

Implementation Report

Leveraging Interoperable Electronic Health Record (EHR) Data for Distributed Analyses in Clinical Research: Technical Implementation Report of the HELP Study

Julia Palm^{1*}; Kutaiba Saleh^{2*}; André Scherag¹; Danny Ammon²

¹Institute of Medical Statistics, Computer and Data Sciences, Jena University Hospital, Jena, Germany

²Data Integration Center, Jena University Hospital, Jena, Germany

*these authors contributed equally

Corresponding Author:

Julia Palm

Institute of Medical Statistics, Computer and Data Sciences, Jena University Hospital

Bachstraße 18

Jena 07743

Germany

Phone: 49 3641-9-396961

Fax: 49 3641-9-396952

Email: julia.palm@med.uni-jena.de

Abstract

Background: The Medical Informatics Initiative (MII) Germany established 38 data integration centers (DIC) in university hospitals to improve health care and biomedical research through the use of electronic health record (EHR) data. To showcase the value of these DIC, the HELP (Hospital-wide Electronic Medical Record Evaluated Computerized Decision Support System to Improve Outcomes of Patients with Staphylococcal Bloodstream Infection) study was initiated as a use case. This study is a clinical trial designed to assess the impact of a computerized decision support system for managing staphylococcal bacteremia.

Objective: In this paper, we present the lessons learned during the use case from a technical perspective. This paper outlines the challenges encountered and solutions developed during our initial implementation of this infrastructure, providing insights applicable to other research platforms using EHR data. These insights are organized into 3 key areas: study-specific data definition and modeling, interoperable data integration and transformation, and distributed data extraction and analysis.

Methods: An interdisciplinary team of clinicians, computer scientists, and statisticians created a catalog of items to identify data elements necessary for the study's evaluation and developed a domain-specific information model. DIC developed extract-transform-load pipelines to collect the disparate, site-specific EHR data and to transform it into a common data format. Health Level Seven International (HL7) Fast Healthcare Interoperability Resources (FHIR) and the MII's core dataset profiles were adopted for consistent data representation across sites. Additionally, data not present in EHRs was gathered using structured electronic case report forms. Analysis scripts were then distributed to the sites to preprocess the data locally, followed by a central analysis of the preprocessed data to generate the final overall results.

Implementation (Results): Our analysis revealed significant heterogeneity in data quality and implementation of interoperability standards, requiring substantial harmonization efforts. The development of analysis scripts and data extraction processes demanded multiple iterative cycles and close collaboration with local data experts. Despite these challenges, the successful implementation demonstrated the feasibility of distributed EHR analyses while highlighting the importance of thorough data quality assessment, realistic timeline planning, and multidisciplinary expertise.

Conclusions: The HELP study highlights challenges and opportunities in leveraging EHR data for clinical research, particularly in the absence of mandatory data standards and resource-intensive data harmonization efforts. Despite limitations in data availability and quality, progress in digitization and interoperability frameworks offers hope for future improvements. Lessons learned from this study can inform the development of standardized methodologies and infrastructures for sustainable EHR data integration in research.

Keywords: clinical decision support system; data collection methods; electronic health records; health information interoperability; software design

Introduction

Context

The digitization of the health sector has significantly increased interest in using electronic health record (EHR) data for biomedical research, leading to the development of new infrastructures for data integration and sharing. In line with this trend, the Medical Informatics Initiative (MII) Germany established 38 data integration centers (DIC) in university hospitals to improve health care and biomedical research based on EHR data [1,2]. To demonstrate the usefulness of these DIC, several use cases were designed, encompassing various studies that leveraged the DIC infrastructure. One notable use case was the HELP (Hospital-wide Electronic Medical Record Evaluated Computerized Decision Support System to Improve Outcomes of Patients with Staphylococcal Bloodstream Infection) [3] study, which aimed to enhance the clinical management of *Staphylococcus* bacteremia. This condition requires prompt and guideline-compliant treatment. While involving an infectious disease specialist can improve patient outcomes, immediate access to these specialists is not always available. From June 2020 to October 2022, the HELP trial explored whether a computerized decision support system (CDSS) could safely enhance the standard of care for these patients. Conducted as a multicenter, noninferiority, interventional stepped-wedge cluster randomized controlled trial at 5 German university hospitals, the CDSS assisted physicians in managing bacteremia until an infectious disease specialist became available. It was deployed as a website with an interactive decision tree, accessible via smartphones, desktops, or laptops. An archived version of the CDSS is available online [4].

Problem Statement

Clinical evaluation of interventions such as the HELP CDSS through traditional randomized controlled trials is resource-intensive, particularly due to extensive data collection and management requirements. However, much of this necessary data already exists within EHR but remains largely inaccessible due to proprietary data formats and fragmented storage across multiple systems. The MII addresses this challenge by collecting, harmonizing, and transforming this

data into interoperable standards, making it readily available for studies like HELP. By using EHR data for the evaluation of interventions in clinical trials, we can blend the characteristics of traditional randomized controlled trials with the practicality of being embedded within the clinical routine. While this setup imposes some restrictions on data availability and quality, it may more closely mirror real-world clinical scenarios, potentially contributing to an improved ecological validity of the trial [5]. Mc Cord et al [6] recently showed that randomized trials using routinely collected data show lower, but likely more realistic, effect estimates than traditional trials. Since this effect persists even when controlling for data quality, the authors suggest that trials using routinely collected data are less prone to overoptimistic treatment adherence and related biases. Even if one may not agree to use EHR data for the most important outcomes, secondary use of EHR data has the potential to reduce the documentation burden associated with traditional trials, offering a more pragmatic study design [7].

While the clinical results of the trial are published elsewhere [8], this manuscript details the lessons learned from the technical perspective.

