate the practicability and adaptability of our proposed topic content, we exemplarily present how these can be embedded in higher education training and share external, introductory hands-on material (Table 1).

Results

Regulations and Concepts

Topic 1: Privacy and Ethics—“Data Privacy and Ethics Are the Most Important Assets in the Clinical Domain.”

Introduction

Health information is sensitive and hence needs to be highly protected and should not be generously shared. Sharing regulations and data privacy matters are defined in the European General Data Protection Regulation (GDPR) [13]. The implementation of the GDPR is an *ongoing* process as the quickly evolving technology, data, and scientific practices demand continuous improvement, which include periodic

adaptations of the technical and legal aspects [14,15]. In terms of ethics and with the rise of novel technologies, like artificial intelligence (AI), the possible re-identification of data, such as images and genomic information, is a major concern [16,17].

Insight

Anonymization is one important way to keep data private. It can also be achieved for high-dimensional data by changing patient-specific identifiers through removal, substitution, distortion, generalization, or aggregation [18]. In contrast, data pseudonymization is another de-identification procedure by which personally identifiable information fields within a data record are replaced by one or more artificial identifiers or pseudonyms [19]. To overcome the paucity of annotated medical data in real-world settings and (fully) save the patients' anonymity, *synthetic data* generation is used to increase the diversity in data sets and to enhance the robustness and adaptability of AI models [20]. To conform with ethical regulations in a research context, medical data are only available in a highly controlled manner and according to strict procedures. New concepts, such as “systemic oversight” [21] or “embedded ethics” [22], might be needed to tackle the new data-driven developments around “medical big data” and AI in health care. To engage with the adoption of broad consent, systemic oversight was suggested as an approach, in which mechanisms like auditing mechanisms, expert advice, and public engagement initiatives (among others) should be adapted as additional layers to the newly arising ecosystem of health data [21]. Recently, embedded ethics was jointly suggested by ethicists and developers to address ethical issues via an iterative and continuous process from the outset of development, which could be an effective means of integrating robust ethical considerations into practical development [22]. A digital representation of information encoded in signed consent forms is needed to facilitate common data use and sharing, as already implemented in an MII informed consent template [23].

Impact

As a researcher in medical informatics, it is inevitable to be informed and knowledgeable about the fact that patients own their medical records and any use of those data requires great care. In Germany, health care providers can only use the data for first medical use. Secondary use, like research, needs to be approved by either broad or individual consent, which can be made available via the electronic health record (EHR). In addition to digitization efforts, it is still a considerable hurdle to convince patients to make their data available for medical research because personal skepticism commonly makes the entire data acquisition process more difficult [24]. Here, well-received external communication, transparency, and increased awareness are necessary for substantial improvements. In general, it is a balance between privacy, patient needs, and the use of data for the common good versus economic interests [25]. In particular, one should be aware of the specific legal regulations that apply within the country and additionally get in touch with the relevant data protection departments. Following this, a plan for infrastructure that meets these regulations and that contains, for example, a trustee for the electronic recording of patient consent and anonymization or direct pseudonymization processes to collect the data needs to

be developed. Risk assessments for potential data leakage, approvals by ethics committee, as well as consultation with a data protection officer are essential considerations to further assure data security.

Topic 2: EHR and Clinical Information Systems—“Get to Know Your Clinical Information System to Understand the Required Data.”

Introduction

Hospitals run clinical information systems (CIS) to collect, store, and alter clinical data about patients. A CIS, independent of the specialization and specific vendor, covers many clinical subdomains and integrates patient-related data to support doctors in their daily routine. Without a doubt, medical data are only useful if meaningful information can be derived from them. This requires high-quality data sets, seamless communication across IT systems, and standard data formats that can be processed by humans and machines [2]. Typical challenges in clinical IT implementations, especially for patient recruitment systems, were recently evaluated by Fitzer et al [26] for 10 German university hospitals, including requirements for data, infrastructure, and workflow integration. The implementation of an EHR, including an individual's medical data in a bundled form, into the CIS is a key aspect to prevent low reliability and poor user-friendliness of EHRs, which has recently been shown to affect time pressure among medical staff [27]. For example, in Scandinavia, the United States, and the United Kingdom, the Open Notes initiative [28] facilitates patients' access to EHRs and health data sharing via “PatientsKnowBest” to give health care professionals and families direct access to medical information [29].

Insight

An EHR is used primarily for the purposes of setting objectives and planning patient care, documenting the delivery of care, and assessing the outcomes of care [30]. EHRs have so far consisted of unstructured, narrative text as well as structured, coded data. Thus, it will be necessary to implement more systematic terminologies and codes so that the data contained in these records can be reused in clinical research, health care management, health services planning, and government reporting in an improved manner [31,32]. Since the domain of medical informatics is rather new, there are many possibilities for software solutions to improve EHR-related issues [33]. Exemplary for the EHR domain, the Systematized Nomenclature of Medicine and Clinical Terms (SNOMED CT) is utilized to develop comprehensive high-quality clinical content [34]. It provides a standardized way to represent clinical phrases captured by the clinician and enables automatic interpretation of these, which is showcased in a “five-step briefing” [35]. Interestingly, the number of annual publications on this subject has decreased since 2012. However, the need for a formal semantic representation of free text in health care remains, and automatic encoding into a compositional ontology could be a solution [36]. In terms of usability and user acceptance, evaluations and improvements of EHRs and clinical decision support systems (CDSS) are currently ongoing [37], for which already well-received examples can be attributed to CeoSYS [38] or the IPSS-M Risk Calculator [39]. Moreover, the actions

of patients directly contributing to their own EHR records are also being evaluated. The study by Klein et al [40] indicates that such an approach facilitates the development of individual solutions for each patient, which in turn requires a flexible EHR during the course of a treatment process. Additionally, it was argued that data incorporation via different devices can also facilitate the convenient utilization of the application and, hence, may increase secondary use.

Impact

Modern CIS support the interaction by doctors and patients with the recorded patient data (eg, using the EHR or patient portals, eHealth platforms). It is important to understand the basic architecture, especially challenges [26], of the hospital IT infrastructure to know where data are located and how they can be retrieved and integrated. Major improvements can be made when supporting international standards for data exchange. Beyond standard EHR, this includes interoperability standards like Fast Healthcare Interoperability Resources (FHIR; see Topic 6) and standard data models like the Observational Medical Outcomes Partnership (OMOP; see Topic 7). These criteria should be considered with every new order of clinical systems.

Topic 3: Data Provenance—“Trace Your Data, Even Within Large-scale Efforts.”

Introduction

Meaningful and standardized metadata facilitate the interpretation of, retrieval of, and access to data [41]. When explainable data are processed with interoperable tools, scientists can create automated and reusable workflows and provide access to reproducible research outcomes and data analysis pipelines [42].

Insight

Data provenance describes the history of digital objects, where they came from, how they came to be in their present state, and who or what acted upon them [43]. In health care, provenance maintains the integrity of digital objects (eg, the results of data analyses engender greater trust if their provenance shows how they were obtained). In addition, it can be used to deliver auditability and transparency, specifically, in learning health systems, and it is applicable across a range of applications [44]. Inau et al [45] argued that the lessons learned from “FAIRification” processes in other domains will also support evidence-based clinical practice and research transparency in the era of big medical data and open research. Further work demonstrated that a findable, accessible, interoperable, reusable (FAIR) research data management plan can provide a data infrastructure in the hospital for machine-actionable digital objects [46]. Recently, the openEHR approach was also suggested for creating FAIR-compliant clinical data repositories as an alternative representation [47].

Impact

Key data management requirements are defined by the FAIR guiding principles [48]. Since data protection laws led to additional requirements for data privacy and data security, the FAIR-Health principles focused on defining additional requirements for information on the sample material used from

biobanks, for provenance information, and incentive schemes [49]. Further work is needed to establish provenance frameworks in health research infrastructures [50].

Topic 4: Data Sharing—“If Data Won’t Come to the Model, the Model Must Go to the Data.”

Introduction

Cross-sectional medical data-sharing is critical in modern clinical practice and medical research, in which the challenge of privacy-preserving transfer and utility needs to be addressed [51]. In order to facilitate high reuse of the data, a decentralized computational scheme that treats the available data as part of a federated (virtual) database, avoiding centralized data collection, processing, and raw data exchanges, is still needed in many countries to analyze large and widespread clinical data [52].

Insight

One possible solution for this federated learning approach is DataSHIELD [53]. In particular, orchestrating privacy-protected analyses of “medical big data” from different resources is applicable within R and DataSHIELD [54]. Here, the developed computerized models represent mathematical concepts or trained machine learning (ML)–based approaches to solve a specific task. In this sense, the model is applied to distributed data sets of the protected (clinical) server infrastructure, and the user only sees the model results but does not retrieve any medical records. Moreover, implementations in other programming languages (eg, Python, Julia) have been introduced in the genomic domain and beyond [55]. Further concepts, such as Personal Health Train, specifically follow the FAIR principles during distributed analyses [56]. Secure multiparty computation (SMPC) is also a viable technology for solving clinical use cases that require cross-institution data exchange and collaboration [57]. Current limitations are thought to be addressed in a stepwise manner [58] or as blockchain [59].

Impact

By using approaches for distributed analyses, researchers are able to train, test, and validate their models on large-scale real-world clinical data. In combination with standardized data formats, these 2 concepts facilitate the use of those models in clinical routine, potentially in the form of a CDSS. This provides a basis for secondary use of observational data in the context of clinical trials, which show particular potential for identifying data characteristics in small cohorts (eg, identification of the individual patient risk for rare diseases or comorbidities).

Harmonization and Processing

Topic 5: Extract, Transform, and Load (ETL)—“ETL Processes Are Computational Approaches for Data Harmonization and Data Unification.”

Introduction

Data handling in medical informatics remains a major challenge. Even though most data in medicine are available electronically, the data often lack interoperability [60]. As a first step to actually use the data, processes to extract, transform, and load (ETL) are needed to obtain harmonized data from different data systems or clinical entities. One important example, among

many others, reflects the uniform representation of the date and time in a common format (eg, Year-Month-Date, not Date-Month-Year). The ETL process is therefore a crucial, individual step toward data unification in large clinical systems, which must be secure, safe, and accurate [61].

Insight

The design of an ETL process faces several challenges, including the following: (1) The ETL process should be able to process huge amounts of data at once [62]; (2) the ETL process should be repeatable—if the source data change, the ETL process needs to be rerun to process the source data (Observational Health Data Sciences and Informatics [OHDSI]) [63]; (3) expert-level anonymization methodologies might be integrated into ETL workflows whenever possible [61]; and (4) there is a need to check for loss of data and compromised data integrity. The latter was highlighted in a recent study, in which inaccurate cohort identification took place because erroneous vocabulary mappings of a common data model were used (eg, ETL programming bugs and errors not captured during the quality assurance stages) [64]. Common solutions to implement ETL processes are code-based (eg, FHIR-to-OMOP [65]) or via Pentaho Data Integration, which is one of many ETL tools. Further subsequent processing may also include loading data into research data repositories, like OMOP (see Topic 7), tranSMART, and Talend Open Studio, which is a central component of the Integrated Data Repository Toolkit [66].

