

1 **Streamlining Intersectoral Provision of Real-World Health Data: A** 2 **Service Platform for Improved Clinical Research and Patient Care**

3 **Katja Hoffmann^{1*}, Igor Nesterow¹, Yuan Peng¹, Elisa Henke¹, Daniela Barnett³, Cigdem**
4 **Klengel¹, Mirko Gruhl¹, Martin Bartos², Frank Nüßler², Richard Gebler¹, Sophia Grummt¹,**
5 **Anne Seim¹, Franziska Bathelt⁴, Ines Reinecke³, Markus Wolfien^{1,5}, Jens Weidner¹, Martin**
6 **Sedlmayr¹**

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

9 ²Department of Information Technology, Klinikum Chemnitz gGmbH, Chemnitz, Germany

10 ³Data Integration Center, Center for Medical Informatics, University Hospital Carl Gustav Carus,
11 Dresden, Germany

12 ⁴Thiem-Research GmbH, Cottbus, Germany

13 ⁵Center for Scalable Data Analytics and Artificial Intelligence, Dresden/Leipzig, Germany

14 * **Correspondence:**

15 Katja Hoffmann

16 katja.hoffmann@tu-dresden.de

17 **Keywords: Service Platform, Intersectoral Data Sharing, Health Data Space, Real World Data,**
18 **Interoperability, Data harmonization, Research Data Infrastructure, Secondary Use of Clinical**
19 **Data**

20 **Abstract**

21 **Introduction:** Obtaining real-world data from routine clinical care is of growing interest for scientific
22 research and personalized medicine. Despite the abundance of medical data across various facilities —
23 including hospitals, outpatient clinics, and physician practices — the intersectoral exchange of
24 information remains largely hindered due to differences in data structure, content, and adherence to
25 data protection regulations. In response to this challenge, the Medical Informatics Initiative (MII) was
26 launched in Germany, focusing initially on university hospitals to foster the exchange and utilization
27 of real-world data through the development of standardized methods and tools, including the creation
28 of a common core dataset. Our aim, as part of the Medical Informatics Research Hub in Saxony

29 (MiHUBx), is to extend the MII concepts to non-university healthcare providers in a more seamless
30 manner to enable the exchange of real-world data among intersectoral medical sites.

31 **Methods:** We investigated what services are needed to facilitate the provision of harmonized real-
32 world data for cross-site research. On this basis, we designed a Service Platform Prototype that hosts
33 services for data harmonization, adhering to the globally recognized Health Level 7 (HL7) Fast
34 Healthcare Interoperability Resources (FHIR) international standard communication format and the
35 Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). Leveraging
36 these standards, we implemented additional services facilitating data utilization, exchange and
37 analysis. Throughout the development phase, we collaborated with an interdisciplinary team of experts
38 from the fields of system administration, software engineering and technology acceptance to ensure
39 that the solution is sustainable and reusable in the long term.

40 **Results:** We have developed the pre-built packages “ResearchData-to-FHIR”, “FHIR-to-OMOP” and
41 “Addons”, which provide the services for data harmonization and provision of project-related real-
42 world data in both the FHIR MII Core dataset format (CDS) and the OMOP CDM format as well as
43 utilization and a Service Platform Prototype to streamline data management and use.

44 **Conclusion:** Our development shows a possible approach to extend the MII concepts to non-university
45 healthcare providers to enable cross-site research on real-world data. Our Service Platform Prototype
46 can thus pave the way for intersectoral data sharing, federated analysis, and provision of SMART-on-
47 FHIR applications to support clinical decision making.

48 1 Introduction

49 For scientific research, there is a high interest in using data from routine clinical care, so-called real-
50 world data (RWD) (1–4). Although large amounts of RWD are available in various institutions, such
51 as hospitals, outpatient clinics, and physician practices, the intersectoral data exchange between sites
52 is hindered by their heterogeneity in terms of structure, content, and compliance with data protection
53 regulations (2,5). To address this challenge, the German Medical Informatics Initiative (MII) was
54 launched in 2018, initially focusing on university hospitals to foster the exchange and utilization of
55 RWD (6,7). At that time, Data Integration Centers (DIC) were established at the medical sites of the
56 university hospitals, and standardized solutions were developed for effective data use and exchange in
57 both healthcare and research, with a focus on interoperability and data harmonization. For example, all
58 university hospitals in the MII defined a dataset description, the MII Core dataset (CDS) (8) using the

59 Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR) international standard
60 communication format (9). The FHIR MII CDS consists of basic modules (e.g., Person, Case,
61 Diagnosis, Procedure, Laboratory test results, medication), and extension modules (e.g., Oncology,
62 Pathology results, Molecular Genetics, Intensive care) (8). This forms the foundation for cross-site data
63 exchange and the integration of third-party applications via SMART-on-FHIR technology (10).

64 To facilitate the data exchange of RWD for scientific purposes, the HL7 International (11) and the
65 Observational Health Data Sciences and Informatics (OHDSI) (12,13) community have announced
66 collaboration in 2021 (14). The OHDSI community develops the Observational Medical Outcomes
67 Partnership (OMOP) common data model (CDM) (15), as well as tools for data quality assessment and
68 analysis (16). Common to all standardized data formats is the need to develop individual processes for
69 the extraction, transformation and loading (ETL) of data from different data sources, which remains a
70 major challenge (5). Nevertheless, hospitals have succeeded in overcoming these initial hurdles and
71 making their own RWD available in a harmonized form for research (17,18) and possibly also for
72 patient care.

73 Since certain diseases, such as cardiovascular diseases, diabetes, allergies, and mental illnesses, are
74 often not treated at university hospitals and there is insufficient data available, especially for rare
75 diseases, it is essential to obtain RWD as well from non-university healthcare providers to ensure a
76 comprehensive and diverse dataset for research studies. In order to make RWD available for research
77 by non-university healthcare providers, the Medical Informatics Hub in Saxony (MiHUBx) was
78 founded in 2021 (19). Among the project goals is to investigate whether the concepts of the MII can
79 be transferred to non-university healthcare providers. As a first result within this project, Bathelt et al.
80 (2022) demonstrated the possibility to utilize an existing portable and standardized infrastructure from
81 a university hospital setting and transferred it to non-university sites to support feasibility requests for
82 participation in multicentre studies (1). However, the work of Bathelt et al. was limited in terms of data
83 availability, terminologies used, and harmonized data formats, so that RWD from non-university sites
84 are still insufficiently available for research. Yet, due to limited human and economic resources and
85 expertise in HL7 FHIR and OMOP CDM, it is hardly possible for non-university healthcare providers
86 to develop the required services for data harmonization and provision in standardized data formats
87 themselves. It is therefore necessary to provide the services in such a way that they can be conveniently
88 deployed and easily used by data providers for different studies. In this paper, all tools for data

89 harmonization, provision and management, such as databases, programs for ETL processes, analysis
90 tools and other applications, are referred to as services.

91 The aim of this work is the development of pre-built packages that contain services and support the
92 harmonization and provision of RWD in FHIR MII CDS and OMOP CDM format and thereby provide
93 reuse potential for various projects. Another goal of this work is the development of a versatile and
94 modular Service Platform Prototype that facilitates project administration, service management, data
95 management and analysis. In this context, we focus on the following two research questions:

- 96 1. What services are needed to facilitate the provision of harmonized RWD for cross-site
97 research?
- 98 2. How can the necessary services (from 1) be technically compiled so that they can be used by
99 hospitals with few resources and limited expertise in HL7 FHIR and OMOP CDM to make
100 harmonized RWD accessible for research studies?