Methods

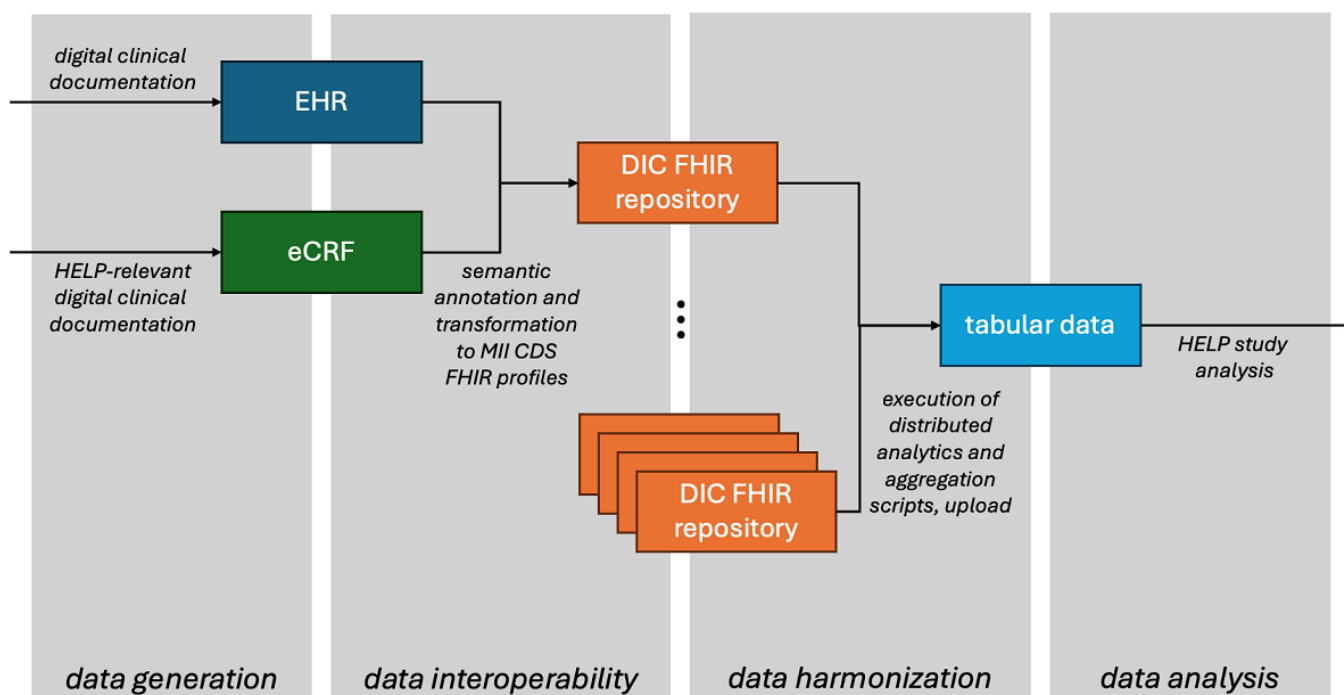
Aims and Objectives

Our study was designed as a proof of concept of a distributed analysis approach combining EHR data with structured electronic case report form (eCRF) data to evaluate outcomes of a randomized clinical trial. This implementation aimed to illustrate how the secondary use of EHR data can streamline clinical trial processes.

Blueprint Summary

To gather data for analysis, the trial used a hybrid data collection approach and relied on the DIC structures for data extraction, transformation, and harmonization (Figure 1). Data available from EHRs were collected directly, while additional data were gathered via eCRFs. All data, regardless of source, were collected in the DIC and transformed into an interoperable, standardized format.

Figure 1. Overall data flow in the HELP (Hospital-wide Electronic Medical Record Evaluated Computerized Decision Support System to Improve Outcomes of Patients with Staphylococcal Bloodstream Infection) study, based on interoperable formats created in data integration centers (DIC). EHR: electronic health record; eCRF: electronic case report form; FHIR: Fast Healthcare Interoperability Resources; MII CDS: core dataset of the Medical Informatics Initiative.



Technical Design

Study-Specific Data Definition and Modeling

At the outset of the study, a comprehensive, study-specific data dictionary, a catalog of items (COI) was developed to serve as a reference for each DIC, detailing the specific data that needed to be extracted from the EHR. The initial version of the COI was developed in several modeling sessions involving at least 2 infectious disease specialists, at least 1 statistician, and at least 2 medical informaticians from the DIC. In these sessions, under the guidance of DIC staff, the following results were achieved:

1. Clinical definition of evidence-based treatment procedures for suspected staphylococcal bloodstream infections (prepared with relevant studies by medical staff from different disciplines).
2. Development of a visualized algorithm for the treatment steps in BPMN format (based on modeling expertise of DIC staff).
3. Identification of critical data and time points that are essential for the evaluation of the patient's status as well as the study results (prepared based on treatment and analysis expertise of medical and statistical staff with different hierarchy levels).
4. Mapping of these data and time points to a COI (prepared using COI development templates by DIC staff).

The interdisciplinary collaboration allowed infectious disease specialists and experts in medical microbiology to contribute their expertise and ensure comprehensive requirements gathering, while DIC staff coordinated the collection and specification of the COI to ensure consistency, completeness,

and suitability for further technical specification. Initially, an extensive spreadsheet was created, containing all identified items. This spreadsheet included detailed information on each item, such as its name, data source, and categorization. For illustrative purposes, this tabular COI has been published [9]. In the next step, the items needed to be further specified. Precise names, properties, and relationships between items can be documented and visualized by developing a domain-specific information model [10]—a process also used in the MII [11]. Given the complexity of clinical data, information models ensure clarity, relationships, and visualization between required items, allowing for semantic enrichment of information that can be useful in later analysis. Various notations can be used for information modeling; in the HELP study, the ART-DECOR tool (version 2.2), developed for the standardization of health data, was used [12]. This approach ensured that details such as the dosage of relevant antibiotics for patients were represented correctly [13]. Thus, domain-specific modeling is the first step toward creating an interoperable data basis across multiple locations.

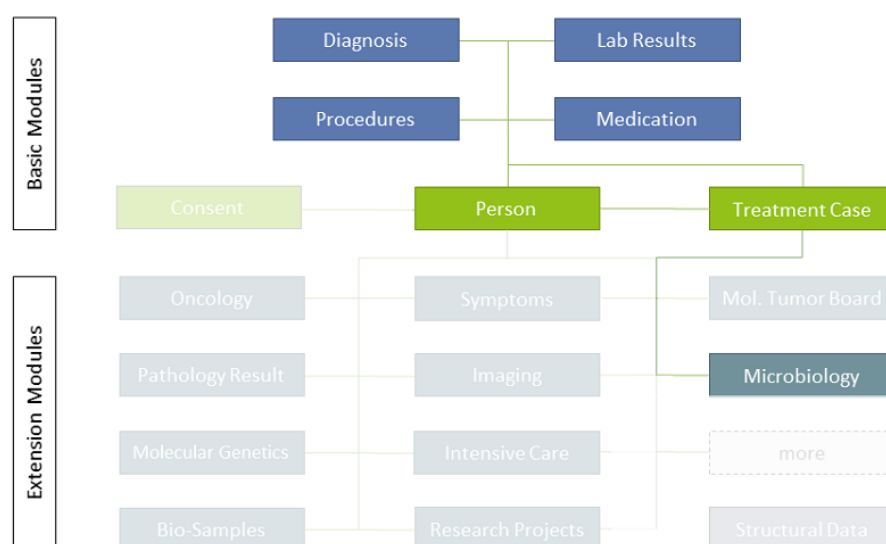
Interoperable Data Integration and Transformation

To facilitate multicenter biomedical research projects based on secondary use of EHR data with the newly established DIC [14], the contents of the EHR in the German university hospitals had to be made accessible gradually. Beginning in 2017, the MII Interoperability Working Group therefore started to define the MII core dataset (CDS; see Figure 2), which separates basic, consistently required data elements of EHR from discipline-specific elements through modularization [11]. Implementing the CDS, the DIC started by

providing access to data from the basic modules across all locations and then successively included the extension modules based on current research projects at each location.

The HELP study was part of these efforts, using data from both basic and extension modules of the CDS.

Figure 2. Illustration of modules of the Medical Informatics Initiative (MII) core dataset. The HELP (Hospital-wide Electronic Medical Record Evaluated Computerized Decision Support System to Improve Outcomes of Patients with Staphylococcal Bloodstream Infection) study used the basic modules “person,” “treatment case,” “diagnosis,” “procedures,” “lab results,” “medication” [15], and the extension module “microbiology” [16].



The step-by-step implementation of the CDS is due to the wide variety of primary documentation systems (hospital information systems, intensive care systems, laboratory information systems, etc) from different vendors operating in German hospitals, whose data together comprise the contents of an EHR. DIC have been designed to eliminate the need for data scientists to collect and merge proprietary formats from different systems and locations repeatedly for every project.