Impact

Since ETL processes are at the core of data handling, all risks associated with the ETL process need to be thoroughly checked, identified, and assessed, and contingency plans to mitigate these risks should be in place [67]. Once the ETL processes are executed, the resulting data will be *trusted* by researchers, who heavily rely on comprehensively checked data integrity to be able to conduct their research on this basis.

Topic 6: FHIR—“Set FHIR to Gain a Communication Standard for Real-time Applications at the Device-to-Device Level.”

Introduction

Interoperability levels can be divided into technical, syntactic, semantic, and organizational interoperability [2]. Semantic and syntactic interoperability can be ensured by communication exchange standards, such as the FHIR [68] standard of Health Level 7 (HL7) and medical terminologies. A suitable starting point for the basic procedures is offered by FHIR drills [69] or fire.ly [70].

Insight

FHIR is one of many communication standards but will likely change the domain of clinical IT significantly [71,72]. As a communication standard, FHIR harmonizes data formats coming from different CIS and enables data exchange between institutions via a RESTful approach [73]. Moreover, FHIR is used to connect devices with each other, which means, in particular, that the Integrating the Healthcare Enterprise (IHE) [74] standard has been revised to support HL7 messaging as well. In turn, IHE has been developing an open-source device

tool set for home and hospital use that recently enabled device control capabilities, a capability accelerated during the COVID-19 pandemic to allow nurses and physicians to operate ventilators and infusion devices outside the contaminated patient room [75].

Impact

Utilizing FHIR in multiple applications already shows its versatile and flexible use (eg, in mobile health applications [76], electrocardiogram monitoring [77], or wearable devices and precision medicine in digital health [72]). In particular, the SMART-on-FHIR technology enables third-party app development for health care applications [78] and encompasses feasible, secure, and time- and resource-efficient solutions [79,80].

Topic 7: OMOP—“Use Common Data Models as Well-defined Representations of Large-scale Research Projects.”

Introduction

Data harmonization enables research teams to run real-world observational studies based on heterogeneous data across country borders. Thus, harmonized data embedded in a common data model (CDM), which is an agreement about the utilization of standardized terminologies for data representation, is crucial to exchange data and results on a large scale. To foster reliability and trust in the results of observational research on real-world data, it is essential to utilize CDMs whenever possible to ensure a high degree of data analysis reproducibility.

Insight

Several CDMs exist for that purpose; the OMOP CDM from the OHDSI community is one of the most promising and established approaches. In comparison with other CDMs, such as the Sentinel CDM or Informatics for Integrating Biology and the Bedside (i2b2), the OMOP CDM has broader terminology coverage [81]. The importance of the OMOP CDM increased a lot over the last years [82], not least since the European Medicines Agency initiated the Data Analysis and Real World Interrogation Network (DARWIN) [83] project to establish a research network in Europe to gain real-world evidence based on OMOP. Moreover, representations of genomic data [84], oncology [85], and imaging projects [86] are also suitable. In addition, the common representation of the data in OMOP semantic interoperability is ensured by utilizing international terminologies and vocabularies, such as SNOMED-CT, the International Statistical Classification of Diseases and Related Health Problems (ICD), the Logical Observation Identifiers Names and Codes (LOINC), and RxNorm to represent every clinical fact in OMOP. Additionally, the open-source OHDSI software stack provides standardized methodology and libraries for data analyses (Athenahene, Atlas, HADES) and training (EHDEN Academy) [87], as well as a framework to assess and improve data quality to foster reliability and trust in research results [88].

Impact

The OMOP CDM is one possibility to represent and analyze clinical data on a research scale. Definition of new cohorts

within OMOP enables researchers to quickly investigate questions spanning multiple research entities. Collectively, both FHIR and OMOP can define the structure and relations of the clinical data corpus, and the individual EHRs provide content to these standardized data reservoirs. In comparison, OMOP is commonly used for static large-scale data analysis of research data, and FHIR is more suitable for rapid data integration scenarios (ie, for real-time applications and analysis). In summary, it is important to know and utilize newly established standards to participate in broader clinical networks for research. This way, all information within the EHR is comparable across different clinical sites and research settings.

Evaluation, Visualization, and Dissemination

Topic 8: Data Quality—“Guarantee High Quality and Then Publish the Data.”

Introduction

What is meant to be appropriate data quality for health informatics research? In this domain, data quality depends on the quality of single data elements, data completeness, data conformance, and data plausibility aspects that may considerably determine the validity and veracity of analysis results [89,90]. Moreover, data quality across different institutional entities and even health sectors requires additional efforts concerning the different personnel, instruments, and more [91]. High-quality data at hand is one fundamental requirement that is often difficult or impossible to achieve, which is why the generation of synthetic data can be an alternative that satisfies privacy problems as well as research needs when data are expensive, scarce, or unavailable by augmentation [92].

Insight

First, a major problem is that clinical data have to be electronically recorded, accessed, and standardized in order to run quality assessment processes [26]. In addition, it would be important to design and use the same data quality tool, standard operating procedures, or ETL mapping rules in all involved institutions. However, in real-life scenarios, there is a lack of both centrally coordinated data quality indicators and formalization of plausibility rules, as well as a repository for automatic querying of the rules, especially in ETL processes [93]. Although numerous data quality evaluation frameworks exist, no clear and widespread approach has been adopted so far [67,94-96]. Even after a well-chosen data quality procedure is properly implemented, clinical data as such cannot be published along with the performed study. As an alternative, synthetic data generation models function in the following 2 different ways: (1) The model is trained, for example, using real-world data and, once trained, will not require any data in the future (model-based approaches), and (2) the model is constantly fed with data to generate synthetic data (data-driven approaches). There are 3 different categories of algorithms used in the generation of synthetic data: probabilistic models, such as Bayesian networks [97] and Copulas [98]; ML, such as Classification and Regression Trees (CART); and deep learning methods, such as a generative adversarial network (GAN) [99-101] and variational autoencoder (VAE) [102].

Impact

A combination of appropriate data quality evaluation and synthetic data generation highly facilitates the development of accurate AI models, which are essential in medical studies [103]. Thus, a corpus of high-quality synthetic data with many patients can be reused by other AI experts for model development and benchmarking. Moreover, it is essential to create an infrastructure that is used across a large community of hospitals; maps the entire treatment process electronically; and only generates interoperable, structured data based on FHIR (Topic 6) and OMOP (Topic 7) in accordance with the FAIR principles (Topic 3). Afterward, one can finally run quality assessment processes.

Topic 9: Clinical Decision Support Systems—“Bring Insights, Not Additional Work, Back to the Clinics via a CDSS and Other User-Centric Applications.”

Introduction

CDSS are computer systems designed to assist the medical staff with decision-making tasks about individual patients and based on clinical data [104]. The decision-making process is still, and will remain, on the shoulders of the physician [105]. The categories of CDSS include knowledge-based systems that make use of clinical rules, nonknowledge-based systems (eg, AI-based systems), and hybrid CDSS that likewise utilize clinical models and knowledge in combination with AI.

Insight

The use of a CDSS in a well-implemented clinical workflow has many positive aspects. It may lead to fewer error rates [106], accelerate rare disease diagnosis [107], increase radiologists' job satisfaction [108], offer personalized cancer treatment [109], or help with real-time cardiovascular risk assessment [110]. Interestingly, computerized alerting systems, which are one of the most disseminated CDSS, can decrease drug-drug interactions significantly [111]. On the other hand, if done improperly, a CDSS can cause alert fatigue by creating too many alerts. If a system is not context-sensitive, alerts can even be inappropriate [112]. According to Olakotan et al [112], influencing factors of a well-designed CDSS need to include aspects about the (1) technology (eg, usability, alert presentation, workload, and data entry), (2) human (eg, training, knowledge, skills, attitude, and behavior), (3) organization (eg, rules and regulations, privacy, and security), and (4) and process (eg, waste, delay, tuning, and optimization). To avoid a lack of transparency and facilitate acceptance by physicians, especially with nonknowledge-based systems, current CDSS seek to use explainable AI approaches; however, the selection of methods used to present explanations in an informative and efficient (*clinically useful*) manner remains challenging [113]. Of note, a CDSS may also have a negative influence on the performance of physicians, especially if inadequate suggestions occur more often, which cannot be compensated with explanations [114]. However, one among many other prominent approaches to obtain such explanations via ML-based feature selection and ranking can be found in the work from Wolfien et al [115]. In terms of an OMOP-based implementation in research, there is patient-level prediction (PLP), which is designed to foster the

clinical decision-making process concerning diagnoses or treatment pathways based on the EHR of the patient and the current clinical guideline. It is used to answer questions, such as identifying patients among a larger population at higher risk of a certain outcome (eg, occurrence of cancer, severe side effects, or death) by using data in standardized formats (eg, as previously described via OMOP CDM). Once the model is designed, the covariates will be extracted from the respective CDM of the target person within the cohort, and the respective outcome will be predicted (eg, via PLP [116,117] or other customized prediction algorithms). Importantly, the results from model prediction should first be internally validated with previously unseen data and afterward compared with established scoring systems (eg, Framingham Risk Score [118], SCORE2 [119]) to connect with already known domain-specific contexts and to prove its benefit in clinical practice. An additional validation with external data, as part of a multicenter study, can be seen as highly beneficial, in which the already presented topics of federated learning (Topic 4) and OMOP (Topic 7) could significantly foster such an essential scenario [120].

Impact

Collectively, a CDSS increases patient safety, assists in clinical management, and can be cost-effective [104]. In general, findings of even erroneous CDSS can be used to guide the design of new CDSS alerts. However, the existing risks cannot be solved solely on a technical basis and require an interdisciplinary effort. In particular, continuous, clear communication between IT professionals (developers) and health professionals (end users) during the design process is key. Only a profound understanding of the needs and requirements of either of the involved parties can lead to well-designed systems that are actually able to support and relieve physicians in doing their job.

Topic 10: Visualizations—“Improved Dissemination of Local and External Data From Computational Models by Well-defined Interactive Visualizations.”

Introduction

Large volumes of data collected from patient registries, health centers, genomic databases, and public records can potentially improve the efficiency and quality of health care via enhancing the interoperability of medical systems, assisting in clinical decision-making, and delivering feedback on effective procedures [121]. However, each and every raw data point must go through different analytical processes until they become useful and interpretable at the point of care.