101 **2 Materials and methods**

102 **2.1 Materials for data harmonization and provision**

103 **2.1.1 Compilation and integrated implementation of existing software resources**

104 For the harmonization of RWD from source systems into the basic modules of the FHIR MII CDS
105 version 1.0 (20), and into the OMOP CDM format (15), we have selected the following materials, as
106 these are successfully used at the University Hospital Dresden. In addition, we investigated the
107 applicability of other established tools. For this purpose, we used the following developed concepts of
108 the MII and the software tools published by the MII consortium Medical Informatics in Research and
109 Care in University Medicine (MIRACUM) (21) as MIRACOLIX Tools (22):

110 **Clinical Data Repository:** A Clinical Data Repository (CDR) is a database in which patient-centered
111 healthcare data from various IT systems (e.g., electronic health record (EHR), laboratory system,
112 biobank) are stored in a site-specific data model.

113 **FHIR Server BLAZE:** A FHIR Server is a software solution that stores and manages FHIR resources.
114 It acts as a bridge connecting healthcare applications and systems, allowing them to exchange patient
115 information in a consistent and structured format or to answer population-wide aggregate queries
116 quickly. The FHIR Server BLAZE (23) was initially developed within the German Biobank Alliance

117 project (24), aimed at high-throughput performance (25). Blaze comes with a built-in feature to
118 authenticate requests against an OpenID Connect provider.

119 **MIRACUM FHIR Gateway:** The MIRACUM FHIR Gateway (26) is a PostgreSQL database with a
120 table for storing FHIR Resources, which are represented in JSON format. It serves as a temporary
121 storage.

122 **ETL Process DWH-TO-FHIR:** The ETL Process DWH-TO-FHIR extracts research data, which is
123 provided as database views, from the data warehouse (DWH) of the site, constructs FHIR resources
124 according to the FHIR MII CDS structure definition v1.0 and loads them into a *MIRACUM FHIR*
125 *Gateway*. The application DWH-TO-FHIR provides Basic Authentication (27) to enforce access
126 controls to the FHIR resources of the FHIR server.

127 **ETL Process ROTATOR:** The ROTATOR (fROm gaTewAy TO seRver) is an application that reads
128 FHIR resources from the MIRACUM FHIR Gateway and loads them onto a FHIR Server, such as
129 BLAZE.

130 **OHDSI Tools:** The OHDSI Tools are provided as Docker containers by Gruhl et al. (28). The basis of
131 the OHDSI Tools is the OMOP CDM PostgreSQL database, which is divided into the standardized
132 OMOP CDM data tables v5.3 and the OMOP CDM standardized vocabularies (state February 2023),
133 such as SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms), ICD-10-GM
134 (International Classification of Diseases, German Modification) and OPS (“Operationen- und
135 Prozedurenschlüssel”, surgery and procedure key). Furthermore, the OHDSI Tools by Gruhl et al. (28)
136 includes additional dockerized tools for analyzing research data in OMOP CDM: 1) the R-based
137 application ACHILLES that can be used for data characterization and visualization, 2) the web-based
138 application ATLAS that can be used for cohort definition and scientific analysis, and 3) the R-based
139 application Data Quality Dashboard (DQD) that can be used for data quality analysis.

140 **ETL Process FHIR-TO-OMOP:** The ETL Process FHIR-TO-OMOP extracts FHIR resources from
141 a FHIR Server (e.g., the FHIR Server BLAZE) or from the MIRACUM FHIR Gateway, transforms
142 them into the standardized format of OMOP CDM and loads them into an OMOP CDM database
143 (29,30). The application provides HTTP Basic Authentication (27) to enforce access controls to the
144 FHIR resources of the FHIR server.

145 **TRANSITION Database:** The TRANSITION Database is a relational database that was originally
146 developed for the semantic mapping of system-specific documented diagnoses to Orpha codes for rare
147 diseases (23). It is deployed as a PostgreSQL database v14 (22), and offers tables with required
148 terminology bindings (e.g., FHIR value sets), which can essentially support the semantic mapping of
149 the RWD to required code systems. For example, the semantic mapping of the gender (e.g. value "1"
150 for female) to a FHIR ValueSet (e.g. code "female" (24)) can be achieved via the TRANSITION
151 Database. As an example of the database structure, the table for the semantic mapping of the vital
152 parameters can be found as a csv file in the Supplementary Materials (S1 file).

153 **Keycloak:** Since medical data needs to be protected in an enhanced manner, the Keycloak server offers
154 significant advantages (e.g. centralized Identity Management, Customizable Authentication Flows) for
155 the medical domain, particularly in terms of securing sensitive patient data, ensuring compliance with
156 healthcare regulations, and facilitating interoperability among diverse healthcare IT systems. The
157 Authentication Server Keycloak (34) can be used as an OpenID (35), OAuth 2.0 (36), or SAML
158 Connect provider to validate requests to the FHIR endpoints.

159 **2.1.2 Original Contributions - Expanding with Custom Software Additions**

160 We analyzed the already established data provision pipeline at University Hospital Dresden by
161 consulting experts in the fields of data integration, provision, protection and security. From this, we
162 derived a generic process for data harmonization and provision. Subsequently, we identified missing
163 materials, which we have developed and adapted. These are technically described below and further
164 explained in the Results Section.

165 **Research Data Repository:** In order to be able to provide the ETL process DWH-TO-FHIR to other
166 sites, a database structure is required which represents medical data and follows the logic of the basic
167 modules of the FHIR MII CDS specification v1.0 (20). For this reason, we developed the Research
168 Data Repository (RDR), deployed as a PostgreSQL database v14 (32). This repository encompasses
169 tables for various FHIR resource types, like *Patient*, *Encounter*, *Condition*, *Observation*, *Procedure*,
170 *Medication*, and *MedicationAdministration*, and is used for the cross-project and project-related
171 storage of research data from routine clinical care (RWD).

172 **Structural Mapping Guideline MII CDS:** To reduce implementation time and to facilitate the
173 specific data mapping, which must be done at each site, the Structural Mapping Guideline MII CDS is
174 developed (S2 File).

175 **The ETL process RDR-to-FHIR:** To streamline the ETL process to provide RWD in FHIR MII CDS
176 format, the sub-processes DWH-to-FHIR and ROTATOR were merged into the ETL process RDR-to-
177 FHIR. RDR-to-FHIR loads project-related RWD from the RDR, constructs FHIR resources according
178 to the FHIR MII CDS specification v1.0 and loads directly into the FHIR Server without buffering via
179 the MIRACUM FHIR Gateway. To facilitate process execution, a RESTful API client was
180 implemented to send data to the FHIR API. To enable OAuth 2.0 (36) authentication against the FHIR
181 server BLAZE, we have added this authentication method to the application.

182 **ETL Process FHIR-TO-OMOP:** To enable OAuth 2.0 authentication against the FHIR server
183 BLAZE, we added this authentication method to this application.

184 **2.2 Pre-built Packages for data harmonization and provision**

185 We have developed pre-built packages, e.g. for the transformation of RWD to FHIR, to facilitate
186 creating, running and connecting the multiple services (see section 2.1). In close collaboration with an
187 interdisciplinary team of software developers, database engineers as well as infrastructure, security,
188 and usability experts, we defined specific applications and composed the services to easy-to-install,
189 pre-configurable installation packages based on Docker v24 (37) and Docker Compose v2 (38). We
190 made our decisions based on previous experience and preliminary works. For testing purposes, we
191 provided a test dataset.

192 To facilitate data and system administration, we have included the following additional materials in
193 our packages, which can be used optionally.

194 **pgAdmin:** The open-source administration and management tool pgAdmin4 (39) provides a user-
195 friendly web interface for the efficient administration of PostgreSQL databases used for RDR,
196 TRANSITION database, Keycloak, and OMOP CDM.