Within the MII, a consensus was reached to use Health Level Seven International (HL7) Fast Healthcare Interoperability Resources (FHIR) for the technical representation and exchange of CDS data across DIC [11]. Despite some limitations for secondary data use, this standard was chosen based on its ability to support a wide variety of research questions and based on its relevance as a common standard for EHR in the German and European health care system in the future [17]. Adopting a health care interoperability standard enables site- and vendor-independent EHR data representation, explicating data properties through a variety of metadata [17-19]. To address specifics of the German health care system (eg, national terminologies), secondary use requirements (eg, custom search parameters), and FHIR usage guidelines, the MII CDS modules were developed as FHIR profiles derived from the international standard. These profiles were balloted and published in the form of implementation guides for application in the DIC [11]. For the

HELP study, the relevant modules of the MII CDS could be identified based on the established COI (Figure 2).

The subsequent phase thus involved integrating the data elements required for the HELP study from the various primary documentation systems in all participating locations and transforming them to the FHIR format of the MII CDS modules in the DIC. At the time of the HELP study, the majority of hospitals in Germany did not yet have FHIR interfaces for their EHR systems and thus hardly any interoperable access to the data required for use. The most important task for the DIC in the process of being established was therefore the identification and connection of data source systems. The relevant EHR data are captured in different IT systems, which also had to be connected in different ways: the hospital information system for patient master data, patient movement data, diagnoses, and procedures; the laboratory information system for laboratory tests and microbiology results; and the clinical workplace system or the e-medication system for medication administrations. These connections were established by implementing ETL (extract-transform-load) pipelines for data integrations either directly from the persistence layers or database systems of the source systems or by processing already exchanged message formats for communication between the clinical IT systems—in particular HL7v2, which is often used in Germany, with many proprietary extensions. The selection of the specific connection method is highly location- and vendor-dependent.

Due to the hybrid data collection approach of the study, which included the use of eCRFs for data not accessible in the EHR, the data collected via eCRFs also had to be transformed into the FHIR format, for which a separate eCRF2FHIR transformation tool was developed.

Distributed Data Extraction and Analysis

The FHIR standard's use of JSON or XML as notation, its alignment with contemporary web API (Application Programming Interface) practices, and its resource-based data modeling approach are optimized for modern implementation as well as for the precise representation of health care information. The nested JSON data structure is particularly useful to accommodate a number of the FAIR (Findable, Accessible, Interoperable, Reusable) principles [20], especially that data should be enhanced with "rich metadata" and be "richly described with a plurality of accurate and relevant attributes." Even beyond the use of FHIR in health care settings, JSON and similar nested data formats have been gaining popularity for big data and open data repositories [21,22]. In the health care context, FHIR benefits significantly from the flexibility of nested structures. This is because the dynamic nature of health care data, with its varying elements and data types, is more accurately represented in nested structures than in traditional tabular data formats.

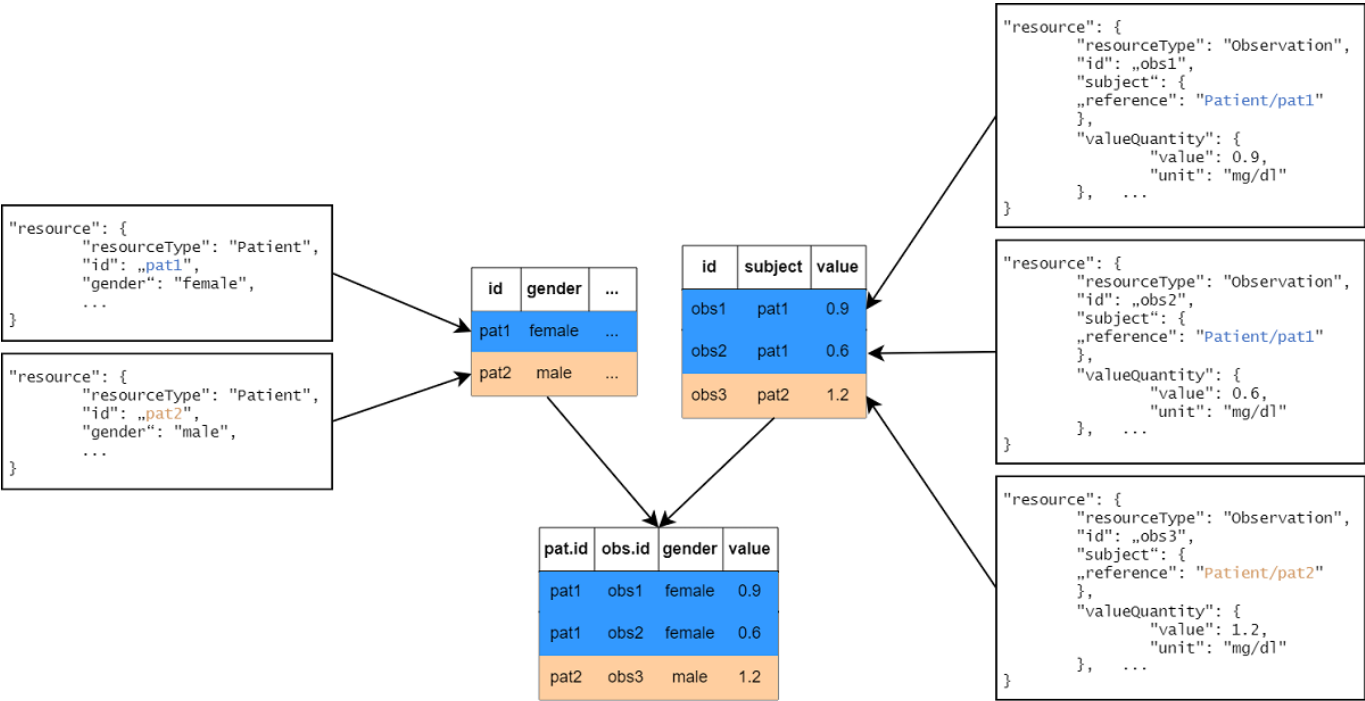
The statistical analysis in our study, however, required tabular data, as most common data analysis techniques do. To make the FHIR data available for this statistical analysis, we therefore transformed the FHIR data into tables using the *fhircrackr* package (version 2.1) [23] within the R statistical computing environment (version 4.1) [24]. Despite all harmonization efforts, both the primary data from the EHR and their FHIR implementation were expected to be at least partly different across the 5 study sites. Consequently, we developed individual data extraction and tabularization scripts

for each DIC, taking into account the local characteristics of the data. The scripts, versioned and distributed via a GitLab repository, queried the required data from the respective FHIR server, tabularized it, and performed local aggregation and anonymization steps.