Insight

R and Python are 2 versatile open-source programming languages that have gained popularity for different purposes, such as preprocessing (eg, tidyverse), statistical tests (eg, dplyr), ML and deep learning (eg, mlr package, caret), visualization (eg, ggplot), and writing reports directly using knitr and R markdown (RStudio education [122]). Like R, Python offers different libraries for data science tasks (eg, open mined [123]) in addition to a library specifically for health predictive models, namely PyHealth [124]. Another versatile visualization functionality is offered for both languages via R Shiny [125]

and Plotly Dash [126]. These 2 platforms enable data scientists to create interactive web applications directly from a script. The applications can be extended using embedded CSS themes, HTML widgets, and Javascript actions. There is already evidence that implementing clinical dashboards or CDSS for immediate access to current patient information can improve processes and patient outcomes [127], especially if the data sets are further evaluated and refined [128]. Similar to FHIR, OHDSI provides tools for analyzing data in the OMOP CDM, which are written in R and use Shiny for the visualization. As a plus, data already stored in the OMOP CDM format can be used in systematic studies, patient-level analysis, and population-based estimations from scratch. The cBioPortal is one prime example of a web resource for exploring, visualizing, and analyzing multidimensional data, which reduces molecular profiling data from cancer tissues and cell lines into readily understandable genetic, epigenetic, gene expression, and proteomic events [129]. It was recently demonstrated how cBioPortal can be extended and integrated with other tools to a comprehensive and easily deployable software solution that supports the work of a molecular tumor board [130] and even deliver meaningful scientific insights [131]. Another translational research platform for the construction and integration of modern clinical research charts is Informatics for i2b2, which is also at the heart of clinical research [132,133].

Impact

Computational approaches and data analyses are tightly connected with medical research; the visualization of such complex data for clinicians in a routine setting especially plays a larger role. The current developments of translational research platforms, such as cBioPortal and i2b2, enable swift translation of research results into the clinic, if adequately adopted and enough trained people supervise the process.

Discussion

Overview

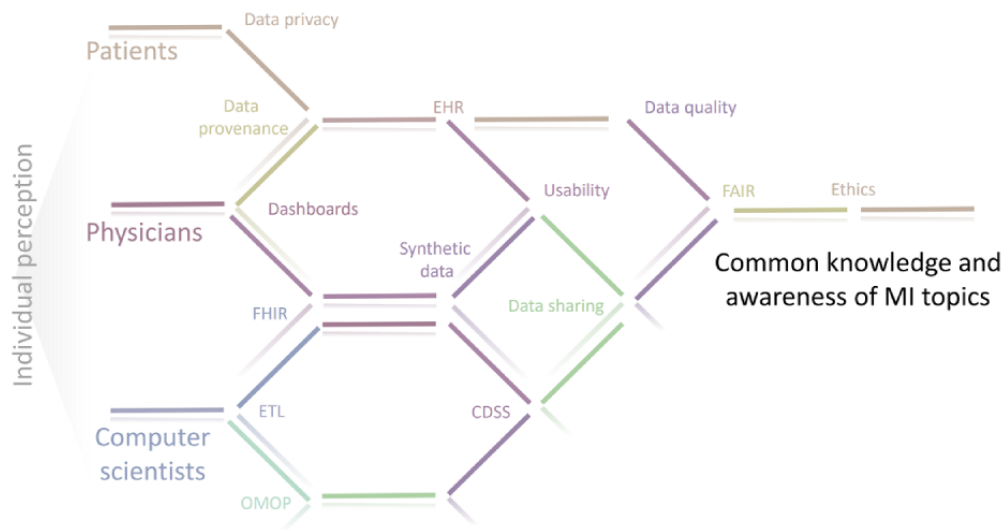
The need for qualified IT specialists in medical informatics has increased continuously in recent years and will continue to grow in the future. On the other hand, medical informatics in Germany

faces problems with the promotion of young researchers. These current developments mean that vacancies in IT in hospitals and the health care industry can often not be filled or only after very long vacancies. In addition, these positions often have to be filled with nonspecialist staff due to a lack of applications. To keep track of these recent developments and provide a basis for interdisciplinary communication, we provide our list of 10 topics that could be used by different stakeholders individually (Figure 2). With a particular focus in medicine, improved interdisciplinary communication has already been shown to positively impact patient outcomes and enhance employee engagement [134].

Furthermore, medical informatics has developed rapidly in recent years. This applies, for example, to new methods, techniques, tools, framework conditions, and organizational structures, especially in the field of medical data science. In particular, definitions of standards and a national digitized data corpus, namely the German Core Dataset [135], were agreed upon. The actual assessment and collection of digitized data in local university hospitals are utilized in so-called data integration centers. These interoperable research data infrastructures enable rapid multisite research, for example, with complex COVID-19 research data sets (German Corona Consensus Dataset [GECCO]) [136] including clinical data and data on biosamples from all German university hospitals in pseudonymized form (CODEX) [137,138] or the COVID-19 Data Portal [139]. The subsequent formation of the Network University Medicine (NUM) strengthens the existing interaction between research and patient care, stabilizes existing structures, and creates new structures that ensure more effective feedback and close cooperation between the clinics. The presented examples of NUM and CODEX, among others [140], attempt a central approach to bundle and harmonize necessary resources like broad consent or the elektronische Patientenakte (ePa), which is the implementation of EHR as a national entity to ultimately facilitate an interconnected health care system.

Finally, all those involved in medical informatics are called upon to engage in lifelong learning and continuously acquire further qualifications.

Figure 2. Exemplary outcome visualization of the underlying study, in which the color coding reflects the initial colors of the proposed sections; it starts with an individual perception of the term medical informatics (MI) based on the individual’s background and ends with acquisition of common domain knowledge for current important topics. CDSS: clinical decision support system; EHR: electronic health record; ETL; extract, transform, and load; FAIR; findable, accessible, interoperable, reusable; FHIR; Fast Healthcare Interoperability Resources; OMOP: Observational Medical Outcomes Partnership.



Exemplary Implementation of the Addressed Topics in the German Medical Informatics in Research and Care in University Medicine Consortium

This article offers newcomers to medical informatics a first introduction and a wealthy overview of current IT-related topics in research and patient care. Nevertheless, there is also a need for further qualification of employees through new, innovative offers for training, further education, and further training. As part of the MII [11], all consortia were asked to develop and set up appropriate offers and formats. The Medical Informatics in Research and Care in University Medicine (MIRACUM) consortium [141] has reacted and set up the part-time training and further education program “Biomedical Informatics and Data Science” [142] and introduced it at the Mannheim University of Applied Sciences in October 2020. The program includes a time-flexible and individually adaptable part-time online master’s course, as well as certificate courses and programs for further scientific education. In addition to the establishment and continuous further development of a

cloud-based learning platform, many new digital and target group-oriented learning resources and application-oriented learning environments were developed and introduced for the master's program.

All 10 topics listed in this article are reflected in the curriculum of the master’s degree and have been offered and dealt with in-depth in the individual courses for more than 2 years. The demand for the master’s program and certificate courses is high, and the evaluation has shown that these topic-specific foci correspond to the training and further education needs of the target groups. One particular aspect that was not covered in the final topics refers to the underlying infrastructure needed to provide the data storage and processing backbone. This aspect would have been too technical for a more broadly set, introductory article, such as this article. A starting point for more in-depth information about this aspect can be obtained from further literature [143,144]. However, to offer a practical start to the 10 topics, we provide links to well-known tutorials and hands-on materials (Table 1).

Table 1. Summary of tutorials and hands-on material about medical informatics standards and applications.

Topic number	Name	Description	Link
2	SNOMED CT ^a	This 5-step briefing presents a high-level overview of SNOMED CT, how it works, and the benefits of use.	[35]
4	DataSHIELD	This tutorial introduces users to DataSHIELD commands and syntax in R/R Studio.	[145]
5	ETL ^b	This provides introductory material to get from the native/raw data to the OMOP ^c CDM ^d one needs to create an ETL process.	[63]
6	FHIR ^e training	This contains a series of FHIR tutorials for those just beginning to learn the new specification.	[70]
6	SMART App Gallery	The SMART platform is composed of open-standard, open-source tools for developers building apps, and a publicly accessible gallery.	[78]
7	EHDEN Academy	This contains a series of tutorials for OMOP CDM and additional OHDSI ^f tools (eg, PLP ^g [117]).	[87]
8	Synthetic data generation	This is a hands-on tutorial from the ODI ^h [146] showing how to use Python to create synthetic data	[147]
10	R Studio education	This provides an introduction to basic R programming.	[122]
10	Python Dash	This tutorial helps develop data visualization interfaces.	[148]

^aSNOMED CT: Systematized Nomenclature of Medicine and Clinical Terms.

^bETL: extract, transform, and load.

^cOMOP: Observational Medical Outcomes Partnership.

^dCDM: common data model.

^eFHIR: Fast Healthcare Interoperability Resources.

^fOHDSI: Observational Health Data Sciences and Informatics.

^gPLP: patient-level prediction.

^hODI: Open Data Institute.

Conclusion

We suggest a set of 10 topics to ease the start for researchers and clinicians to become engaged with basic concepts in health informatics research. We provide current review articles for more in-depth reading about the specific topic and present practical hands-on material. The presented topics likewise serve as a broad overview of the medical informatics research domain but also guide individuals and their specific interests. For example, a computer scientist familiar with CDSS development could more easily connect with important aspects, such as data

privacy, FHIR, and specific EHRs that are highly relevant for daily work. In contrast, medical experts can obtain an overview of behind-the-scenes technologies, like ETL processes and underlying data quality approaches that are finally visualized as a summarizing clinical dashboard. For readers, we provided a first step toward an improved understanding of a lively and quickly expanding field, but more novel technologies and practical knowledge are ahead. Suggestions and contributions to improve the current topics can be made at GitHub, which will likewise enable content and readers to stay current [12].

Acknowledgments

This work was supported by the Federal Ministry of Health (BMG) and the German Federal Ministry of Education and Research (BMBF) within the Medical Informatics Initiative Medical Informatics in Research and Care in University Medicine (MIRACUM) Consortium (FKZ: 01ZZ180L [Dresden]; FKZ: 01ZZ180A [Erlangen]; FKZ: 01ZZ1801M [Greifswald]). The article processing charge was funded by the joint publication funds of the Technische Universität (TU) Dresden, including the Carl Gustav Carus Faculty of Medicine; Saxon State and University Library (SLUB) Dresden; and the Open Access Publication Funding of the German Research Foundation (DFG).

The funding sources had no involvement in the conduct of the research and preparation of the article.