197 **Portainer:** The open source container administration tool Portainer community edition (40) simplifies
198 the management of containerized applications that are used to provide all the services described in the
199 section 2.1. Portainer provides a web interface for interacting with Docker containers (37) to create,
200 manage, and deploy containers without the need for extensive knowledge of Docker commands.

201

202 **2.3 Service Platform Prototype**

203 To support research projects and system administration, we have developed a web-based Service
204 Platform Prototype that facilitates project-related data provision for research projects. Based on
205 experience with the design of user interfaces for interactive systems, an initial Low-Fidelity Prototype
206 of the service platform was created to serve as a basis for discussion for a joint, interdisciplinary
207 Graphical User Interface (GUI) design. By incorporating feedback from usability experts, this Low-
208 Fidelity Prototype was iteratively transformed into a High-Fidelity Prototype. To this end, regular joint
209 meetings were held with software developers and usability experts to ensure that the interaction
210 principles for design solutions in accordance with ISO 9241-110 (41) were taken into account and
211 implemented in the service platform interface (task appropriateness, self-descriptiveness, conformity
212 to expectations, learnability, controllability, robustness against errors, user retention).

213 Our Service Platform Prototype was developed using Quarkus v3.4.3 (42). The frontend provides a
214 form-based interface for users to specify the services and configurations needed for their project. The
215 backend, managing the service deployment, is responsible for the instantiation of Docker containers,
216 network setup, configuration management, and error handling. Angular v16.2.8 (43) was used for the
217 frontend, complemented by Angular Bootstrap v16.0.0 (44). Communication between the
218 microservices (e.g., to manage projects and Docker containers) is orchestrated by the Java Docker API
219 v3.3.3 (45). Communication between the services is facilitated by RabbitMQ (46), a message broker
220 that ensures reliable and real-time messaging.

221 **3 Results**

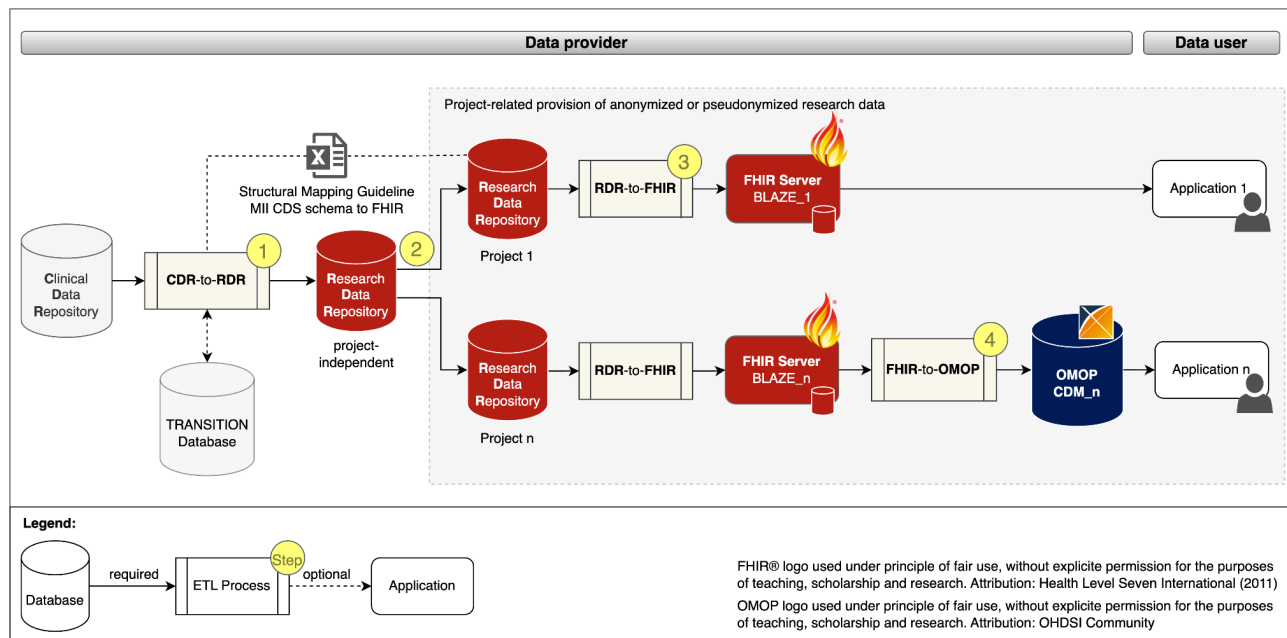
222 **3.1 A generic data harmonization and provision process to foster data availability**

223 We developed a generic process for the project-related provision of research data, which is shown in
224 Figure 1 and described below.

225 The data provision process begins with a project-independent and site-specific ETL pre-process CDR-
226 TO-RDR (cf. Figure 1 (Step 1)) that extracts RWD from the CDR, transforms it with the help of the
227 *Structural Mapping Guideline* and the *TRANSITION Database* to the data structure of the *Research*
228 *Data Repository (RDR)* and loads it into a *project-independent* instance of the *RDR*.

229 According to the MII concepts, the organizational starting point for the provision of data for a research
 230 project is a request from the data user. After the technical and legal feasibility check, the required data
 231 for the study is requested from the *project-independent RDR* instance and masked, pseudonymized or
 232 anonymized in the subsequent process step to ensure that the research data does not allow any
 233 conclusions to be drawn about the identity of a patient (Figure 1 (Step 2)).

234 To provide the project-related data in FHIR MII CDS format, instances of *RDR-TO-FHIR* and the
 235 *FHIR Server BLAZE* are created and the ETL process is executed (Figure 1 (Step 3)). To provide the
 236 project-related data in the OMOP CDM format, instances of *FHIR-TO-OMOP* and the *OHDSI tools*
 237 (cf. OHDSI tools) are created and the ETL process is executed (Figure 1 (Step 4)).