A crucial step in the data analysis was the classification of patients into intervention versus control conditions, followed by the identification and extraction of all study-relevant data items for those patients. In the stepped-wedge design, clinical wards are assigned to the intervention and control conditions, instead of individual patients [25]. Thus, the classification of a patient into either the control or intervention group was based on the ward where the patient was treated when they received their first blood culture result indicating a possible *Staphylococcus* infection. This required querying data from different data sources represented at each DIC and was implemented by the following overall workflow: (1) process questionnaire data from the eCRFs to identify all patients participating in the trial; (2) analyze encounter data to track patient movement through wards; (3) review medical microbiology reports to identify the first blood culture positive for *Staphylococcus*; (4) link the medical microbiology report dates with the encounter data to determine whether the patient was on a control or intervention ward at the time of the positive blood culture result; and (5) download and tabularize all data for the primary or secondary endpoint analysis as well as for the clinical description of the patient cohort.

FHIR resources containing the relevant data were processed by type, with a separate table created for each resource type during the initial tabularization step. The information from different resources was then connected by joining these tables based on the references linking the original FHIR resources (see Figure 3).

Figure 3. One table was created per resource type; afterwards, tables were joined based on references.



The statistical analysis of the trial data was divided into 2 parts. The first analysis step was conducted locally at each DIC to remove patient-identifying information and aggregate the data where possible, thereby minimizing the risk of re-identification. The data was processed based on its importance for the trial evaluation. Data for the primary and secondary endpoints were maintained at the patient level to maximize statistical power in the central analysis. These data were reduced to the minimal set of variables necessary for statistical analysis and sent to Jena for central analysis using generalized linear mixed models [26]. Conversely, data needed for the clinical description of the study population were aggregated at each site (eg, to mean and SD or counts) and then centrally combined using meta-analytic methods. All data were transferred to the central analysis site manually via a securely provided cloud, since the central MII infrastructure for data sharing was not established yet [27].

repeated consultations with infectious disease specialists and DIC employees to clarify specific datasets.

We initially planned to use project-specific FHIR profiles for the HELP use case and started with their development. However, both the amount of work involved in generating profiles derived from the CDS and tailored to the project and the unclear benefits—as compliance with the stricter requirements of project-specific profiles at the sites could not be verified—made it questionable whether the time saved in the analysis by project-specific FHIR implementation guides justified these efforts. Consequently, we did not complete the profiling work, but the results and experiences were incorporated into the further development of the CDS extension modules. The reliance on interoperability standards for data analysis must be investigated further in future projects [28].

Implementation (Results)

Outcomes

Study-Specific Data Definition and Modeling

Identification and explicit description of the data to be used in the study constituted the first step, but creating a finalized, binding version of the COI proved to be an iterative and complex process. A major challenge was the varying status of available data and its assessment regarding clinical priorities, which required a continuous refinement of the COI even during the data integration and transformation phase. Throughout the data extraction process, we encountered numerous occasions where the data found in the EHR did not match the items from the COI exactly, requiring

Interoperable Data Integration and Transformation

Both the transformation of proprietary, vendor- and location-specific health care data into interoperability standards and the processing of such interoperable data for analysis in a research study required the development of software tools: ETL pipelines, semantic annotation and mapping programs, structure conversions for data analysis, and data analysis scripts. Project-specific software development, however, inadvertently causes errors and bugs, most of which can only be found when the software is used practically in the intended context. Consequently, software development required significant time during the study, which is unsurprising given that many of the tasks the software addressed were novel to this project. However, this time investment is only justifiable if the software can be reused in future projects; otherwise, it is not sustainable.

The main obstacle during data extraction and transformation was the varying quality of the EHR data, which we encountered in 3 distinct forms:

1. In cases where data are entered manually into the EHR, the person documenting it may be rushed or may not consider it important to ensure accuracy. As a result, manual data entries can contain errors, which may be corrected later or remain, leading to inconsistencies or contradictions in the extracted data.
2. Many entries in EHR are only given in free text form and do not provide semantically annotated information. This prevented the processing of the contained information in the HELP study and led to the described approach of hybrid documentation.
3. Even structured data are not always unconditionally machine-readable. An example is the coding of measurement units in laboratory data. In communication between clinicians, variations in how a measurement unit is presented, how a formula is formatted, or even minor errors in notation are typically understood in context due to the clinician's background knowledge. [Figure 4](#), for example, represents 4 different notations of alanine aminotransferase (ALT) and aspartate aminotransferase (AST) measurement units encountered in an EHR. While any clinician can easily see that the unit described is $\mu\text{mol}/(\text{l} * \text{s})$, computers are unable to interpret these discrepancies without explicit standardization.

Figure 4. Examples for units of measurement for alanine aminotransferase (ALT) and aspartate aminotransferase (AST) in German laboratory results, captured from a PDF.

umol/l*s

μmol/l*s

μmol/s.l

fmol/l#s

Furthermore, relevant data were captured at different locations and by different individuals using different software, which can also change over time [29]. One possible countermeasure to those differences would have been a stricter specification of uniform data, which can be achieved through project-specific FHIR profiling with more extensive constraints than those provided by the MII CDS. However, this was not feasible within the HELP study (cf previous paragraph).

Distributed Data Extraction and Analysis

As a result of the heterogeneity of the primary data described earlier, the data provided by the 5 DIC showed substantial heterogeneity as well. Although all DIC theoretically follow a comparable IT and service architecture [11], their practical implementations vary significantly, often due to local requirements. Differences emerged even at the initial access stage: some DIC required analysis scripts to be delivered in Docker containers, while others only accepted plain R scripts to be executed by DIC employees. Additionally, some DIC had installed a FHIR repository dedicated solely to the HELP use case, containing data only for study participants, whereas other DIC used a single, comprehensive FHIR repository with

all available data. This necessitated individualized querying routines that were developed in a highly iterative workflow and close collaboration with each DIC.

On the next level, the implementation of the FHIR resources differed significantly (see [Figure 5](#)). Although all resources were supposed to conform to the same profiles defined in the CDS, we identified two major sources of variation. First, the DIC differed in the primary data sources connected at the time of the study. For example, diagnoses originated from data exports for subsequent billing by the hospitals at some DIC and from clinical documentation during patient treatment at others. This resulted in varying FHIR resources, as specific data elements were not available in some sources (eg, the diagnosis date in billing data) and led to different interpretations (eg, diagnoses optimized for billing purposes after treatment versus purposes during treatment). Second, the implementation of profile-conformant FHIR resources was at different stages. For instance, the CDS provides a multilevel model for representing encounter data that aligns with the administrative structure of a German hospital. However, at the time of the study, only 1 of the 5 DIC had fully implemented this complex model, while the other 4 used more rudimentary, simplified representations.