Authors' Contributions

MW conceptualized the study, curated the data, and wrote the original manuscript draft. MW also defined the initial topics 1 and 2; MZ defined the initial topics 3 and 4; YP defined the initial topics 5 and 6; IR defined the initial topics 7 and 8; and NA defined the initial topics 8, 9, and 10. MS provided the resources and supervised the study. The topics were revised and extended by KF, AK, SG, DK, KLH, ICJ, CS, JS, TS, PS, and DW. MW, NA, YP, MZ, IR, and MS performed the formal analysis, and MW, NA,

and MS created the visualizations. NA, YP, MZ, IR, and MS wrote, reviewed, and edited the manuscript, and all authors read and agreed to the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Wyatt J, Liu JLY. Basic concepts in medical informatics. *J Epidemiol Community Health* 2002 Nov;56(11):808-812 [FREE Full text] [doi: [10.1136/jech.56.11.808](https://doi.org/10.1136/jech.56.11.808)] [Medline: [12388565](https://pubmed.ncbi.nlm.nih.gov/12388565/)]
2. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. *NPJ Digit Med* 2019 Aug 20;2(1):79 [FREE Full text] [doi: [10.1038/s41746-019-0158-1](https://doi.org/10.1038/s41746-019-0158-1)] [Medline: [31453374](https://pubmed.ncbi.nlm.nih.gov/31453374/)]
3. Beck S, Bergenholz C, Bogers M, Brasseur T, Conradsen ML, Di Marco D, et al. The Open Innovation in Science research field: a collaborative conceptualisation approach. *Industry and Innovation* 2020 Aug 04;29(2):136-185 [doi: [10.1080/13662716.2020.1792274](https://doi.org/10.1080/13662716.2020.1792274)]
4. MacLeod M. What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. *Synthese* 2016 Oct 7;195(2):697-720 [doi: [10.1007/s11229-016-1236-4](https://doi.org/10.1007/s11229-016-1236-4)]
5. Zajac S, Woods A, Tannenbaum S, Salas E, Holladay CL. Overcoming challenges to teamwork in healthcare: a team effectiveness framework and evidence-based guidance. *Front. Commun* 2021 Mar 17;6:1 [doi: [10.3389/fcomm.2021.606445](https://doi.org/10.3389/fcomm.2021.606445)]
6. Cunningham E, Smyth B, Greene D. Collaboration in the time of COVID: a scientometric analysis of multidisciplinary SARS-CoV-2 research. *Humanit Soc Sci Commun* 2021 Oct 19;8(1):1 [doi: [10.1057/s41599-021-00922-7](https://doi.org/10.1057/s41599-021-00922-7)]
7. Menezes S, Murray-Johnson K, Smith H, Trautmann H, Azizi M. Making science communication inclusive: an exploratory study of choices, challenges and change mechanisms in the United States from an emerging movement. *JCOM* 2022;21(05):1-22 [doi: [10.22323/2.21050203](https://doi.org/10.22323/2.21050203)]
8. Specht A, Crowston K. Interdisciplinary collaboration from diverse science teams can produce significant outcomes. *PLoS One* 2022 Nov 29;17(11):e0278043 [FREE Full text] [doi: [10.1371/journal.pone.0278043](https://doi.org/10.1371/journal.pone.0278043)] [Medline: [36445918](https://pubmed.ncbi.nlm.nih.gov/36445918/)]
9. Ten Simple Rules. *PLOS*. 2020 Apr 22. URL: <https://collections.plos.org/collection/ten-simple-rules/> [accessed 2023-07-01]
10. Kohane IS, Aronow BJ, Avillach P, Beaulieu-Jones BK, Bellazzi R, Bradford RL, Consortium For Clinical Characterization Of COVID-19 By EHR (4CE), et al. What every reader should know about studies using electronic health record data but may be afraid to ask. *J Med Internet Res* 2021 Mar 02;23(3):e22219 [FREE Full text] [doi: [10.2196/22219](https://doi.org/10.2196/22219)] [Medline: [33600347](https://pubmed.ncbi.nlm.nih.gov/33600347/)]
11. About the initiative. Medical Informatics Initiative. URL: <https://www.medizininformatik-initiative.de/en/about-initiative> [accessed 2023-07-01]
12. Ten Topics for Medical Informatics. GitHub. 2023 Mar 31. URL: https://github.com/mwolfien/ten_topics [accessed 2023-01-23]
13. EU data protection rules. European Commission. URL: https://commission.europa.eu/law/law-topic/data-protection/eu-data-protection-rules_en [accessed 2022-12-12]
14. Coppen R, van Veen E, Groenewegen P, Hazes J, de Jong J, Kievit J, et al. Will the trilogue on the EU Data Protection Regulation recognise the importance of health research? *Eur J Public Health* 2015 Oct 10;25(5):757-758 [FREE Full text] [doi: [10.1093/eurpub/ckv149](https://doi.org/10.1093/eurpub/ckv149)] [Medline: [26265364](https://pubmed.ncbi.nlm.nih.gov/26265364/)]
15. Lopes IM, Guarda T, Oliveira P. General Data Protection Regulation in health clinics. *J Med Syst* 2020 Jan 10;44(2):53 [doi: [10.1007/s10916-020-1521-0](https://doi.org/10.1007/s10916-020-1521-0)] [Medline: [31925566](https://pubmed.ncbi.nlm.nih.gov/31925566/)]
16. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics* 2021 Sep 15;22(1):122 [FREE Full text] [doi: [10.1186/s12910-021-00687-3](https://doi.org/10.1186/s12910-021-00687-3)] [Medline: [34525993](https://pubmed.ncbi.nlm.nih.gov/34525993/)]
17. Liaw S, Liyanage H, Kuziemyky C, Terry AL, Schreiber R, Jonnagaddala J, et al. Ethical use of electronic health record data and artificial intelligence: recommendations of the Primary Care Informatics Working Group of the International Medical Informatics Association. *Yearb Med Inform* 2020 Aug 17;29(1):51-57 [FREE Full text] [doi: [10.1055/s-0040-1701980](https://doi.org/10.1055/s-0040-1701980)] [Medline: [32303098](https://pubmed.ncbi.nlm.nih.gov/32303098/)]
18. Meurers T, Bild R, Do KM, Prasser F. A scalable software solution for anonymizing high-dimensional biomedical data. *Gigascience* 2021 Oct 04;10(10):1 [FREE Full text] [doi: [10.1093/gigascience/giab068](https://doi.org/10.1093/gigascience/giab068)] [Medline: [34605868](https://pubmed.ncbi.nlm.nih.gov/34605868/)]
19. Zuo Z, Watson M, Budgen D, Hall R, Kennelly C, Al Moubayed N. Data anonymization for pervasive health care: systematic literature mapping study. *JMIR Med Inform* 2021 Oct 15;9(10):e29871 [FREE Full text] [doi: [10.2196/29871](https://doi.org/10.2196/29871)] [Medline: [34652278](https://pubmed.ncbi.nlm.nih.gov/34652278/)]
20. Chen RJ, Lu MY, Chen TY, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. *Nat Biomed Eng* 2021 Jun 15;5(6):493-497 [FREE Full text] [doi: [10.1038/s41551-021-00751-8](https://doi.org/10.1038/s41551-021-00751-8)] [Medline: [34131324](https://pubmed.ncbi.nlm.nih.gov/34131324/)]
21. Vayena E, Blasimme A. Health research with big data: time for systemic oversight. *J Law Med Ethics* 2018 Mar 27;46(1):119-129 [FREE Full text] [doi: [10.1177/1073110518766026](https://doi.org/10.1177/1073110518766026)] [Medline: [30034208](https://pubmed.ncbi.nlm.nih.gov/30034208/)]
22. McLennan S, Fiske A, Tigard D, Müller R, Haddadin S, Buyx A. Embedded ethics: a proposal for integrating ethics into the development of medical AI. *BMC Med Ethics* 2022 Jan 26;23(1):6 [FREE Full text] [doi: [10.1186/s12910-022-00746-3](https://doi.org/10.1186/s12910-022-00746-3)] [Medline: [35081955](https://pubmed.ncbi.nlm.nih.gov/35081955/)]