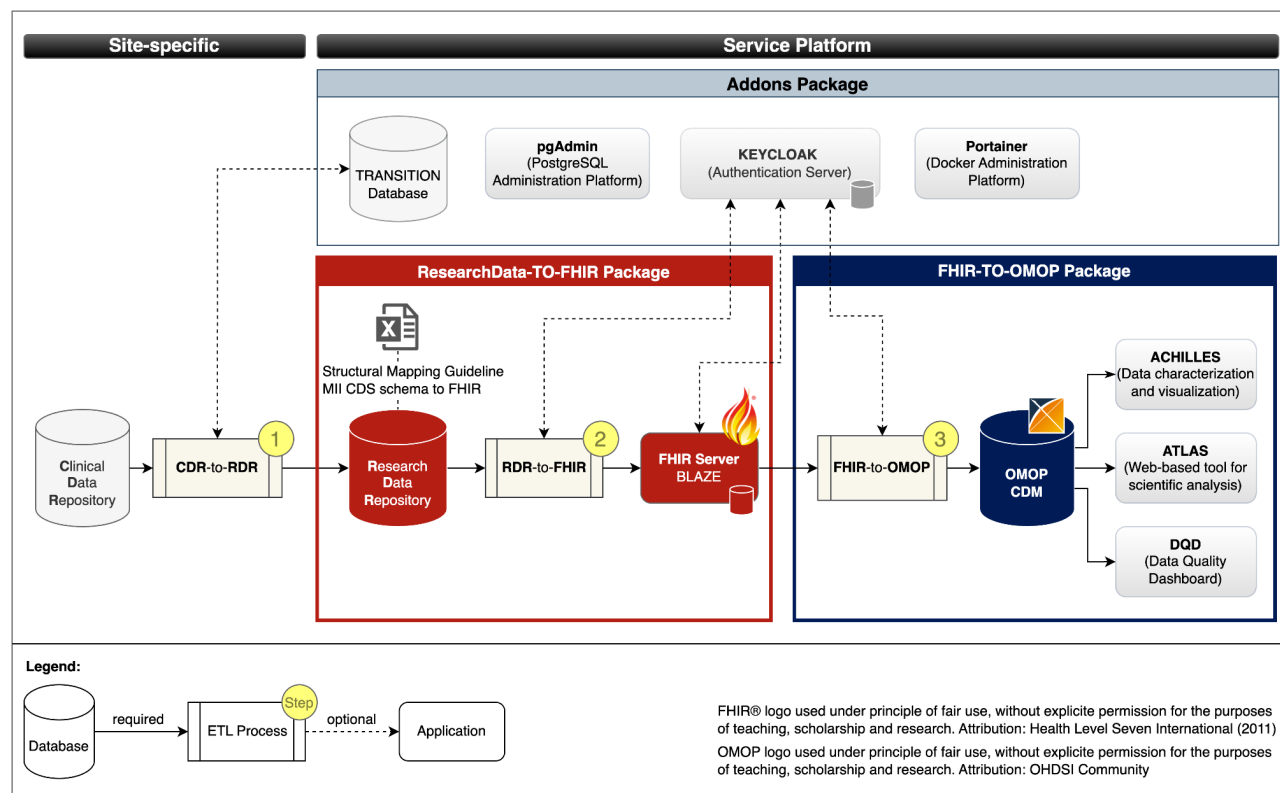
238

239 **Figure 1. Project-related provision of research data for providers and users.** Fast Healthcare Interoperability Resources
 240 (FHIR); Observational Medical Outcomes Partnership (OMOP); Common Data Model (CDM); Extract Transform Load
 241 (ETL); Clinical Data Repository (CDR); Research Data Repository (RDR); Medical Informatics Initiative (MII); Core
 242 dataset (CDS)

243 3.2 Pre-built Packages to facilitate interoperability on the fly

244 We developed the following three pre-build packages: 1) ResearchData-TO-FHIR Package for
 245 converting RWD to FHIR resources (cf. red bordered box in Figure 2), 2) FHIR-TO-OMOP Package
 246 for converting the FHIR resources to the OMOP CDM format (cf. blue bordered box in Figure 2), and
 247 3) the Addons Package for deployment of optional services to simplify database, container, and security

248 management (cf. gray bordered box in Figure 2). Figure 2 illustrates the composition of our
 249 implemented services and the data flows.



250

251 **Figure 2. Services and data flows.** Fast Healthcare Interoperability Resources (FHIR); Observational Medical Outcomes
 252 Partnership (OMOP); Common Data Model (CDM); Extract Transform Load (ETL); Clinical Data Repository (CDR);
 253 Research Data Repository (RDR): Medical Informatics Initiative (MII); Core dataset (CDS)

254 **ResearchData-TO-FHIR Package:** The ResearchData-TO-FHIR Package v2.2.0 (36) provides a
 255 Docker compose specification that allows to automatically retrieve the images of the *RDR*, the *FHIR*
 256 *server BLAZE*, and the *RDR-to-FHIR* (cf. section 2.1 and red bordered box in Figure 3), creates Docker
 257 containers as instances of the images that can be used to provide project-related research data in FHIR
 258 MII CDS format. For testing purposes, the package provides a test dataset. The package also includes
 259 the Structural Mapping Guideline (cf. Research Data Repository). By adjusting environment variables,
 260 the installation is customizable.

261 **FHIR-TO-OMOP Package:** The FHIR-TO-OMOP Package v1.1.0 (37) provides a Docker compose
 262 specification that allows to automatically retrieve the images of the *OHDSI Tools* and the *FHIR-TO-*
 263 *OMOP* (cf. section 2.1 and blue bordered box in Figure 3), creates Docker containers as instances of
 264 the images that can be used to provide project-related research data in OMOP CDM format. Through

265 environment variables, the installation is customizable and the synthetic dataset from the
266 ResearchData-TO-FHIR package can be used for testing purposes.

267 **Addons Package:** The Addons Package v1.0.0 (38) provides Docker compose specifications that
268 allow to optionally deploy the PostgreSQL administration platform *pgAdmin*, the *TRANSITION*
269 *database*, the container administration tool *Portainer* and the Authentication server *Keycloak* with a
270 demo configuration for protecting the FHIR Server Blaze (cf. ResearchData-TO-FHIR Package).

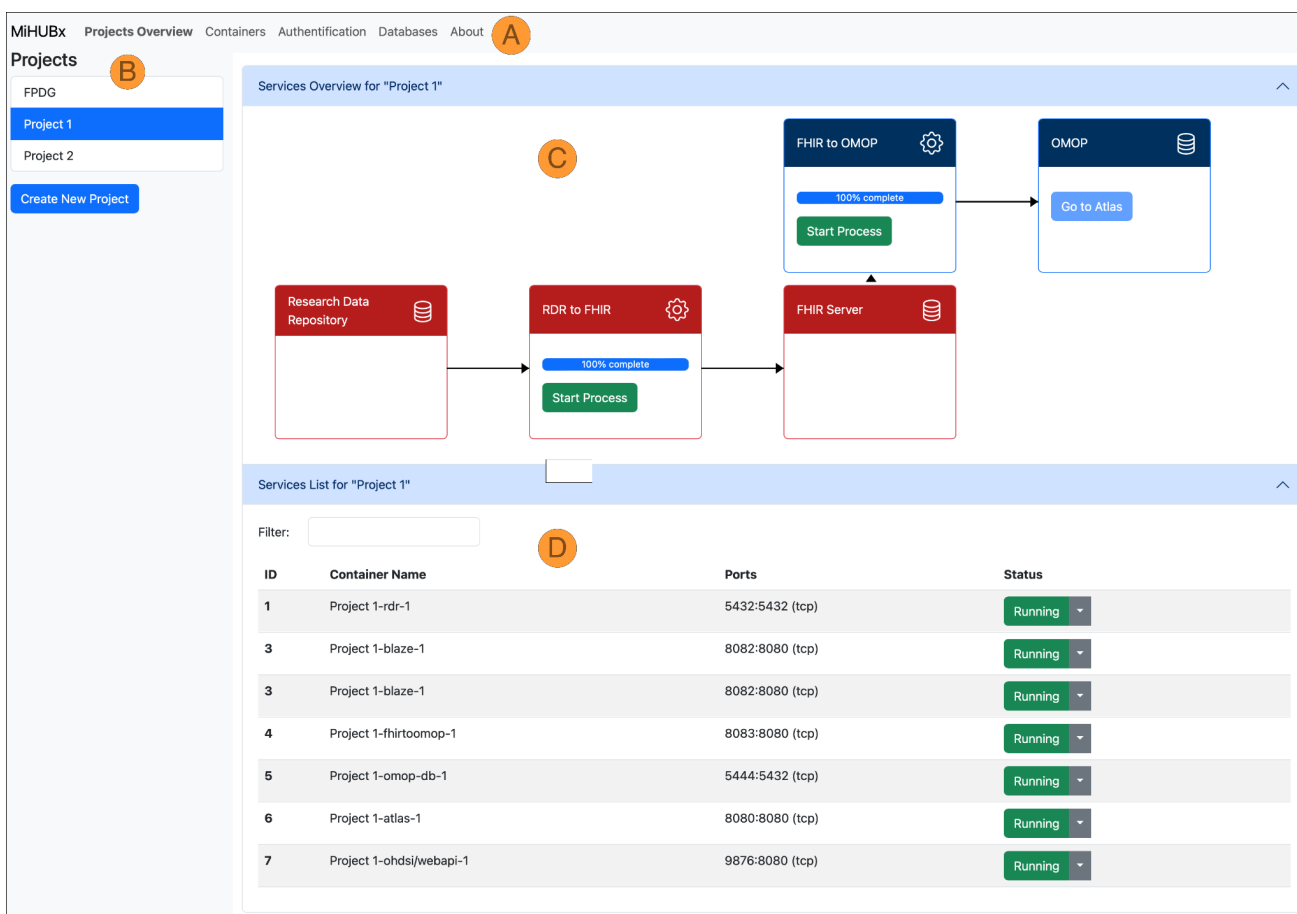
271 The source code of the ResearchData-TO-FHIR package v2.2.1 (47), the FHIR-TO-OMOP package
272 v1.1.0 (48) and the Addons package v1.0.0 (49) includes instructions for installation and usage as well
273 as for further developments.