Figure 5. Two exemplary FHIR (Fast Healthcare Interoperability Resources) resources, illustrating potential differences despite adhering to the same FHIR profile. Elements that exist in both resources but contain different codes are highlighted in yellow, while elements present in only 1 resource are marked in red. The resources have been shortened and anonymized.

```

1 {
2   "resourceType": "Observation",
3   "status": "final",
4   "category": [{
5     "coding": [{
6       "system": "http://terminology.hl7.org/CodeSystem/observation-category",
7       "code": "laboratory",
8       "display": "Laboratory"
9     }],
10    {
11      "system": "http://loinc.org",
12      "code": "26436-6",
13      "display": "Laboratory studies (set)"
14    }
15  ]
16 },
17 {
18   "code": {
19     "coding": [{
20       "system": "https://uk2.de/LAB/Vexus/Swisslab/Observation/Code",
21       "version": "1.0.0",
22       "code": "zB11L1",
23       "display": "Bilirubin (gesamt)"
24     }],
25     {
26       "system": "http://loinc.org",
27       "code": "1975-2"
28     }
29   ]
30 },
31 "subject": {
32   "reference": "Patient/d34badtn658q32werfdvixfgu8j8345"
33 },
34 "effectiveDateTime": "2021-01-30T01:01:00+01:00",
35 "issued": "2021-01-30T01:01:00+01:00",
36 "valueQuantity": {
37   "value": 0.6,
38   "unit": "mg/dL",
39   "system": "http://units-of-measure.org",
40   "code": "mg/dL"
41 }
42 }

```

```

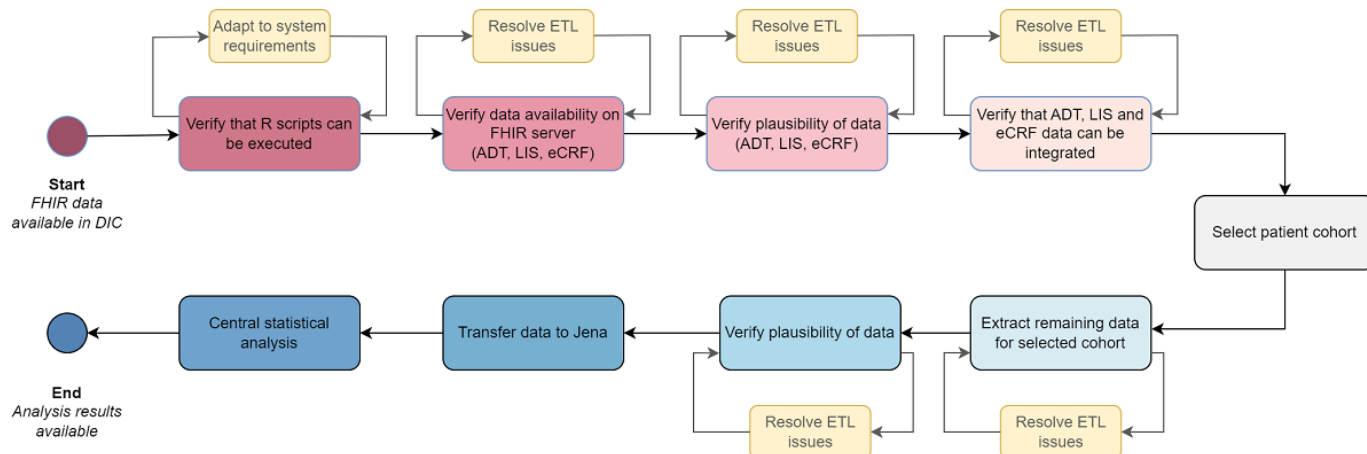
1 {
2   "resourceType": "Observation",
3   "status": "final",
4   "category": [{
5     "coding": [{
6       "system": "http://terminology.hl7.org/CodeSystem/observation-category",
7       "code": "laboratory",
8       "display": "Laboratory"
9     }],
10    {
11      "system": "http://loinc.org",
12      "code": "26436-6",
13      "display": "Laboratory studies (set)"
14    }
15  ]
16 },
17 {
18   "code": {
19     "coding": [{
20       "system": "http://loinc.org",
21       "code": "14631-6",
22       "display": "Bilirubin total [Moles/volume] in Serum or Plasma"
23     }],
24     {
25       "system": "http://www.uk1.de/CodeSystem/IIM/analyt",
26       "code": "B11L_3",
27       "display": "B11L_3"
28     }
29   ]
30 },
31 "subject": {
32   "reference": "Patient/deb5d5f1e5fbdddb1e80adfa6231elf9alc0a2720a6aed72fb18ec6c0cd60c2"
33 },
34 "encounter": {
35   "reference": "Encounter/cb5077b7c795955f8b4692ffe38cd0b9710315247594437b0484547e1cbd3543"
36 },
37 "effectiveDateTime": "2020-10-13T08:22:48+01:00",
38 "valueQuantity": {
39   "value": 6.1,
40   "unit": "µmol/L",
41   "system": "http://units-of-measure.org",
42   "code": "µmol/L"
43 }
44 }

```

These sources of data heterogeneity, which persisted even after being transformed to an interoperable format, had to be addressed through harmonization during data extraction and analysis. According to Nguyen [30], interoperability involves the transactional sharing of health care data, while data harmonization integrates data from multiple sources to preserve context and normalize meanings across datasets. FHIR, as an interoperability standard, allows for unified data representation from different sources but does not address data harmonization, which requires understanding the context and meaning of the data. While FHIR resources can provide context in the form of metadata (eg, indicating whether a diagnosis originated from billing or clinical information),

managing these different contexts in the analysis is a task of harmonization. In multicenter projects like HELP, it is crucial to recognize that no single person can anticipate the heterogeneous data across all study sites. This is because understanding such data requires site-specific information about the clinical routines and the distinct features of the individual information systems used at each location. In our case, the harmonization therefore required numerous iterations of communication and script adaptation with the respective data providers. To ensure feasibility, we broke the extraction process down into manageable subtasks as depicted in Figure 6. Each subtask was iterated with each DIC until it produced plausible results.

Figure 6. Iterative data extraction and analysis process in the HELP (Hospital-wide Electronic Medical Record Evaluated Computerized Decision Support System to Improve Outcomes of Patients with Staphylococcal Bloodstream Infection) study. FHIR: Fast Healthcare Interoperability Resources; DIC: data integration center; ADT: admission, discharge, and transfer data; LIS: laboratory information system; eCRF: electronic case report form; ETL: extract-transform-load.



The infrastructure for our distributed analysis needed to be straightforward to implement, flexible, and compliant with

strict privacy regulations for health data. The requirement for stringent data protection precluded the use of any cloud

computation services. Additionally, the need for simplicity ruled out established privacy-preserving solutions like DataSHIELD [31], as setting up a dedicated OPAL server at each DIC was not feasible for this study. Instead, we opted for a straightforward infrastructure by distributing plain R scripts via GitLab and Docker images through Docker Hub. Overall, the script development process was extremely time-consuming, partly because the technical infrastructure and FHIR mapping were still under construction during the analysis. While this helped identify problems in DIC development early on, it required more iterations than anticipated.

Lessons Learned

Overview

The lessons learned presented in this paper emerged from systematic observations and collaborative discussions throughout our study's implementation phase. Our experience demonstrates that distributed analyses using EHR data are feasible but require careful planning and consideration of key challenges. While these insights were primarily derived from our experiences within the MII, they extend beyond this framework and offer valuable guidance for researchers planning clinical trials that incorporate EHR data. Based on our experience implementing distributed analyses, we identified 3 fundamental considerations that warrant careful attention during the project-planning phase. These insights are particularly relevant for study coordinators and researchers who aim to leverage EHR data in clinical research, regardless of their institutional context.

The Precise Definition and Quality Assessment of Required Data Elements are Paramount

While interoperable EHR formats facilitate data exchange, they do not inherently resolve underlying data quality issues. Rather, standardization tends to illuminate existing quality challenges, making thorough preliminary data quality assessment essential. Researchers must realistically evaluate whether available EHR data meets the quality thresholds necessary for their specific research objectives.