23. Bild R, Bialke M, Buckow K, Ganslandt T, Ihrig K, Jahns R, et al. Towards a comprehensive and interoperable representation of consent-based data usage permissions in the German medical informatics initiative. *BMC Med Inform Decis Mak* 2020 Jun 05;20(1):103 [FREE Full text] [doi: [10.1186/s12911-020-01138-6](https://doi.org/10.1186/s12911-020-01138-6)] [Medline: [32503529](https://pubmed.ncbi.nlm.nih.gov/32503529/)]
24. Stamm T, Bott N, Thwaites R, Mosor E, Andrews MR, Borgdorff J, et al. Building a value-based care infrastructure in Europe: the Health Outcomes Observatory. *NEJM Catalyst* 2021 Jun 09;1-14 [doi: [10.1056/CAT.21.0146](https://doi.org/10.1056/CAT.21.0146)]
25. Househ M, Grainger R, Petersen C, Bamidis P, Merolli M. Balancing between privacy and patient needs for health information in the age of participatory health and social media: a scoping review. *Yearb Med Inform* 2018 Aug 22;27(1):29-36 [FREE Full text] [doi: [10.1055/s-0038-1641197](https://doi.org/10.1055/s-0038-1641197)] [Medline: [29681040](https://pubmed.ncbi.nlm.nih.gov/29681040/)]
26. Fitzer K, Haeuslschmid R, Blasini R, Altun FB, Hampf C, Freiesleben S, et al. Patient recruitment system for clinical trials: mixed methods study about requirements at ten university hospitals. *JMIR Med Inform* 2022 Apr 20;10(4):e28696 [FREE Full text] [doi: [10.2196/28696](https://doi.org/10.2196/28696)] [Medline: [35442203](https://pubmed.ncbi.nlm.nih.gov/35442203/)]
27. Vehko T, Hyppönen H, Puttonen S, Kujala S, Ketola E, Tuukkanen J, et al. Experienced time pressure and stress: electronic health records usability and information technology competence play a role. *BMC Med Inform Decis Mak* 2019 Aug 14;19(1):160 [FREE Full text] [doi: [10.1186/s12911-019-0891-z](https://doi.org/10.1186/s12911-019-0891-z)] [Medline: [31412859](https://pubmed.ncbi.nlm.nih.gov/31412859/)]
28. OpenNotes. URL: <https://www.opennotes.org/> [accessed 2023-07-01]
29. Grote-Westrick M. New German digital project paves the way for online access to personal electronic health records. *The BMJ Opinion*. 2021 Feb 18. URL: <https://blogs.bmj.com/bmj/2021/02/18/new-german-digital-project-paves-the-way-for-online-access-to-personal-electronic-health-records/> [accessed 2023-07-01]
30. Häyrynen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform* 2008 May;77(5):291-304 [doi: [10.1016/j.ijmedinf.2007.09.001](https://doi.org/10.1016/j.ijmedinf.2007.09.001)] [Medline: [17951106](https://pubmed.ncbi.nlm.nih.gov/17951106/)]
31. Brender J, Nøhr C, McNair P. Research needs and priorities in health informatics. *Int J Med Inform* 2000 Sep;58-59:257-289 [doi: [10.1016/s1386-5056\(00\)00092-7](https://doi.org/10.1016/s1386-5056(00)00092-7)] [Medline: [10978926](https://pubmed.ncbi.nlm.nih.gov/10978926/)]
32. Haux R, Ammenwerth E, Herzog W, Knaup P. Health care in the information society. A prognosis for the year 2013. *Int J Med Inform* 2002 Nov 20;66(1-3):3-21 [doi: [10.1016/s1386-5056\(02\)00030-8](https://doi.org/10.1016/s1386-5056(02)00030-8)] [Medline: [12453552](https://pubmed.ncbi.nlm.nih.gov/12453552/)]
33. de Hoop T, Neumuth T. Evaluating electronic health record limitations and time expenditure in a German medical center. *Appl Clin Inform* 2021 Oct 22;12(5):1082-1090 [FREE Full text] [doi: [10.1055/s-0041-1739519](https://doi.org/10.1055/s-0041-1739519)] [Medline: [34937102](https://pubmed.ncbi.nlm.nih.gov/34937102/)]
34. Millar J. The need for a global language - SNOMED CT introduction. *Studies in Health Technology and Informatics* 2016;225:683-685 [doi: [10.3233/978-1-61499-658-3-683](https://doi.org/10.3233/978-1-61499-658-3-683)]
35. 5-Step briefing. SNOMED International. URL: <https://www.snomed.org/snomed-ct/five-step-briefing> [accessed 2023-07-01]
36. Gaudet-Blavignac C, Foufi V, Bjelogrić M, Lovis C. Use of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) for processing free text in health care: systematic scoping review. *J Med Internet Res* 2021 Jan 26;23(1):e24594 [FREE Full text] [doi: [10.2196/24594](https://doi.org/10.2196/24594)] [Medline: [33496673](https://pubmed.ncbi.nlm.nih.gov/33496673/)]
37. Schaaf J, Sedlmayr M, Sedlmayr B, Prokosch H, Storf H. Evaluation of a clinical decision support system for rare diseases: a qualitative study. *BMC Med Inform Decis Mak* 2021 Feb 18;21(1):65 [FREE Full text] [doi: [10.1186/s12911-021-01435-8](https://doi.org/10.1186/s12911-021-01435-8)] [Medline: [33602191](https://pubmed.ncbi.nlm.nih.gov/33602191/)]
38. Lichtner G, Spies C, Jurth C, Bienert T, Mueller A, Kumpf O, et al. Automated monitoring of adherence to evidenced-based clinical guideline recommendations: design and implementation study. *J Med Internet Res* 2023 May 04;25:e41177 [FREE Full text] [doi: [10.2196/41177](https://doi.org/10.2196/41177)] [Medline: [36996044](https://pubmed.ncbi.nlm.nih.gov/36996044/)]
39. Bernard E, Tuechler H, Greenberg PL, Hasserjian RP, Arango Ossa JE, Nannya Y, et al. Molecular International Prognostic scoring system for myelodysplastic syndromes. *NEJM Evidence* 2022 Jun 28;1(7):1 [doi: [10.1056/evidoa2200008](https://doi.org/10.1056/evidoa2200008)]
40. Klein TM, Augustin M, Otten M. How should electronic health records be designed? A cross-sectional study in patients with psoriasis. *BMC Med Inform Decis Mak* 2019 Nov 12;19(1):218 [FREE Full text] [doi: [10.1186/s12911-019-0926-5](https://doi.org/10.1186/s12911-019-0926-5)] [Medline: [31718653](https://pubmed.ncbi.nlm.nih.gov/31718653/)]
41. Mayer G, Müller W, Schork K, Uszkoreit J, Weidemann A, Wittig U, et al. Implementing FAIR data management within the German Network for Bioinformatics Infrastructure (de.NBI) exemplified by selected use cases. *Brief Bioinform* 2021 Sep 02;22(5):1 [FREE Full text] [doi: [10.1093/bib/bbab010](https://doi.org/10.1093/bib/bbab010)] [Medline: [33589928](https://pubmed.ncbi.nlm.nih.gov/33589928/)]
42. Palmblad M, Lamprecht AL, Ison J, Schwämmle V. Automated workflow composition in mass spectrometry-based proteomics. *Bioinformatics* 2019 Feb 15;35(4):656-664 [FREE Full text] [doi: [10.1093/bioinformatics/bty646](https://doi.org/10.1093/bioinformatics/bty646)] [Medline: [30060113](https://pubmed.ncbi.nlm.nih.gov/30060113/)]
43. Xu S, Rogers T, Fairweather E, Glenn A, Curran J, Curcin V. Application of data provenance in healthcare analytics software: information visualisation of user activities. *AMIA Jt Summits Transl Sci Proc* 2018;2017:263-272 [FREE Full text] [Medline: [29888084](https://pubmed.ncbi.nlm.nih.gov/29888084/)]
44. Curcin V. Embedding data provenance into the learning health system to facilitate reproducible research. *Learn Health Syst* 2017 Apr 27;1(2):e10019 [FREE Full text] [doi: [10.1002/lrh2.10019](https://doi.org/10.1002/lrh2.10019)] [Medline: [31245557](https://pubmed.ncbi.nlm.nih.gov/31245557/)]
45. Inau ET, Sack J, Waltemath D, Zeleke AA. Initiatives, concepts, and implementation practices of FAIR (Findable, Accessible, Interoperable, and Reusable) data principles in health data stewardship practice: protocol for a scoping review. *JMIR Res Protoc* 2021 Feb 02;10(2):e22505 [FREE Full text] [doi: [10.2196/22505](https://doi.org/10.2196/22505)] [Medline: [33528373](https://pubmed.ncbi.nlm.nih.gov/33528373/)]

46. Queralt-Rosinach N, Kaliyaperumal R, Bernabé CH, Long Q, Joosten SA, van der Wijk HJ, BEAT-COVID Group, COVID-19 LUMC Group. Applying the FAIR principles to data in a hospital: challenges and opportunities in a pandemic. *J Biomed Semantics* 2022 Apr 25;13(1):12 [FREE Full text] [doi: [10.1186/s13326-022-00263-7](https://doi.org/10.1186/s13326-022-00263-7)] [Medline: [35468846](https://pubmed.ncbi.nlm.nih.gov/35468846/)]
47. Frexia F, Mascia C, Lianas L, Delussu G, Sulis A, Meloni V, et al. openEHR Is FAIR-enabling by design. *Stud Health Technol Inform* 2021 May 27;281:113-117 [doi: [10.3233/SHTI210131](https://doi.org/10.3233/SHTI210131)] [Medline: [34042716](https://pubmed.ncbi.nlm.nih.gov/34042716/)]
48. 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 Mar 15;3(1):160018 [FREE Full text] [doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)] [Medline: [26978244](https://pubmed.ncbi.nlm.nih.gov/26978244/)]
49. Holub P, Kohlmayer F, Prasser F, Mayrhofer MT, Schlünder I, Martin GM, et al. Enhancing reuse of data and biological material in medical research: from FAIR to FAIR-Health. *Biopreserv Biobank* 2018 Apr;16(2):97-105 [FREE Full text] [doi: [10.1089/bio.2017.0110](https://doi.org/10.1089/bio.2017.0110)] [Medline: [29359962](https://pubmed.ncbi.nlm.nih.gov/29359962/)]
50. Gierend K, Krüger F, Waltemath D, Fünfgeld M, Ganslandt T, Zeleke AA. Approaches and criteria for provenance in biomedical data sets and workflows: protocol for a scoping review. *JMIR Res Protoc* 2021 Nov 22;10(11):e31750 [FREE Full text] [doi: [10.2196/31750](https://doi.org/10.2196/31750)] [Medline: [34813494](https://pubmed.ncbi.nlm.nih.gov/34813494/)]
51. Scheibner J, Raisaro JL, Troncoso-Pastoriza JR, Ienca M, Fellay J, Vayena E, et al. Revolutionizing medical data sharing using advanced privacy-enhancing technologies: technical, legal, and ethical synthesis. *J Med Internet Res* 2021 Feb 25;23(2):e25120 [FREE Full text] [doi: [10.2196/25120](https://doi.org/10.2196/25120)] [Medline: [33629963](https://pubmed.ncbi.nlm.nih.gov/33629963/)]
52. Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. Federated learning of predictive models from federated Electronic Health Records. *Int J Med Inform* 2018 Apr;112:59-67 [FREE Full text] [doi: [10.1016/j.ijmedinf.2018.01.007](https://doi.org/10.1016/j.ijmedinf.2018.01.007)] [Medline: [29500022](https://pubmed.ncbi.nlm.nih.gov/29500022/)]
53. Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, Jones EM, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. *Int J Epidemiol* 2014 Dec;43(6):1929-1944 [FREE Full text] [doi: [10.1093/ije/dyu188](https://doi.org/10.1093/ije/dyu188)] [Medline: [25261970](https://pubmed.ncbi.nlm.nih.gov/25261970/)]
54. Marcon Y, Bishop T, Avraam D, Escriba-Montagut X, Ryser-Welch P, Wheeler S, et al. Orchestrating privacy-protected big data analyses of data from different resources with R and DataSHIELD. *PLoS Comput Biol* 2021 Mar 30;17(3):e1008880 [FREE Full text] [doi: [10.1371/journal.pcbi.1008880](https://doi.org/10.1371/journal.pcbi.1008880)] [Medline: [33784300](https://pubmed.ncbi.nlm.nih.gov/33784300/)]
55. Lenz S, Hess M, Binder H. Deep generative models in DataSHIELD. *BMC Med Res Methodol* 2021 Apr 03;21(1):64 [FREE Full text] [doi: [10.1186/s12874-021-01237-6](https://doi.org/10.1186/s12874-021-01237-6)] [Medline: [33812380](https://pubmed.ncbi.nlm.nih.gov/33812380/)]
56. Beyan O, Choudhury A, van Soest J, Kohlbacher O, Zimmermann L, Stenzhorn H, et al. Distributed analytics on sensitive medical data: the personal health train. *Data Intelligence* 2020 Jan;2(1-2):96-107 [FREE Full text] [doi: [10.1162/dint_a_00032](https://doi.org/10.1162/dint_a_00032)]
57. Dong X, Randolph DA, Weng C, Kho AN, Rogers JM, Wang X. Developing high performance Secure Multi-Party Computation protocols in healthcare: a case study of patient risk stratification. *AMIA Jt Summits Transl Sci Proc* 2021;2021:200-209 [FREE Full text] [Medline: [34457134](https://pubmed.ncbi.nlm.nih.gov/34457134/)]
58. Vogelsang L, Lehne M, Schoppmann P, Prasser F, Thun S, Scheuermann B, et al. A Secure Multi-Party Computation protocol for time-to-event analyses. *Stud Health Technol Inform* 2020 Jun 16;270:8-12 [doi: [10.3233/SHTI200112](https://doi.org/10.3233/SHTI200112)] [Medline: [32570336](https://pubmed.ncbi.nlm.nih.gov/32570336/)]
59. Zhou J, Feng Y, Wang Z, Guo D. Using Secure Multi-Party Computation to protect privacy on a permissioned blockchain. *Sensors (Basel)* 2021 Feb 23;21(4):1540 [FREE Full text] [doi: [10.3390/s21041540](https://doi.org/10.3390/s21041540)] [Medline: [33672175](https://pubmed.ncbi.nlm.nih.gov/33672175/)]
60. Negro-Calduch E, Azzopardi-Muscat N, Krishnamurthy RS, Novillo-Ortiz D. Technological progress in electronic health record system optimization: Systematic review of systematic literature reviews. *Int J Med Inform* 2021 Aug;152:104507 [FREE Full text] [doi: [10.1016/j.ijmedinf.2021.104507](https://doi.org/10.1016/j.ijmedinf.2021.104507)] [Medline: [34049051](https://pubmed.ncbi.nlm.nih.gov/34049051/)]
61. Prasser F, Spengler H, Bild R, Eicher J, Kuhn KA. Privacy-enhancing ETL-processes for biomedical data. *Int J Med Inform* 2019 Jun;126:72-81 [FREE Full text] [doi: [10.1016/j.ijmedinf.2019.03.006](https://doi.org/10.1016/j.ijmedinf.2019.03.006)] [Medline: [31029266](https://pubmed.ncbi.nlm.nih.gov/31029266/)]
62. Helgheim BI, Maia R, Ferreira JC, Martins AL. Merging data diversity of clinical medical records to improve effectiveness. *Int J Environ Res Public Health* 2019 Mar 03;16(5):769 [FREE Full text] [doi: [10.3390/ijerph16050769](https://doi.org/10.3390/ijerph16050769)] [Medline: [30832447](https://pubmed.ncbi.nlm.nih.gov/30832447/)]
63. Blacketer C, Voss E. Chapter 6 Extract Transform Load. *The Book of OHDSI*. 2021 Jan 11. URL: <https://ohdsi.github.io/TheBookOfOhdsi/ExtractTransformLoad.html> [accessed 2023-07-01]
64. Guo G, Jonnagaddala J, Farshid S, Huser V, Reich C, Liaw ST. Comparison of the cohort selection performance of Australian Medicines Terminology to Anatomical Therapeutic Chemical mappings. *J Am Med Inform Assoc* 2019 Nov 01;26(11):1237-1246 [FREE Full text] [doi: [10.1093/jamia/ocz143](https://doi.org/10.1093/jamia/ocz143)] [Medline: [31545380](https://pubmed.ncbi.nlm.nih.gov/31545380/)]
65. Peng Y, Henke E, Reinecke I, Zoch M, Sedlmayr M, Bathelt F. An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM. *Int J Med Inform* 2023 Jan;169:104925 [FREE Full text] [doi: [10.1016/j.ijmedinf.2022.104925](https://doi.org/10.1016/j.ijmedinf.2022.104925)] [Medline: [36395615](https://pubmed.ncbi.nlm.nih.gov/36395615/)]
66. Bauer CRKD, Ganslandt T, Baum B, Christoph J, Engel I, Löbe M, et al. Integrated Data Repository Toolkit (IDRT). *Methods Inf Med* 2018 Jan 08;55(02):125-135 [doi: [10.3414/me15-01-0082](https://doi.org/10.3414/me15-01-0082)]
67. Liaw S, Guo JGN, Ansari S, Jonnagaddala J, Godinho MA, Borelli AJ, et al. Quality assessment of real-world data repositories across the data life cycle: A literature review. *J Am Med Inform Assoc* 2021 Jul 14;28(7):1591-1599 [FREE Full text] [doi: [10.1093/jamia/ocaa340](https://doi.org/10.1093/jamia/ocaa340)] [Medline: [33496785](https://pubmed.ncbi.nlm.nih.gov/33496785/)]
68. HL7 FHIR - Release 5. URL: <https://www.hl7.org/fhir/> [accessed 2023-07-01]