274 **3.3. Service Platform Prototype to enable an easy to use and modular** 275 **infrastructure**

276 While the pre-built packages and their services, which are integrated into the Service Platform
277 Prototype, were described in 3.2, the Service Platform Prototype v1.0.0 (50) itself is described below.

278 We implemented two key services: 1) the Project Management Service and 2) the Container
279 Management Service. The Project Management Service allows researchers to create and manage their
280 research projects. The Container Management Service automates the installation, operation and
281 administration of services for each research project, such as the launch of ETL-processes and the
282 display of technical details, such as unique identifiers, Docker names, assigned communication ports
283 or operational status, to inform users about the current activity and availability of services.

284 The frontend layout contains a navigation bar (Figure 3, A), a sidebar (Figure 3, B) and a main content
285 area (Figure 3, C+D). The navigation bar offers links to subpages that provide services and further
286 information on database administration (i.e. via *pgAdmin*), authentication (i.e. *Keycloak*) and container
287 administration (i.e. *Portainer*). The research projects are listed in the sidebar and new projects can be
288 created via clicking on the respective button. The main content area is divided horizontally. The upper
289 area displays the project-related services and offers functions for use and management, such as starting
290 and stopping the Docker containers, starting the ETL processes, accessing the web frontend of the
291 services and receiving further information (Figure 3, C). The project-related containers are listed in the
292 lower area, where further technical details are displayed, e.g. information on Docker containers,
293 images, communication ports and operating status (Figure 3, D).



294

295 **Figure 3. Schematic representation of the Service Platform Prototype.** Fast Healthcare Interoperability Resources
296 (FHIR); Observational Medical Outcomes Partnership (OMOP); Research Data Repository (RDR); Identifier (ID); Medical
297 Informatics Hub in Saxony (MiHUBx)

298 The demo server, as an instance of the Service Platform Prototype, is available at [https://tu-](https://tu-dresden.de/med/demoserver)
299 [dresden.de/med/demoserver](https://tu-dresden.de/med/demoserver). This repository (50) also contains the developer documentation,
300 including initial installation instructions.

301 4 Discussion

302 Our aim was to determine which services and processes are required to enable the provision of RWD
303 for cross-site research and patient care, and how the services can be made available and usable to non-
304 university healthcare providers with limited resources in a low-threshold manner. In particular, we
305 determined which services are required for the harmonization of healthcare data into HL7 FHIR
306 standard-based MII CDS format for intersectoral exchange, and to the OMOP CDM format for national
307 or even international cross-site research. From this, we derived a process for the project-related
308 provision of RWD based on the already established data provision pipelines of the university sites (21)

309 in Germany. On this basis, we developed the pre-built packages *ResearchData-TO-FHIR*, *FHIR-TO-*
310 *OMOP*, and *Addons* for data harmonization and provision to facilitate interoperability on the fly. To
311 streamline the use of the particular services, especially for non-university healthcare providers, we
312 developed a versatile and modular Service Platform Prototype that demonstrates the administration of
313 research projects based on RWD.

314 To the best of our knowledge, our work is the first that shows how services for data harmonization,
315 provision, and analytics can be provided to non-university healthcare providers in a low-threshold
316 manner. The proposed Pre-built Packages and Service Platform Prototype streamline the process of
317 setting up research project environments and reduce the time and technical expertise required to
318 provide RWD for research studies and feasibility inquiries, such as those conducted by the German
319 Portal for Medical Research (Deutsches Forschungsdatenportal für Gesundheit, FDPG) (18,51). Even
320 though admission to the FDPG is currently only possible for sites that participate in the MII (mainly
321 universities), our work is also highly relevant for non-university service providers, as feasibility studies
322 can also be carried out using the OHDSI tools. We also consider it likely that our work may be of use
323 to the European Health Data Space (EHDS) (52) in the future, as our platform can be extended with
324 appropriate services for data harmonization.

325 Although our work is a big step towards intersectoral provision of RWD, it has some limitations due
326 to the following factors:

327 1) Healthcare providers use site-specific clinical information systems to store RWD, which have
328 limited standardized interfaces for the data exchange. Therefore, the RWD from possibly multiple data
329 sources must be harmonized into the data format of our provided Research Data Repository in a pre-
330 processing step by the sites themselves, which can be a challenge depending on the type and scope of
331 data storage, the available human and economic resources and the knowledge of medical informatics.
332 The legally binding interface specifications currently developed and established in Germany, such as
333 the HL7 FHIR standard-based *Information technology systems in hospitals* (Informationstechnische
334 Systeme im Krankenhäusern, ISIK) (53) and *Medical Information Objects* (Medizinische
335 Informationsobjekte, MIO) (54), could provide a remedy here, as services based on these standards
336 could be developed and made available via our platform to automatically convert the RWD into the
337 FHIR MII CDS format for research purposes.

338 2) The FHIR MII CDS v1.0, established as the standard for data exchange within the MII, presents
339 certain limitations, particularly in its dataset specifications for specific medical fields like oncology

340 and ophthalmology. Although plans are in place to refine profiles for oncology in the upcoming FHIR
341 MII CDS v2, there remains a need to develop more appropriate dataset specifications for specific
342 medical domains. These improvements are crucial for effectively incorporating such specifications into
343 similar service platforms, especially in key areas like *Observations, Imaging Studies, Diagnostic*
344 *Reports, Procedures, and Medication Administrations*.

345 3) In addition, the services/applications cannot yet be installed "out of the box" via the frontend. In
346 order to further minimize the technical hurdle, this is an important goal for the future.

347 Despite the limitations, our Pre-built Packages together with the Service Platform Prototype can
348 already be used to provide data for specific research projects in a time-saving manner. We believe that
349 our research represents a significant contribution in research data management, offering an efficient,
350 user-friendly, and reproducible way to establish project-specific data provision pipelines. This may
351 make our work interesting not only for non-university service providers, but also for university sites.
352 In addition, our Service Platform Prototype can serve as a foundation for third-party applications, e.g.
353 based on SMART-on-FHIR, which can be used not only for research but also for patient care.

354 Next, we will implement the Service System Platform at two non-university hospitals in Germany as
355 part of a pilot study. In order to achieve a high level of acceptance among end users, we will test its
356 functionality and usability. As part of the roll-out, an accompanying acceptance analysis in the form
357 of an observation protocol and in-depth face-to-face interviews will be conducted to ensure the solution
358 is usable and sustainable. For this purpose, we collaborate closely with experts from the fields of
359 technology acceptance and usability. Thanks to our heterogeneous teams, the interdisciplinary
360 perspective can have a supportive effect in order to strengthen user-friendliness and thereby actual
361 usage of the Service System Platform.

362 **5 Conclusions**

363 In conclusion, the developed Service Platform Prototype together with the Pre-built Packages represent
364 an essential step forward in managing and facilitating medical research studies, with an emphasis on
365 data harmonization, resource efficiency, data security, and collaborative effectiveness.

366 **6 Conflict of Interest**

367 The authors declare that the research was conducted in the absence of any commercial or financial
368 relationships that could be construed as a potential conflict of interest.