Project Timelines Should Accommodate Iterative Development and Refinement Cycles

The notion of plug-and-play solutions for analyzing EHR data across different institutions proves unrealistic in practice. Scientific research inherently involves trial and error, particularly when working with heterogeneous EHR data sources. Successfully navigating these iterations requires sufficient time allocation in project planning.

Team Composition Significantly Influences Project Success

Effective distributed EHR analysis demands a multidisciplinary approach, combining team members with complementary expertise. Essential roles include local data stewards with intimate knowledge of institutional data structures, alongside specialists who understand various aspects of EHR data

utilization. This diversity of perspectives enables comprehensive problem-solving and robust methodology development.

Discussion

Principal Results

The HELP study was designed as a use case for the secondary use of EHR data, limited to 5 university hospitals participating in the MII in Germany. Therefore, the lessons learned in this paper apply specifically to this context. The study commenced in the early phases of DIC development, which introduced challenges that may be resolved over time. A key issue was the absence of mandatory data standards in Germany's health care, which complicated our efforts. However, the implementation of national standards is anticipated in the future [32], and such standards are essential for efficient multicenter secondary EHR data usage. In this regard, the DIC were pioneers in applying HL7 FHIR in German university hospitals. Nonetheless, the processing of FHIR remains subject to optimization, as we were unable to develop and apply project-specific information models and FHIR profiles for the HELP study as intended. This was largely due to the extensive workload required for profiling as well as data integration and transformation, which fell on small teams. In particular, not every EHR data use project can be expected to have the resources and expertise in FHIR profiling to be able to technically express project-specific data requirements.

For some phases of the study, other methods of data harmonization could also have been considered. For example, a Common Data Model can be used to harmonize cross-site datasets, with Observational Medical Outcomes Partnership (OMOP) [33] playing a particularly important role in the area of health data. However, due to the mandatory use of international terminologies and the incompatibility of national specifications or site-specific data properties, various relevant data items could not be represented in OMOP for the German HELP study either, as has already been shown for the MII CDS [34,35]. Examples include the vital status within 90 days, blood culture results, or hospital readmission after discharge, which would have required additional logic or custom extensions in OMOP, while FHIR offered more flexible options through resources like Observation and Encounter. Common Data Models, in turn, remain highly relevant for use in internationally conducted observational studies.

It is evident that the scalability of any comparable cross-facility research project is impaired by the divergent clinical documentation in facilities and their specialist departments, which can vary greatly in details due to different manufacturers, different billing forms, and individual decisions, especially complicating distributed analyses. The enhancement of secondary use can only be achieved through the implementation of billing-independent ontologies such as SNOMED CT for the semantic annotation of clinical data [17], as is envisaged for the European Health Data Space (EHDS [36]), for instance.

In the HELP study, data availability posed a significant challenge, particularly regarding patient movement between wards as well as medication records. These data types were often difficult to use due to their manual documentation and their frequent updates in many settings. A broader issue was the lack of standardized formats for certain data, with some not being available digitally at all. To address these limitations, we supplemented the existing data collection with additional eCRF. However, during the course of the study, digitization of clinical documentation increased, offering hope that such challenges may be mitigated in future research.

In addition to data availability challenges, we encountered significant data quality issues during both data transformation to interoperable standards and subsequent analysis. While several formal frameworks exist for assessing and improving EHR data quality—notably the framework proposed by Kahn et al [37], which has gained traction in the MII [38,39]—these were not yet robustly established for FHIR-based data at the time of our study. Our experience underscores the critical need for standardized data quality assessment methodologies, and future research in the field should systematically explore and implement comprehensive frameworks like that proposed by Kahn et al.

While a distributed analysis approach, as used in our study, is not statistically ideal—centralized, patient-level analyses provide greater flexibility and statistical power—it was the only feasible method given the ethical and data protection constraints induced by not being able to obtain informed consent from every participant. In the future, our manual approach to distributed analysis will likely be replaced by more formal frameworks for distributed analysis of EHR data, incorporating FHIR-based APIs. Examples include the Personal Health Train [40] and the Federated Learning and Analysis in Medicine platform [41]. These platforms, however, were not available at the time of the HELP study [42].

Comparison to Other EHR Data Applications

Since the start of the HELP use case, several other use cases have emerged within the MII. Zoch et al [43] highlight

lessons learned from studies on rare diseases conducted in the MII context, many of which align with our findings. For instance, they also emphasize the importance of multidisciplinary teams in successful data use projects and the need for a solid understanding of EHR data sources. Building on their insights, we took a broader view of secondary data use, abstracting beyond the specific MII structures. Additionally, we place greater emphasis on the iterative nature required for projects of this kind.

Extending the scope beyond Germany, many countries have undertaken initiatives similar to the MII to develop infrastructures for the secondary use of EHR data. Notable examples include the Dutch Health-RI initiative, the Swiss Personalized Health Network, and the US National Patient-Centered Clinical Research Network. These initiatives share a common goal: to collect and standardize EHR data for research purposes. In Europe, these efforts will soon be consolidated under the EHDS [36], which aims to create a unified infrastructure for the secondary use of EHR data across the continent. It is our hope to share the lessons learned from this project and use them to shape the integration efforts of the various approaches.

Conclusions

In this paper, we have shared key insights from conducting a clinical trial that leveraged EHR data. By applying these lessons and fostering cross-institutional collaboration, we can advance toward more sophisticated and efficient use of interoperable EHR systems in clinical research. This approach, while challenging, has the potential to enhance our understanding of patient care and clinical outcomes, providing valuable insights that can inform both clinical practice and future research endeavors. As the field continues to evolve, the lessons learned from current implementations will serve as important guideposts for developing robust and sustainable approaches to EHR data integration in clinical research.

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (grant numbers for the 5 study centers are Jena University Hospital: 01ZZ1803C, Halle University Hospital: 01ZZ1803N, Leipzig University Hospital: 01ZZ1803D, Aachen University Hospital: 01ZZ1803B, and Essen University Hospital: 01ZZ1803P). The German Federal Ministry of Education and Research also funded part of the project as part of the funding program “quantum technologies – from basic research to market”; contract number: 13N15711.

Data Availability

The tabular catalog of items and HELP (Hospital-wide Electronic Medical Record Evaluated Computerized Decision Support System to Improve Outcomes of Patients with Staphylococcal Bloodstream Infection) dataset in ART-DECOR are openly available as cited in the manuscript. The *fhircrackr* package is also openly accessible and actively developed on GitHub [44].

Conflicts of Interest

None declared.