69. FHIR Drills. URL: <https://fhir-drills.github.io/index.html> [accessed 2023-07-01]
70. FHIR Training Courses and Workshops for Health IT Professionals. fire.ly. URL: <https://fire.ly/training/> [accessed 2023-07-01]
71. Andersen B, Kasparick M, Ulrich H, Franke S, Schlamelcher J, Rockstroh M, et al. Connecting the clinical IT infrastructure to a service-oriented architecture of medical devices. *Biomed Tech (Berl)* 2018 Feb 23;63(1):57-68 [doi: [10.1515/bmt-2017-0021](https://doi.org/10.1515/bmt-2017-0021)] [Medline: [29272252](https://pubmed.ncbi.nlm.nih.gov/29272252/)]
72. Lehne M, Luijten S, Vom Felde Genannt Imbusch P, Thun S. The use of FHIR in digital health - a review of the scientific literature. *Stud Health Technol Inform* 2019 Sep 03;267:52-58 [doi: [10.3233/SHTI190805](https://doi.org/10.3233/SHTI190805)] [Medline: [31483254](https://pubmed.ncbi.nlm.nih.gov/31483254/)]
73. Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems* 2013:1 [doi: [10.1109/CBMS.2013.6627810](https://doi.org/10.1109/CBMS.2013.6627810)]
74. IHE International. URL: <https://www.ihe.net/> [accessed 2023-07-01]
75. Sloane EB, Cooper T, Silva R. MDIRA: IEEE, IHE, and FHIR Clinical Device and Information Technology Interoperability Standards, bridging Home to Hospital to "Hospital-in-Home". *SoutheastCon 2021* 2021:1 [doi: [10.1109/SoutheastCon45413.2021.9401934](https://doi.org/10.1109/SoutheastCon45413.2021.9401934)]
76. Lamprinakos G, Mousas AS, Kapsalis AP, Kaklamani DI, Venieris IS, Boufifis AD, et al. Using FHIR to develop a healthcare mobile application. 2014 Presented at: 4th International Conference on Wireless Mobile Communication and Healthcare; November 3–5, 2014; Athens, Greece [doi: [10.4108/icst.mobihealth.2014.257232](https://doi.org/10.4108/icst.mobihealth.2014.257232)]
77. Benhamida A, Kanas A, Vincze M, Papp KT, Abbassi M, Kozlovsky M. SaECG: a new FHIR Data format revision to enable continuous ECG storage and monitoring. *2020 IEEE 20th International Symposium on Computational Intelligence and Informatics (CINTI)* 2021:1 [doi: [10.1109/CINTI51262.2020.9305828](https://doi.org/10.1109/CINTI51262.2020.9305828)]
78. SMART App Gallery. URL: <https://apps.smarthealthit.org/apps/featured> [accessed 2023-07-01]
79. 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 Sep;23(5):899-908 [FREE Full text] [doi: [10.1093/jamia/ocv189](https://doi.org/10.1093/jamia/ocv189)] [Medline: [26911829](https://pubmed.ncbi.nlm.nih.gov/26911829/)]
80. Wesley D, Blumenthal J, Shah S, Littlejohn RA, Pruitt Z, Dixit R, et al. A novel application of SMART on FHIR architecture for interoperable and scalable integration of patient-reported outcome data with electronic health records. *J Am Med Inform Assoc* 2021 Sep 18;28(10):2220-2225 [FREE Full text] [doi: [10.1093/jamia/ocab110](https://doi.org/10.1093/jamia/ocab110)] [Medline: [34279660](https://pubmed.ncbi.nlm.nih.gov/34279660/)]
81. Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform* 2016 Dec;64:333-341 [FREE Full text] [doi: [10.1016/j.jbi.2016.10.016](https://doi.org/10.1016/j.jbi.2016.10.016)] [Medline: [27989817](https://pubmed.ncbi.nlm.nih.gov/27989817/)]
82. Reinecke I, Zoch M, Reich C, Sedlmayr M, Bathelt F. The usage of OHDSI OMOP - a scoping review. *Stud Health Technol Inform* 2021 Sep 21;283:95-103 [doi: [10.3233/SHTI210546](https://doi.org/10.3233/SHTI210546)] [Medline: [34545824](https://pubmed.ncbi.nlm.nih.gov/34545824/)]
83. Data Analysis and Real World Interrogation Network (DARWIN EU). European Medicines Agency. URL: <https://www.ema.europa.eu/en/about-us/how-we-work/big-data/data-analysis-real-world-interrogation-network-darwin-eu> [accessed 2023-07-01]
84. Peng Y, Nassirian A, Ahmadi N, Sedlmayr M, Bathelt F. Towards the representation of genomic data in HL7 FHIR and OMOP CDM. *Stud Health Technol Inform* 2021 Sep 21;283:86-94 [doi: [10.3233/SHTI210545](https://doi.org/10.3233/SHTI210545)] [Medline: [34545823](https://pubmed.ncbi.nlm.nih.gov/34545823/)]
85. Ahmadi N, Peng Y, Wolfien M, Zoch M, Sedlmayr M. OMOP CDM can facilitate data-driven studies for cancer prediction: a systematic review. *Int J Mol Sci* 2022 Oct 05;23(19):11834 [FREE Full text] [doi: [10.3390/ijms231911834](https://doi.org/10.3390/ijms231911834)] [Medline: [36233137](https://pubmed.ncbi.nlm.nih.gov/36233137/)]
86. Park C, You SC, Jeon H, Jeong CW, Choi JW, Park RW. Development and validation of the Radiology Common Data Model (R-CDM) for the international standardization of medical imaging data. *Yonsei Med J* 2022;63(Suppl):S74 [doi: [10.3349/ymj.2022.63.s74](https://doi.org/10.3349/ymj.2022.63.s74)]
87. EHDEN Academy. URL: <https://academy.ehden.eu/> [accessed 2023-07-01]
88. Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR. Increasing trust in real-world evidence through evaluation of observational data quality. *J Am Med Inform Assoc* 2021 Sep 18;28(10):2251-2257 [FREE Full text] [doi: [10.1093/jamia/ocab132](https://doi.org/10.1093/jamia/ocab132)] [Medline: [34313749](https://pubmed.ncbi.nlm.nih.gov/34313749/)]
89. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *EGEMS (Wash DC)* 2016 Sep 11;4(1):1244 [FREE Full text] [doi: [10.13063/2327-9214.1244](https://doi.org/10.13063/2327-9214.1244)] [Medline: [27713905](https://pubmed.ncbi.nlm.nih.gov/27713905/)]
90. Kamdje-Wabo G, Gradinger T, Löbe M, Lodahl R, Seuchter SA, Sax U, et al. Towards structured data quality assessment in the German Medical Informatics Initiative: initial approach in the MII demonstrator study. *Stud Health Technol Inform* 2019 Aug 21;264:1508-1509 [doi: [10.3233/SHTI190508](https://doi.org/10.3233/SHTI190508)] [Medline: [31438205](https://pubmed.ncbi.nlm.nih.gov/31438205/)]
91. Tayefi M, Ngo P, Chomutare T, Dalianis H, Salvi E, Budrionis A, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. *WIREs Comp Stat* 2021 Feb 14;13(6):1 [doi: [10.1002/wics.1549](https://doi.org/10.1002/wics.1549)]
92. Goncalves A, Ray P, Soper B, Stevens J, Coyle L, Sales AP. Generation and evaluation of synthetic patient data. *BMC Med Res Methodol* 2020 May 07;20(1):108 [FREE Full text] [doi: [10.1186/s12874-020-00977-1](https://doi.org/10.1186/s12874-020-00977-1)] [Medline: [32381039](https://pubmed.ncbi.nlm.nih.gov/32381039/)]