369 **7 Author Contributions**

370 KH: Conceptualization, Investigation, Methodology, Software, Visualization, Project administration,
371 Writing – original draft, Writing – review & editing; IN: Conceptualization, Methodology, Software,
372 Project administration, Writing – review & editing; YP: Resources, Software, Writing – review &
373 editing; EH: Resources, Software, Writing – review & editing; DB: Resources, Software, Writing –
374 review & editing; CK: Software, Writing – review & editing; MG: Resources, Writing – review &
375 editing; MB: Validation, Writing – review & editing; FN: Writing – review & editing; RG: Writing –
376 review & editing; SG: Resources, Writing – review & editing; AS: Resources, Writing – review &
377 editing; FB: Conceptualization, Funding acquisition, Resources, Writing – review & editing; IR:
378 Resources, Writing – review & editing; MW: Supervision, Writing – review & editing; JW: Writing –
379 review & editing; MS: Funding acquisition, Supervision, Writing – review & editing.

380 **8 Funding**

381 This research was funded by the German Federal Ministry of Education and Research
382 (<https://www.bmbf.de/en>) within the project “Medical Informatics Hub in Saxony (MiHUBx)”. Grant
383 numbers 01ZZ2101A (Dresden) and 01ZZ2101E (Chemnitz). The funders had no role in study design,
384 data collection and analysis, decision to publish, or preparation of the manuscript.

385 **9 Acknowledgments**

386 Not applicable.

387 **10 Reference**

- 388 1. Bathelt F, Reinecke I, Peng Y, Henke E, Weidner J, Bartos M, et al. Opportunities of Digital
389 Infrastructures for Disease Management—Exemplified on COVID-19-Related Change in
390 Diagnosis Counts for Diabetes-Related Eye Diseases. *Nutrients*. 2022 May 11;14(10):2016.
- 391 2. Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges
392 and opportunities. *BMC Medical Research Methodology*. 2022 Nov 5;22(1):287.
- 393 3. Orsini LS, Berger M, Crown W, Daniel G, Eichler HG, Goettsch W, et al. Improving
394 Transparency to Build Trust in Real-World Secondary Data Studies for Hypothesis Testing—
395 Why, What, and How: Recommendations and a Road Map from the Real-World Evidence
396 Transparency Initiative. *Value in Health*. 2020 Sep;23(9):1128–36.

- 397 4. Kalra D. The importance of real-world data to precision medicine. *Personalized Medicine*.
398 2019 Mar;16(2):79–82.
- 399 5. Wolfien M, Ahmadi N, Fitzer K, Grummt S, Heine KL, Jung IC, et al. Ten Topics to Get
400 Started in Medical Informatics Research. *J Med Internet Res*. 2023 Jul 24;25:e45948.
- 401 6. Gehring S, Eulenfeld R. German Medical Informatics Initiative: Unlocking Data for Research
402 and Health Care. *Methods Inf Med*. 2018 Jul;57(S 01):e46–9.
- 403 7. Semler S, Wissing F, Heyder R. German Medical Informatics Initiative: A National Approach
404 to Integrating Health Data from Patient Care and Medical Research. *Methods Inf Med*. 2018
405 Jul;57(S 01):e50–6.
- 406 8. Medical Informatics Initiative. The Medical Informatics Initiative’s core data set [Internet].
407 [cited 2023 Dec 6]. Available from: [https://www.medizininformatikinitiative.de/en/medical-](https://www.medizininformatikinitiative.de/en/medical-informatics-initiatives-core-data-set)
408 [informatics-initiatives-core-data-set](https://www.medizininformatikinitiative.de/en/medical-informatics-initiatives-core-data-set)
- 409 9. HL7.org, Health Level Seven International Fast Healthcare Interoperability Resources (HL7
410 FHIR) [Internet]. [cited 2023 Dec 15]. Available from: <http://hl7.org/fhir/>
- 411 10. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-
412 based, interoperable apps platform for electronic health records. *Journal of the American*
413 *Medical Informatics Association*. 2016 Sep 1;23(5):899–908.
- 414 11. HL7 International [Internet]. [cited 2024 Jan 13]. Available from: <https://www.hl7.org>
- 415 12. OHDSI.org [Internet]. [cited 2024 Jan 13]. Available from: <https://www.ohdsi.org>
- 416 13. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health
417 Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud*
418 *Health Technol Inform*. 2015;216:574–8.
- 419 14. OHDSI.org, HL7 International and OHDSI Announce Collaboration to Provide Single
420 Common Data Model for Sharing Information in Clinical Care and Observational Research
421 [Internet]. [cited 2024 Jan 13]. Available from: <https://www.ohdsi.org/ohdsi-hl7-collaboration/>
- 422 15. OMOP Common Data Model – OHDSI [Internet]. [cited 2023 Dec 15]. Available from:
423 <https://www.ohdsi.org/data-standardization/>
- 424 16. OHDSI.org, Software tools [Internet]. [cited 2024 Jan 13]. Available from:
425 <https://www.ohdsi.org/software-tools/>
- 426 17. Prokosch HU, Bahls T, Bialke M, Eils J, Fegeler C, Gruendner J, et al. The COVID-19 Data
427 Exchange Platform of the German University Medicine. In: Séroussi B, Weber P, Dhombres F,
428 Grouin C, Liebe JD, Pelayo S, et al., editors. *Studies in Health Technology and Informatics*

- 429 [Internet]. IOS Press; 2022 [cited 2024 Jan 15]. Available from:
430 <https://ebooks.iospress.nl/doi/10.3233/SHTI220554>
- 431 18. Prokosch HU, Gebhardt M, Gruendner J, Kleinert P, Buckow K, Rosenau L, et al. Towards a
432 National Portal for Medical Research Data (FDPG): Vision, Status, and Lessons Learned. In:
433 Hägglund M, Blusi M, Bonacina S, Nilsson L, Cort Madsen I, Pelayo S, et al., editors. *Studies*
434 *in Health Technology and Informatics* [Internet]. IOS Press; 2023 [cited 2024 Jan 7]. Available
435 from: <https://ebooks.iospress.nl/doi/10.3233/SHTI230124>
- 436 19. MiHUBx: ein digitales Ökosystem für Forschung, Diagnostik und Therapie [Internet].
437 Deutsche Zentrum für Luft und Raumfahrt e.V. - DLR Gesundheitsforschung; [cited 2023 Dec
438 15]. Available from: [https://www.gesundheitsforschung-bmbf.de/de/mihubx-ein-digitales-](https://www.gesundheitsforschung-bmbf.de/de/mihubx-ein-digitales-okosystem-fur-forschung-diagnostik-und-therapie-13054.php)
439 [okosystem-fur-forschung-diagnostik-und-therapie-13054.php](https://www.gesundheitsforschung-bmbf.de/de/mihubx-ein-digitales-okosystem-fur-forschung-diagnostik-und-therapie-13054.php)
- 440 20. Basismodule des Kerndatensatzes der MII | Medizininformatik-Initiative [Internet]. [cited 2023
441 Dec 15]. Available from: [https://www.medizininformatik-initiative.de/de/basismodule-des-](https://www.medizininformatik-initiative.de/de/basismodule-des-kerndatensatzes-der-mii)
442 [kerndatensatzes-der-mii](https://www.medizininformatik-initiative.de/de/basismodule-des-kerndatensatzes-der-mii)
- 443 21. Prokosch HU, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM:
444 Medical Informatics in Research and Care in University Medicine: A Large Data Sharing
445 Network to Enhance Translational Research and Medical Care. *Methods Inf Med*. 2018
446 Jul;57(S 01):e82–91.
- 447 22. Miracum.org, MIRACOLIX Tools [Internet]. [cited 2024 Jan 15]. Available from:
448 <https://www.miracum.org/das-konsortium/datenintegrationszentren/miracolix-tools/>
- 449 23. Kiel A. Blaze FHIR® Server with internal, fast CQL Engine [Internet]. [cited 2023 Dec 15].
450 Available from: <https://sampler.github.io/blaze/>
- 451 24. Schüttler C, Prokosch HU, Hummel M, Lablans M, Kroll B, Engels C, et al. The journey to
452 establishing an IT-infrastructure within the German Biobank Alliance. Ahmed Z, editor. *PLoS*
453 *ONE*. 2021 Sep 22;16(9):e0257632.
- 454 25. Gruendner J, Deppenwiese N, Folz M, Köhler T, Kroll B, Prokosch HU, et al. The Architecture
455 of a Feasibility Query Portal for Distributed COVID-19 Fast Healthcare Interoperability
456 Resources (FHIR) Patient Data Repositories: Design and Implementation Study. *JMIR Med*
457 *Inform*. 2022 May 25;10(5):e36709.
- 458 26. GitHub. FHIR gateway. [Internet]. [cited 2024 Jan 15]. Available from:
459 <https://github.com/miracum/fhir-gateway>
- 460 27. Basic HTTP Authentication Scheme RFC 7617 [Internet]. [cited 2023 Jan 5]. Available from:
461 <https://datatracker.ietf.org/doc/html/rfc7617>