Checklist 1

i-CHECK-DH checklist. i-CHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations. [\[DOCX File \(Microsoft Word File\), 16 KB-Checklist 1\]](#)

References

1. Albashiti F, Thasler R, Wendt T, Bathelt F, Reinecke I, Schreiweis B. Data integration centers-from a concept in the Medical Informatics Initiative to its local implementation in the Network of University Medicine. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. Jun 2024;67(6):629-636. [doi: [10.1007/s00103-024-03879-5](https://doi.org/10.1007/s00103-024-03879-5)] [Medline: [38662020](https://pubmed.ncbi.nlm.nih.gov/38662020/)]
2. Gehring S, Eulenfeld R. German Medical Informatics Initiative: unlocking data for research and health care. Methods Inf Med. Jul 2018;57(S 01):e46-e49. [doi: [10.3414/ME18-13-0001](https://doi.org/10.3414/ME18-13-0001)] [Medline: [30016817](https://pubmed.ncbi.nlm.nih.gov/30016817/)]
3. Hagel S, Gantner J, Spreckelsen C, et al. Hospital-wide ELectionic medical record evaluated computerised decision support system to improve outcomes of Patients with staphylococcal bloodstream infection (HELP): study protocol for a multicentre stepped-wedge cluster randomised trial. BMJ Open. Feb 10, 2020;10(2):e033391. [doi: [10.1136/bmjopen-2019-033391](https://doi.org/10.1136/bmjopen-2019-033391)] [Medline: [32047014](https://pubmed.ncbi.nlm.nih.gov/32047014/)]
4. Pletz M, Hagel S, Kimmig A, et al. HELP CDSS v. 1.0.SMITH - smart medical information technology for healthcare. Zenodo; 2024. URL: <https://zenodo.org/records/10704513> [doi: [10.5281/zenodo.10704513](https://doi.org/10.5281/zenodo.10704513)]
5. van Berkel N, Clarkson MJ, Xiao G, et al. Dimensions of ecological validity for usability evaluations in clinical settings. J Biomed Inform. Oct 2020;110:103553. [doi: [10.1016/j.jbi.2020.103553](https://doi.org/10.1016/j.jbi.2020.103553)] [Medline: [32891762](https://pubmed.ncbi.nlm.nih.gov/32891762/)]
6. Mc Cord KA, Ewald H, Agarwal A, et al. Treatment effects in randomised trials using routinely collected data for outcome assessment versus traditional trials: meta-research study. Br Med J. Mar 3, 2021;372:n450. [doi: [10.1136/bmj.n450](https://doi.org/10.1136/bmj.n450)] [Medline: [33658187](https://pubmed.ncbi.nlm.nih.gov/33658187/)]
7. Mc Cord KA, Al-Shahi Salman R, Treweek S, et al. Routinely collected data for randomized trials: promises, barriers, and implications. Trials. Jan 11, 2018;19(1):29. [doi: [10.1186/s13063-017-2394-5](https://doi.org/10.1186/s13063-017-2394-5)] [Medline: [29325575](https://pubmed.ncbi.nlm.nih.gov/29325575/)]
8. Palm J, Alaid S, Ammon D, et al. Leveraging electronic medical records to evaluate a computerized decision support system for Staphylococcus bacteremia. NPJ Digit Med. Mar 28, 2025;8(1):180. [doi: [10.1038/s41746-025-01569-3](https://doi.org/10.1038/s41746-025-01569-3)] [Medline: [40148479](https://pubmed.ncbi.nlm.nih.gov/40148479/)]
9. HELP study group. HELP Study Data Dictionary/Catalog of Items. Zenodo; 2024. [doi: [10.5281/zenodo.13739833](https://doi.org/10.5281/zenodo.13739833)]
10. Schulz S, Stegwee R, Chronaki C. Standards in healthcare data. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of Clinical Data Science. Springer; 2019:19-36. [doi: [10.1007/978-3-319-99713-1_3](https://doi.org/10.1007/978-3-319-99713-1_3)] ISBN: 978-3-319-99712-4
11. Ammon D, Kurscheidt M, Buckow K, et al. Interoperability Working Group: core dataset and information systems for data integration and data exchange in the Medical Informatics Initiative. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. Jun 2024;67(6):656-667. [doi: [10.1007/s00103-024-03888-4](https://doi.org/10.1007/s00103-024-03888-4)] [Medline: [38753022](https://pubmed.ncbi.nlm.nih.gov/38753022/)]
12. Heitmann K. ART-DECOR: an open-source tool bridging the chasm between clinicians and health IT. Health Level Seven International. 2014. URL: http://www.hl7.org/documentcenter/public/newsletters/HL7_NEWS_20140908.pdf [Accessed 2025-07-21]
13. HELP study group. HELP dataset. ART-DECOR. 2018. URL: <https://art-decor.org/ad/#/help-/datasets/dataset/2.16.840.1.113883.3.1937.777.22.1.1/2018-05-07T19:20:17> [Accessed 2025-07-21]
14. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med. 2019;2:79. [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/)]
15. Sass J, Zabka S, Essenwanger A, Schepers J, Boeker M, Thun S. Fast Healthcare Interoperability Resources (FHIR®) representation of medication data derived from German Procedure Classification Codes (OPS) using identification of medicinal products (IDMP) compliant terminology. Stud Health Technol Inform. May 24, 2021;278:231-236. [doi: [10.3233/SHTI210074](https://doi.org/10.3233/SHTI210074)] [Medline: [34042899](https://pubmed.ncbi.nlm.nih.gov/34042899/)]
16. Rinaldi E, Drenkhahn C, Gebel B, et al. Towards interoperability in infection control: a standard data model for microbiology. Sci Data. Sep 23, 2023;10(1):654. [doi: [10.1038/s41597-023-02560-x](https://doi.org/10.1038/s41597-023-02560-x)] [Medline: [37741862](https://pubmed.ncbi.nlm.nih.gov/37741862/)]
17. Vorisek CN, Lehne M, Klopfenstein SAI, et al. Fast Healthcare Interoperability Resources (FHIR) for interoperability in health research: systematic review. JMIR Med Inform. Jul 19, 2022;10(7):e35724. [doi: [10.2196/35724](https://doi.org/10.2196/35724)] [Medline: [35852842](https://pubmed.ncbi.nlm.nih.gov/35852842/)]
18. Menzel F, Waltemath D, Henkel R. Exploring new possibilities for research data exploration using the example of the German Core Data Set. Stud Health Technol Inform. May 18, 2023;302:749-750. [doi: [10.3233/SHTI230255](https://doi.org/10.3233/SHTI230255)] [Medline: [37203485](https://pubmed.ncbi.nlm.nih.gov/37203485/)]
19. Ulrich H, Behrend P, Wiedekopf J, Drenkhahn C, Kock-Schoppenhauer AK, Ingenerf J. Hands on the Medical Informatics Initiative Core Data Set - lessons learned from converting the MIMIC-IV. Stud Health Technol Inform. Sep 21, 2021;283:119-126. [doi: [10.3233/SHTI210549](https://doi.org/10.3233/SHTI210549)] [Medline: [34545827](https://pubmed.ncbi.nlm.nih.gov/34545827/)]

20. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. Mar 15, 2016;3:160018. [doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)] [Medline: [26978244](https://pubmed.ncbi.nlm.nih.gov/26978244/)]
21. Möller M, Scharlau N, Klettke M. An empirical study of open data JSON files. Presented at: 23rd International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data; Mar 23, 2021; Nicosia, Cyprus. URL: <https://ceur-ws.org/Vol-2840/#short9> [Accessed 2025-07-21]
22. Yusof MK. Efficiency of JSON for data retrieval in Big Data. *Indonesian J Elect Eng Comput Sci*. 2017;7(1):250. [doi: [10.11591/ijeecs.v7.i1.pp250-262](https://doi.org/10.11591/ijeecs.v7.i1.pp250-262)]
23. Palm J, Meineke FA, Przybilla J, Peschel T. “fhircrackr”: An R Package Unlocking Fast Healthcare Interoperability Resources for Statistical Analysis. *Appl Clin Inform*. Jan 2023;14(1):54-64. [doi: [10.1055/s-0042-1760436](https://doi.org/10.1055/s-0042-1760436)] [Medline: [36696915](https://pubmed.ncbi.nlm.nih.gov/36696915/)]
24. The R project for statistical computing. URL: <https://www.r-project.org/> [Accessed 2025-07-21]
25. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *Br Med J*. Feb 6, 2015;350:h391. [doi: [10.1136/bmj.h391](https://doi.org/10.1136/bmj.h391)] [Medline: [25662947](https://pubmed.ncbi.nlm.nih.gov/25662947/)]
26. McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd ed. London, Chapman and Hall; 1989.
27. Hund H, Wettstein R, Kurscheidt M, Schweizer ST, Zilske C, Fegeler C. Interoperability is a process - the data sharing framework. *Stud Health Technol Inform*. Jan 25, 2024;310:28-32. [doi: [10.3233/SHTI230921](https://doi.org/10.3233/SHTI230921)] [Medline: [38269759](https://pubmed.ncbi.nlm.nih.gov/38269759/)]
28. Rosenau L, Behrend P, Wiedekopf J, Gruendner J, Ingenerf J. Uncovering harmonization potential in health care data through iterative refinement of fast healthcare interoperability resources profiles based on retrospective discrepancy analysis: case study. *JMIR Med Inform*. Jul 23, 2024;12:e57005. [doi: [10.2196/57005](https://doi.org/10.2196/57005)] [Medline: [39042420](https://pubmed.ncbi.nlm.nih.gov/39042420/)]
29. Shah SM, Khan RA. Secondary use of electronic health record: opportunities and challenges. *IEEE Access*. 2020;8:136947-136965. [doi: [10.1109/ACCESS.2020.3011099](https://doi.org/10.1109/ACCESS.2020.3011099)]
30. Nguyen A. *Hands-On Healthcare Data: Taming the Complexity of Real-World*. 1st ed. O'Reilly Media; 2022. ISBN: 978-1098112929
31. Marcon Y, Bishop T, Avraam D, et al. Orchestrating privacy-protected big data analyses of data from different resources with R and DataSHIELD. *PLoS Comput Biol*. Mar 2021;17(3):e1008880. [doi: [10.1371/journal.pcbi.1008880](https://doi.org/10.1371/journal.pcbi.1008880)] [Medline: [33784300](https://pubmed.ncbi.nlm.nih.gov/33784300/)]
32. Fuhrmann F, Adjei B, Hartge F. Interop roadmap 2024-2025. INA. 2024. URL: <https://www.ina.gematik.de/en/technical-topics-and-roadmap> [Accessed 2024-10-07]
33. Stang PE, Ryan PB, Racoosin JA, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann Intern Med*. Nov 2, 2010;153(9):600-606. [doi: [10.7326/0003-4819-153-9-201011020-00010](https://doi.org/10.7326/0003-4819-153-9-201011020-00010)] [Medline: [21041580](https://pubmed.ncbi.nlm.nih.gov/21041580/)]
34. Maier C, Lang L, Storf H, et al. Towards implementation of OMOP in a German University Hospital Consortium. *Appl Clin Inform*. Jan 2018;9(1):54-61. [doi: [10.1055/s-0037-1617452](https://doi.org/10.1055/s-0037-1617452)] [Medline: [29365340](https://pubmed.ncbi.nlm.nih.gov/29365340/)]
35. 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*. Jan 2023;169:104925. [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/)]
36. Directorate-General for Health and Food Safety. Proposal for a regulation on the European health data space. EUR-Lex. 2022. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022PC0197> [Accessed 2025-07-21]
37. Kahn MG, Callahan TJ, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *EGEMS (Wash DC)*. 2016;4(1):1244. [doi: [10.13063/2327-9214.1244](https://doi.org/10.13063/2327-9214.1244)] [Medline: [27713905](https://pubmed.ncbi.nlm.nih.gov/27713905/)]
38. Draeger C, Tute E, Schmidt CO, et al. Identifying relevant FHIR elements for data quality assessment in the German Core Data Set. *Stud Health Technol Inform*. May 18, 2023;302:272-276. [doi: [10.3233/SHTI230117](https://doi.org/10.3233/SHTI230117)] [Medline: [37203661](https://pubmed.ncbi.nlm.nih.gov/37203661/)]
39. 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*. Jan 14, 2022;289:240-243. [doi: [10.3233/SHTI210904](https://doi.org/10.3233/SHTI210904)] [Medline: [35062137](https://pubmed.ncbi.nlm.nih.gov/35062137/)]
40. Mou Y, Li F, Weber S, et al. Distributed privacy-preserving data analysis in NFDI4Health with the personal health train. *Proc Conf Res Data Infrastr*. 2023;1:1. [doi: [10.52825/cordi.v1i.282](https://doi.org/10.52825/cordi.v1i.282)]
41. PrivateAIM. URL: <https://privateaim.de/> [Accessed 2024-10-08]
42. Wirth FN, Meurers T, Johns M, Prasser F. Privacy-preserving data sharing infrastructures for medical research: systematization and comparison. *BMC Med Inform Decis Mak*. Aug 12, 2021;21(1):242. [doi: [10.1186/s12911-021-01602-x](https://doi.org/10.1186/s12911-021-01602-x)] [Medline: [34384406](https://pubmed.ncbi.nlm.nih.gov/34384406/)]

43. Zoch M, Gierschner C, Andreeff AK, et al. Secondary use of patient data within decentralized studies using the example of rare diseases in Germany: a data scientist's exploration of process and lessons learned. Digit Health. 2024;10:20552076241265219. [doi: [10.1177/20552076241265219](https://doi.org/10.1177/20552076241265219)] [Medline: [39130526](https://pubmed.ncbi.nlm.nih.gov/39130526/)]
44. POLAR-fhir/fhircrackr. GitHub. URL: <https://github.com/polar-fhir/fhircrackr> [Accessed 2025-07-21]

Abbreviations

CDS: core dataset

CDSS: computerized decision support system

COI: catalog of items

DIC: data integration center

eCRF: electronic case report form

EHR: electronic health record

ETL: extract-transform-load

FHIR: Fast Healthcare Interoperability Resources

HELP: Hospital-wide Electronic Medical Record Evaluated Computerized Decision Support System to Improve Outcomes of Patients with Staphylococcal Bloodstream Infection

HL7: Health Level Seven International

MII: Medical Informatics Initiative

OMOP: Observational Medical Outcomes Partnership

Edited by Jeffrey Klann; peer-reviewed by Markus Wolfien, Martin Boeker, Sunday Oworah; submitted 30.10.2024; final revised version received 06.05.2025; accepted 15.05.2025; published 30.07.2025

Please cite as:

Palm J, Saleh K, Scherag A, Ammon D

Leveraging Interoperable Electronic Health Record (EHR) Data for Distributed Analyses in Clinical Research: Technical Implementation Report of the HELP Study

JMIR Med Inform 2025;13:e68171

URL: <https://medinform.jmir.org/2025/1/e68171>

doi: [10.2196/68171](https://doi.org/10.2196/68171)

© Julia Palm, Kutaiba Saleh, André Scherag, Danny Ammon. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 30.07.2025. 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 JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.