93. Löbe M, Kamdje-Wabo G, Sinza AC, Spengler H, Strobel M, Tute E. Towards harmonized data quality in the Medical Informatics Initiative - current state and future directions. *Stud Health Technol Inform* 2022 Jan 14;289:240-243 [doi: [10.3233/SHTI210904](https://doi.org/10.3233/SHTI210904)] [Medline: [35062137](https://pubmed.ncbi.nlm.nih.gov/35062137/)]
94. Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems* 2015 Dec 11;12(4):5-33 [doi: [10.1080/07421222.1996.11518099](https://doi.org/10.1080/07421222.1996.11518099)]
95. Wahyudi A, Kuk G, Janssen M. A process pattern model for tackling and improving big data quality. *Inf Syst Front* 2018 Jan 25;20(3):457-469 [doi: [10.1007/s10796-017-9822-7](https://doi.org/10.1007/s10796-017-9822-7)]
96. Schmidt CO, Struckmann S, Enzenbach C, Reineke A, Stausberg J, Damerow S, et al. Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R. *BMC Med Res Methodol* 2021 Apr 02;21(1):63 [FREE Full text] [doi: [10.1186/s12874-021-01252-7](https://doi.org/10.1186/s12874-021-01252-7)] [Medline: [33810787](https://pubmed.ncbi.nlm.nih.gov/33810787/)]
97. Kaur D, Sobieski M, Patil S, Liu J, Bhagat P, Gupta A, et al. Application of Bayesian networks to generate synthetic health data. *J Am Med Inform Assoc* 2021 Mar 18;28(4):801-811 [FREE Full text] [doi: [10.1093/jamia/ocaa303](https://doi.org/10.1093/jamia/ocaa303)] [Medline: [33367620](https://pubmed.ncbi.nlm.nih.gov/33367620/)]
98. Kamthe S, Assefa S, Deisenroth M. Copula flows for synthetic data generation. *arXiv Preprint* posted online January 3, 2021. [doi: [10.48550/arXiv.2101.00598](https://doi.org/10.48550/arXiv.2101.00598)]
99. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. *Communications of the ACM* 2020 Oct 22;63(11):139-144 [doi: [10.1145/3422622](https://doi.org/10.1145/3422622)]
100. Walonoski J, Kramer M, Nichols J, Quina A, Moesel C, Hall D, et al. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *J Am Med Inform Assoc* 2018 Mar 01;25(3):230-238 [FREE Full text] [doi: [10.1093/jamia/ocx079](https://doi.org/10.1093/jamia/ocx079)] [Medline: [29025144](https://pubmed.ncbi.nlm.nih.gov/29025144/)]
101. Baowaly M, Lin CC, Liu CL, Chen KT. Synthesizing electronic health records using improved generative adversarial networks. *J Am Med Inform Assoc* 2019 Mar 01;26(3):228-241 [FREE Full text] [doi: [10.1093/jamia/ocy142](https://doi.org/10.1093/jamia/ocy142)] [Medline: [30535151](https://pubmed.ncbi.nlm.nih.gov/30535151/)]
102. Wan Z, Zhang Y, He H. Variational autoencoder based synthetic data generation for imbalanced learning. 2017 IEEE Symposium Series on Computational Intelligence (SSCI) 2017:1 [doi: [10.1109/SSCI.2017.8285168](https://doi.org/10.1109/SSCI.2017.8285168)]
103. Hahn W, Schütte K, Schultz K, Wolkenhauer O, Sedlmayr M, Schuler U, et al. Contribution of synthetic data generation towards an improved patient stratification in palliative care. *J Pers Med* 2022 Aug 04;12(8):1278 [FREE Full text] [doi: [10.3390/jpm12081278](https://doi.org/10.3390/jpm12081278)] [Medline: [36013227](https://pubmed.ncbi.nlm.nih.gov/36013227/)]
104. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med* 2020;3:17 [FREE Full text] [doi: [10.1038/s41746-020-0221-y](https://doi.org/10.1038/s41746-020-0221-y)] [Medline: [32047862](https://pubmed.ncbi.nlm.nih.gov/32047862/)]
105. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med* 2022 Jan 20;28(1):31-38 [doi: [10.1038/s41591-021-01614-0](https://doi.org/10.1038/s41591-021-01614-0)] [Medline: [35058619](https://pubmed.ncbi.nlm.nih.gov/35058619/)]
106. Shen L, Wright A, Lee LS, Jajoo K, Nayor J, Landman A. Clinical decision support system, using expert consensus-derived logic and natural language processing, decreased sedation-type order errors for patients undergoing endoscopy. *J Am Med Inform Assoc* 2021 Jan 15;28(1):95-103 [FREE Full text] [doi: [10.1093/jamia/ocaa250](https://doi.org/10.1093/jamia/ocaa250)] [Medline: [33175157](https://pubmed.ncbi.nlm.nih.gov/33175157/)]
107. Ronicke S, Hirsch MC, Türk E, Larionov K, Tientcheu D, Wagner AD. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. *Orphanet J Rare Dis* 2019 Mar 21;14(1):69 [FREE Full text] [doi: [10.1186/s13023-019-1040-6](https://doi.org/10.1186/s13023-019-1040-6)] [Medline: [30898118](https://pubmed.ncbi.nlm.nih.gov/30898118/)]
108. Eltorai AEM, Bratt AK, Guo HH. Thoracic radiologists' versus computer scientists' perspectives on the future of artificial intelligence in radiology. *J Thorac Imaging* 2020 Jul;35(4):255-259 [doi: [10.1097/RTL.0000000000000453](https://doi.org/10.1097/RTL.0000000000000453)] [Medline: [31609778](https://pubmed.ncbi.nlm.nih.gov/31609778/)]
109. Mazo C, Kearns C, Mooney C, Gallagher WM. Clinical decision support systems in breast cancer: a systematic review. *Cancers (Basel)* 2020 Feb 06;12(2):369 [FREE Full text] [doi: [10.3390/cancers12020369](https://doi.org/10.3390/cancers12020369)] [Medline: [32041094](https://pubmed.ncbi.nlm.nih.gov/32041094/)]
110. Groenhof TKJ, Rittersma ZH, Bots ML, Brandjes M, Jacobs JLL, Grobbee DE, Members of the UCC-CVRM Study Group. A computerised decision support system for cardiovascular risk management 'live' in the electronic health record environment: development, validation and implementation-the Utrecht Cardiovascular Cohort Initiative. *Neth Heart J* 2019 Sep 1;27(9):435-442 [FREE Full text] [doi: [10.1007/s12471-019-01308-w](https://doi.org/10.1007/s12471-019-01308-w)] [Medline: [31372838](https://pubmed.ncbi.nlm.nih.gov/31372838/)]
111. Helmons P, Suijkerbuijk BO, Nannan Panday PV, Kosterink JGW. Drug-drug interaction checking assisted by clinical decision support: a return on investment analysis. *J Am Med Inform Assoc* 2015 Jul;22(4):764-772 [doi: [10.1093/jamia/ocu010](https://doi.org/10.1093/jamia/ocu010)] [Medline: [25670751](https://pubmed.ncbi.nlm.nih.gov/25670751/)]
112. Olakotan O, Mohd Yusof M, Ezat Wan Puteh S. A systematic review on CDSS alert appropriateness. *Stud Health Technol Inform* 2020 Jun 16;270:906-910 [doi: [10.3233/SHTI200293](https://doi.org/10.3233/SHTI200293)] [Medline: [32570513](https://pubmed.ncbi.nlm.nih.gov/32570513/)]
113. Antoniadi AM, Du Y, Guendouz Y, Wei L, Mazo C, Becker BA, et al. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. *Applied Sciences* 2021 May 31;11(11):5088 [doi: [10.3390/app11115088](https://doi.org/10.3390/app11115088)]