- 462 28. Gruhl M, Reinecke I, Sedlmayr M. Specification and Distribution of Vocabularies Among
463 Consortial Partners. *Stud Health Technol Inform*. 2020 Jun 16;270:1393–4.
- 464 29. Peng Y, Henke E, Reinecke I, Zoch M, Sedlmayr M, Bathelt F. An ETL-process design for
465 data harmonization to participate in international research with German real-world data based
466 on FHIR and OMOP CDM. *International Journal of Medical Informatics*. 2023 Jan
467 1;169:104925.
- 468 30. Henke E, Peng Y, Reinecke I, Zoch M, Sedlmayr M, Bathelt F. An ETL-process design for
469 incremental loading German real-world data based on FHIR and OMOP CDM: Algorithm
470 Development and Validation (Preprint). *JMIR Medical Informatics* [Internet]. 2023 Mar 15
471 [cited 2023 Aug 25]; Available from: <http://preprints.jmir.org/preprint/47310/accepted>
- 472 31. Kümmel M, Reinecke I, Gruhl M, Bathelt F, Sedlmayr M. Transition Database for a
473 harmonized mapping of German patient data to the OMOP CDM - The ‘German TDB’
474 [Internet]. 2020 [cited 2023 Nov 15]. Available from: [https://www.ohdsi.org/wp-](https://www.ohdsi.org/wp-content/uploads/2020/05/13_Poster.pdf)
475 [content/uploads/2020/05/13_Poster.pdf](https://www.ohdsi.org/wp-content/uploads/2020/05/13_Poster.pdf)
- 476 32. PostgreSQL: The World’s Most Advanced Open Source Relational Database [Internet]. [cited
477 2023 Dec 15]. Available from: <https://www.postgresql.org>
- 478 33. HL7.org, FHIR Core R4, Administrative Gender [Internet]. [cited 2024 Jan 3]. Available from:
479 <https://hl7.org/fhir/r4/valueset-administrative-gender.html>
- 480 34. Keycloak: Open Source Identity and Access Management [Internet]. Available from:
481 <https://www.keycloak.org>
- 482 35. OpenID. OpenID. San Ramon, CA: OpenID Foundation [Internet]. 2023 [cited 2023 Dec 18].
483 Available from: <https://openid.net/connect/>
- 484 36. OAuth 2.0 RFC 6749, Section 4.4: Client Credentials Flow [Internet]. Available from:
485 <https://datatracker.ietf.org/doc/html/rfc6749#section-4.4>
- 486 37. docker: Make better, secure software from the start [Internet]. [cited 2023 Dec 16]. Available
487 from: <https://www.docker.com>
- 488 38. docker.com, Docker compose [Internet]. [cited 2014 Jan 19]. Available from:
489 <https://docs.docker.com/compose/>
- 490 39. pgAdmin: Open Source administration and development platform for PostgreSQL [Internet].
491 [cited 2023 Dec 29]. Available from: <https://www.pgadmin.org>
- 492 40. Portainer: Container Management Software for Kubernetes and Docker [Internet]. [cited 2023
493 Dec 29]. Available from: <https://www.portainer.io>

- 494 41. Geis T, Tesch G. Basiswissen Usability und User Experience: Aus- und Weiterbildung zum
495 UXQB Certified Professional for Usability and User Experience (CPUX) - Foundation Level
496 (CPUX-F). 2., überarbeitete und aktualisierte Auflage. Heidelberg: dpunkt.verlag; 2023. 312 p.
- 497 42. Quarkus: A Kubernetes Native Java stack tailored for OpenJDK HotSpot and GraalVM, crafted
498 from the best of breed Java libraries and standards. [Internet]. [cited 2023 Dec 16]. Available
499 from: <https://quarkus.io>
- 500 43. Angular: The web development framework for building the future [Internet]. [cited 2023 Dec
501 16]. Available from: <https://angular.io>
- 502 44. Bootstrap widgets: The angular way [Internet]. [cited 2024 Jan 9]. Available from: [https://ng-
503 bootstrap.github.io/](https://ng-bootstrap.github.io/)
- 504 45. gitHub, Java Docker API Client [Internet]. [cited 2024 Jan 8]. Available from:
505 <https://github.com/docker-java/docker-java>
- 506 46. RabbitMQ: Message Broker [Internet]. [cited 2023 Dec 16]. Available from:
507 <https://www.rabbitmq.com>
- 508 47. ResearchData-TO-FHIR Package [Internet]. Available from:
509 <https://gitlab.ukdd.de/pub/mihubx/researchdata-to-fhir-install-pkg>
- 510 48. FHIR-TO-OMOP package [Internet]. Available from: [https://gitlab.ukdd.de/mihubx/fhir-to-
511 omop-install-pkg](https://gitlab.ukdd.de/mihubx/fhir-to-omop-install-pkg)
- 512 49. Addons package [Internet]. Available from: <https://gitlab.ukdd.de/mihubx/addons-install-pkg>
- 513 50. MiHUBx Service Platform Prototype [Internet]. [cited 2024 Jan 23]. Available from:
514 <https://gitlab.ukdd.de/pub/mihubx/mihubx-service-portal>
- 515 51. Das Deutsche Forschungsportal für Gesundheit (FDPG) [Internet]. [cited 2024 Jan 7].
516 Available from: <https://forschen-fuer-gesundheit.de>
- 517 52. Horgan D, Hajduch M, Vrana M, Soderberg J, Hughes N, Omar MI, et al. European Health
518 Data Space—An Opportunity Now to Grasp the Future of Data-Driven Healthcare. Healthcare.
519 2022 Aug 26;10(9):1629.
- 520 53. gematik, ISiK – Informationstechnische Systeme in Krankenhäusern. [Internet]. [cited 2024 Jan
521 7]. Available from: <https://fachportal.gematik.de/informationen-fuer/isik>
- 522 54. KBV, MIO - Medizinische Informationsobjekte. [Internet]. Available from:
523 <https://mio.kbv.de/site/mio>

524

525 **11 Supplementary Material**

526 S1 File: Example table for the semantic mapping of vital parameter:
527 observation_vital_parameter_loinc_code.csv

528 (CSV)

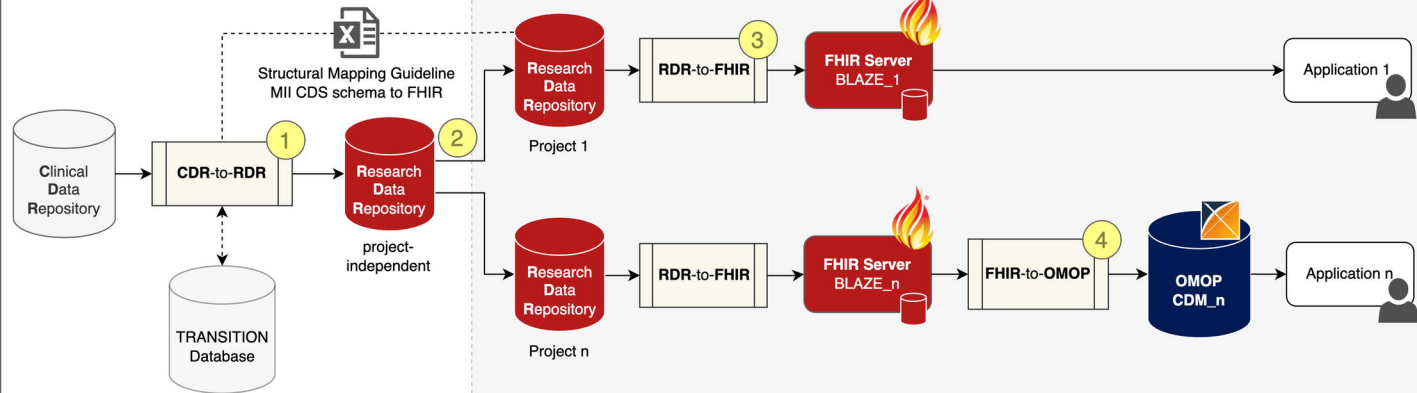
529 S2 File: Structural Mapping Guideline MII CDS schema to FHIR

530 (Excel)

531 **12 Data Availability Statement**

532 The source code of the pre-built packages “ResearchData-to-FHIR v2.2.0”, “FHIR-to-OMOP v1.1.0”,
533 “Addons v1.0.0” and “MiHUBx Service Platform v1.0.0” can be downloaded from the GitLab
534 repository at <https://gitlab.ukdd.de/pub/mihubx>. These repositories include the test datasets, developer
535 documentation and the initial installation instructions.

Project-related provision of anonymized or pseudonymized research data

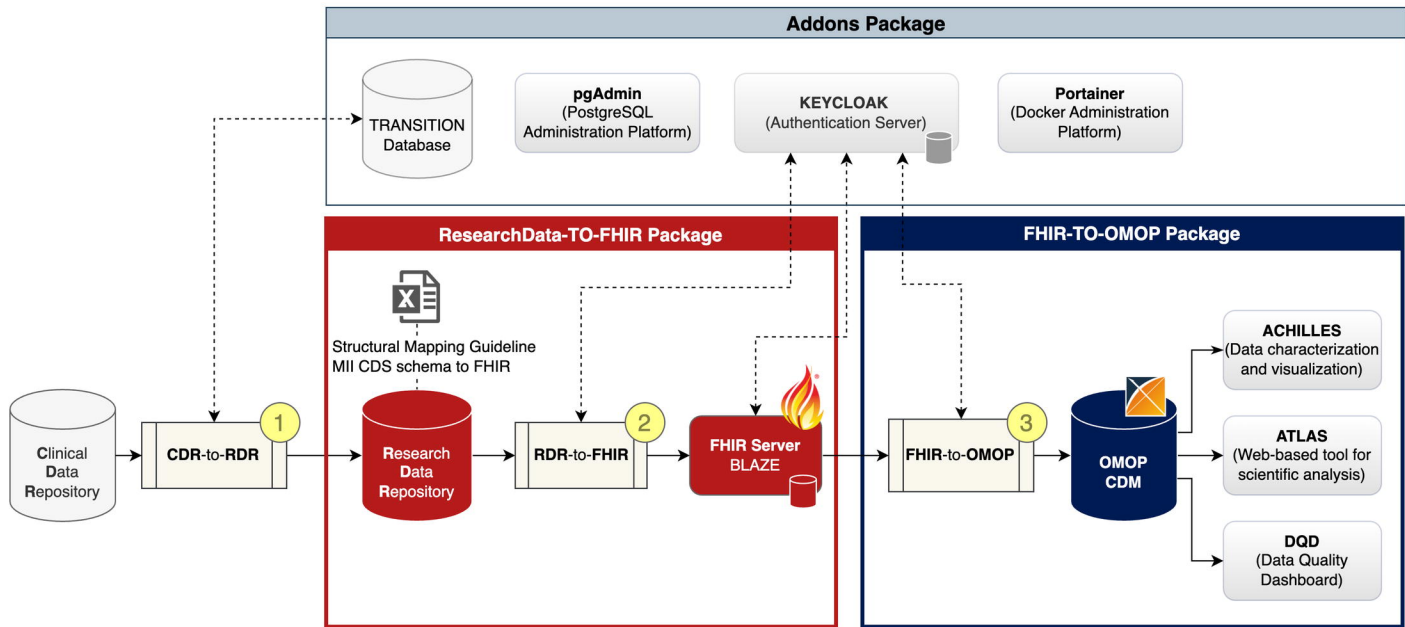


FHIR® logo used under principle of fair use, without explicit permission for the purposes of teaching, scholarship and research. Attribution: Health Level Seven International (2011)

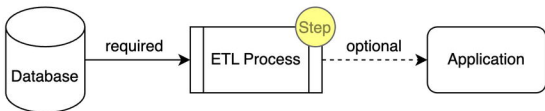
OMOP logo used under principle of fair use, without explicit permission for the purposes of teaching, scholarship and research. Attribution: OHDSI Community

Site-specific

Service Platform



Legend:



FHIR® logo used under principle of fair use, without explicit permission for the purposes of teaching, scholarship and research. Attribution: Health Level Seven International (2011)
 OMOP logo used under principle of fair use, without explicit permission for the purposes of teaching, scholarship and research. Attribution: OHDSI Community

Projects

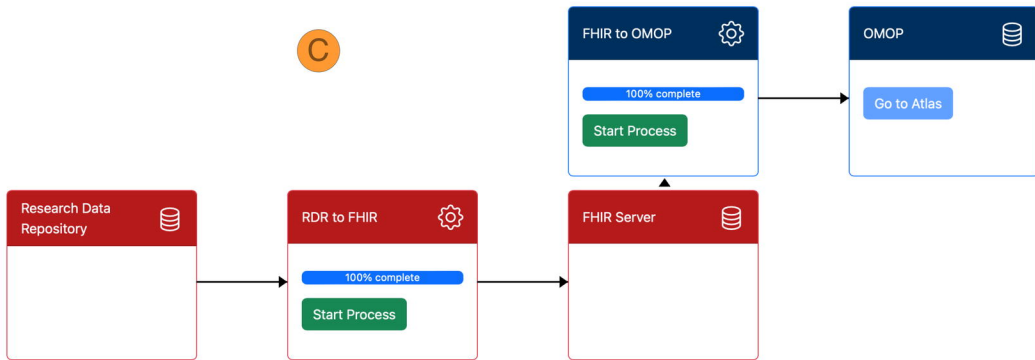
FPDG

Project 1

Project 2

Create New Project

Services Overview for "Project 1"



Services List for "Project 1"

Filter:

ID	Container Name	Ports	Status
1	Project 1-rdr-1	5432:5432 (tcp)	Running
3	Project 1-blaze-1	8082:8080 (tcp)	Running
3	Project 1-blaze-1	8082:8080 (tcp)	Running
4	Project 1-fhirtoomop-1	8083:8080 (tcp)	Running
5	Project 1-omop-db-1	5444:5432 (tcp)	Running
6	Project 1-atlas-1	8080:8080 (tcp)	Running
7	Project 1-ohdsi/webapi-1	9876:8080 (tcp)	Running