114. Jacobs M, Pradier MF, McCoy TH, Perlis RH, Doshi-Velez F, Gajos KZ. How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection. *Transl Psychiatry* 2021 Feb 04;11(1):108 [FREE Full text] [doi: [10.1038/s41398-021-01224-x](https://doi.org/10.1038/s41398-021-01224-x)] [Medline: [33542191](https://pubmed.ncbi.nlm.nih.gov/33542191/)]
115. Wolfien M, Klatt D, Salybekov AA, Li M, Komatsu-Horii M, Gaebel R, et al. Hematopoietic stem-cell senescence and myocardial repair - Coronary artery disease genotype/phenotype analysis of post-MI myocardial regeneration response induced by CABG/CD133+ bone marrow hematopoietic stem cell treatment in RCT PERFECT Phase 3. *EBioMedicine* 2020 Jul;57:102862 [FREE Full text] [doi: [10.1016/j.ebiom.2020.102862](https://doi.org/10.1016/j.ebiom.2020.102862)] [Medline: [32629392](https://pubmed.ncbi.nlm.nih.gov/32629392/)]
116. Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *J Am Med Inform Assoc* 2018 Aug 01;25(8):969-975 [FREE Full text] [doi: [10.1093/jamia/ocy032](https://doi.org/10.1093/jamia/ocy032)] [Medline: [29718407](https://pubmed.ncbi.nlm.nih.gov/29718407/)]
117. Rijnbeek P, Reps J. Chapter 13 Patient-Level Prediction. *The Book of OHDSI*. 2021 Jan 11. URL: <https://ohdsi.github.io/TheBookOfOhdsi/PatientLevelPrediction.html> [accessed 2023-07-01]
118. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998 May 12;97(18):1837-1847 [FREE Full text] [doi: [10.1161/01.cir.97.18.1837](https://doi.org/10.1161/01.cir.97.18.1837)] [Medline: [9603539](https://pubmed.ncbi.nlm.nih.gov/9603539/)]
119. Hageman S, Pennells L, Ojeda F, Kaptoge S, Dorresteyn J, Di Angelantonio E, SCORE2 working group and ESC Cardiovascular Risk Collaboration. SCORE2 models allow consideration of sex-specific cardiovascular disease risks by region. *Eur Heart J* 2022 Jan 25;43(3):241-242 [doi: [10.1093/eurheartj/ehab761](https://doi.org/10.1093/eurheartj/ehab761)] [Medline: [35080243](https://pubmed.ncbi.nlm.nih.gov/35080243/)]
120. Carus J, Nürnberg S, Ückert F, Schlüter C, Bartels S. Mapping cancer registry data to the episode domain of the Observational Medical Outcomes Partnership Model (OMOP). *Applied Sciences* 2022 Apr 15;12(8):4010 [doi: [10.3390/app12084010](https://doi.org/10.3390/app12084010)]
121. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol* 2017 Jan;106(1):1-9 [FREE Full text] [doi: [10.1007/s00392-016-1025-6](https://doi.org/10.1007/s00392-016-1025-6)] [Medline: [27557678](https://pubmed.ncbi.nlm.nih.gov/27557678/)]
122. Beginners. R Studio Education. URL: <https://education.rstudio.com/learn/beginner/> [accessed 2023-07-01]
123. OpenMined. URL: <https://www.openmined.org/> [accessed 2023-07-01]
124. Welcome to PyHealth!. PyHealth. URL: <https://pyhealth.readthedocs.io/en/latest/> [accessed 2023-07-01]
125. Shiny. URL: <https://shiny.rstudio.com/> [accessed 2023-07-01]
126. Hossain S. Visualization of Bioinformatics Data with Dash Bio. Zenodo. 2019 Jul 10. URL: <https://zenodo.org/record/3346213> [accessed 2023-07-01]
127. Clarke S, Wilson ML, Terhaar M. Using dashboard technology and clinical decision support systems to improve heart team efficiency and accuracy: review of the literature. *Stud Health Technol Inform* 2016;225:364-366 [Medline: [27332223](https://pubmed.ncbi.nlm.nih.gov/27332223/)]
128. Esquer Rochin MA, Gutierrez-Garcia JO, Rosales J, Rodriguez L. Design and evaluation of a dashboard to support the comprehension of the progression of patients with dementia in day centers. *Int J Med Inform* 2021 Dec;156:104617 [doi: [10.1016/j.ijmedinf.2021.104617](https://doi.org/10.1016/j.ijmedinf.2021.104617)] [Medline: [34710725](https://pubmed.ncbi.nlm.nih.gov/34710725/)]
129. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013 Apr 02;6(269):p11 [FREE Full text] [doi: [10.1126/scisignal.2004088](https://doi.org/10.1126/scisignal.2004088)] [Medline: [23550210](https://pubmed.ncbi.nlm.nih.gov/23550210/)]
130. Reimer N, Unberath P, Busch H, Börries M, Metzger P, Ustjanzew A, et al. Challenges and experiences extending the cBioPortal for cancer genomics to a molecular tumor board platform. *Stud Health Technol Inform* 2021 Nov 18;287:139-143 [doi: [10.3233/SHTI210833](https://doi.org/10.3233/SHTI210833)] [Medline: [34795098](https://pubmed.ncbi.nlm.nih.gov/34795098/)]
131. Brlek P, Kafka A, Bukovac A, Pećina-Šlaus N. Integrative cBioPortal analysis revealed molecular mechanisms that regulate EGFR-PI3K-AKT-mTOR pathway in diffuse gliomas of the brain. *Cancers (Basel)* 2021 Jun 29;13(13):3247 [FREE Full text] [doi: [10.3390/cancers13133247](https://doi.org/10.3390/cancers13133247)] [Medline: [34209611](https://pubmed.ncbi.nlm.nih.gov/34209611/)]
132. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010;17(2):124-130 [FREE Full text] [doi: [10.1136/jamia.2009.000893](https://doi.org/10.1136/jamia.2009.000893)] [Medline: [20190053](https://pubmed.ncbi.nlm.nih.gov/20190053/)]
133. Castro V, Gainer V, Wattanasin N, Benoit B, Cagan A, Ghosh B, et al. The Mass General Brigham Biobank Portal: an i2b2-based data repository linking disparate and high-dimensional patient data to support multimodal analytics. *J Am Med Inform Assoc* 2022 Mar 15;29(4):643-651 [FREE Full text] [doi: [10.1093/jamia/ocab264](https://doi.org/10.1093/jamia/ocab264)] [Medline: [34849976](https://pubmed.ncbi.nlm.nih.gov/34849976/)]
134. Whittington KD, Walker J, Hirsch B. Promoting interdisciplinary communication as a vital function of effective teamwork to positively impact patient outcomes, satisfaction, and employee engagement. *J Med Imaging Radiat Sci* 2020 Dec;51(4S):S107-S111 [doi: [10.1016/j.jmir.2020.07.002](https://doi.org/10.1016/j.jmir.2020.07.002)] [Medline: [32768292](https://pubmed.ncbi.nlm.nih.gov/32768292/)]
135. The Medical Informatics Initiative's core data set. Medical Informatics Initiative'. URL: <https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set> [accessed 2023-07-01]
136. Sass J, Bartschke A, Lehne M, Essenwanger A, Rinaldi E, Rudolph S, et al. The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. *BMC Med Inform Decis Mak* 2020 Dec 21;20(1):341 [FREE Full text] [doi: [10.1186/s12911-020-01374-w](https://doi.org/10.1186/s12911-020-01374-w)] [Medline: [33349259](https://pubmed.ncbi.nlm.nih.gov/33349259/)]
137. Prokosch H, Bahls T, Bialke M, Eils J, Fegeler C, Gruendner J, et al. The COVID-19 Data Exchange Platform of the German University Medicine. *Stud Health Technol Inform* 2022 May 25;294:674-678 [doi: [10.3233/SHTI220554](https://doi.org/10.3233/SHTI220554)] [Medline: [35612174](https://pubmed.ncbi.nlm.nih.gov/35612174/)]

138. Gruendner J, Deppenwiese N, Folz M, Köhler T, Kroll B, Prokosch H, et al. The architecture of a feasibility query portal for distributed COVID-19 Fast Healthcare Interoperability Resources (FHIR) patient data repositories: design and implementation study. *JMIR Med Inform* 2022 May 25;10(5):e36709 [FREE Full text] [doi: [10.2196/36709](https://doi.org/10.2196/36709)] [Medline: [35486893](https://pubmed.ncbi.nlm.nih.gov/35486893/)]
139. Harrison P, Lopez R, Rahman N, Allen SG, Aslam R, Buso N, et al. The COVID-19 Data Portal: accelerating SARS-CoV-2 and COVID-19 research through rapid open access data sharing. *Nucleic Acids Res* 2021 Jul 02;49(W1):W619-W623 [FREE Full text] [doi: [10.1093/nar/gkab417](https://doi.org/10.1093/nar/gkab417)] [Medline: [34048576](https://pubmed.ncbi.nlm.nih.gov/34048576/)]
140. OECD. Towards an integrated health information system in the Netherlands. Paris, France: OECD Publishing; 2022.
141. Prokosch H, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. *Methods Inf Med* 2018 Jul 17;57(S 01):e82-e91 [doi: [10.3414/me17-02-0025](https://doi.org/10.3414/me17-02-0025)]
142. Part-time online master's degree "Biomedical Informatics and Data Science". Hochschule Mannheim. URL: <https://www.master-bids.hs-mannheim.de/> [accessed 2023-07-01]
143. Ismail L, Materwala H, Karduck AP, Adem A. Requirements of health data management systems for biomedical care and research: scoping review. *J Med Internet Res* 2020 Jul 07;22(7):e17508 [FREE Full text] [doi: [10.2196/17508](https://doi.org/10.2196/17508)] [Medline: [32348265](https://pubmed.ncbi.nlm.nih.gov/32348265/)]
144. Ozaydin B, Zengul F, Oner N, Feldman SS. Healthcare research and analytics data infrastructure solution: a data warehouse for health services research. *J Med Internet Res* 2020 Jun 04;22(6):e18579 [FREE Full text] [doi: [10.2196/18579](https://doi.org/10.2196/18579)] [Medline: [32496199](https://pubmed.ncbi.nlm.nih.gov/32496199/)]
145. v5 Tutorial for DataSHIELD users. Confluence. URL: <https://data2knowledge.atlassian.net/wiki/spaces/DSDEV/pages/714571780/v5+Tutorial+for+DataSHIELD+users> [accessed 2023-07-01]
146. Open Data Institute. URL: <https://theodi.org/> [accessed 2023-07-01]
147. theodi / synthetic-data-tutorial. GitHub. 2022 Apr 12. URL: <https://github.com/theodi/synthetic-data-tutorial> [accessed 2023-07-01]
148. Dash. Real Python. URL: <https://realpython.com/python-dash/> [accessed 2023-07-01]

Abbreviations

- AI:** artificial intelligence
- CART:** Classification and Regression Tree
- CDM:** common data model
- CDSS:** clinical decision support system
- CIS:** clinical information system
- DARWIN:** Data Analysis and Real World Interrogation Network
- EHR:** electronic health record
- ePa:** elektronische Patientenakte
- ETL:** extract, transform, and load
- FAIR:** findable, accessible, interoperable, reusable
- FHIR:** Fast Healthcare Interoperability Resources
- GAN:** generative adversarial network
- GDPR:** General Data Protection Regulation
- GECCO:** German Corona Consensus Dataset
- HL7:** Health Level 7
- i2b2:** Informatics for Integrating Biology and the Bedside
- ICD:** International Statistical Classification of Diseases and Related Health Problems
- IHE:** Integrating the Healthcare Enterprise
- LOINC:** Logical Observation Identifiers Names and Codes
- III:** Medical Informatics Initiative
- MIRACUM:** Medical Informatics in Research and Care in University Medicine
- ML:** machine learning
- NUM:** Network University Medicine
- OHDSI:** Observational Health Data Sciences and Informatics
- OMOP:** Observational Medical Outcomes Partnership
- PLP:** patient-level prediction
- SMPC:** secure multiparty computation
- SNOMED CT:** Systematized Nomenclature of Medicine and Clinical Terms
- VAE:** variational autoencoder

Edited by A Mavragani; submitted 23.01.23; peer-reviewed by I Mircheva, Q Wu, H Oh; comments to author 22.02.23; revised version received 29.03.23; accepted 11.04.23; published 24.07.23

Please cite as:

Wolfien M, Ahmadi N, Fitzer K, Grummt S, Heine KL, Jung IC, Krefting D, Kühn A, Peng Y, Reinecke I, Scheel J, Schmidt T, Schmücker P, Schüttler C, Waltemath D, Zoch M, Sedlmayr M

Ten Topics to Get Started in Medical Informatics Research

J Med Internet Res 2023;25:e45948

URL: <https://www.jmir.org/2023/1/e45948>

doi: [10.2196/45948](https://doi.org/10.2196/45948)

PMID:

©Markus Wolfien, Najia Ahmadi, Kai Fitzer, Sophia Grummt, Kilian-Ludwig Heine, Ian-C Jung, Dagmar Krefting, Andreas Kühn, Yuan Peng, Ines Reinecke, Julia Scheel, Tobias Schmidt, Paul Schmücker, Christina Schüttler, Dagmar Waltemath, Michele Zoch, Martin Sedlmayr. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 24.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.