



# Simplifying Multimodal Clinical Research Data Management: Introducing an Integrated and User-friendly Database Concept

Anna Schweinar<sup>1,2</sup> Franziska Wagner<sup>1,3</sup> Carsten Klingner<sup>1,3</sup> Sven Festag<sup>4</sup> Cord Spreckelsen<sup>4</sup>  
Stefan Brodoehl<sup>1,3</sup>

<sup>1</sup>Biomagnetic Center, University Hospital Jena, Friedrich Schiller University, Jena, Germany

<sup>2</sup>Else Kröner Graduate School for Medical Students "JSAM," Jena University Hospital, Jena, Germany

<sup>3</sup>Department of Neurology, Jena University Hospital, Jena, Germany

<sup>4</sup>Institute of Medical Statistics, Computer and Data Sciences, Jena University Hospital, Jena, Thüringen, Germany

**Address for correspondence** Stefan Brodoehl, MD, Department of Neurology, University Hospital Jena, Am Klinikum 1, 07747 Jena, Germany (e-mail: stefan.brodoehl@med.uni-jena.de).

Appl Clin Inform 2024;15:234–249.

## Abstract

**Background** Clinical research, particularly in scientific data, grapples with the efficient management of multimodal and longitudinal clinical data. Especially in neuroscience, the volume of heterogeneous longitudinal data challenges researchers. While current research data management systems offer rich functionality, they suffer from architectural complexity that makes them difficult to install and maintain and require extensive user training.

**Objectives** The focus is the development and presentation of a data management approach specifically tailored for clinical researchers involved in active patient care, especially in the neuroscientific environment of German university hospitals. Our design considers the implementation of FAIR (Findable, Accessible, Interoperable, and Reusable) principles and the secure handling of sensitive data in compliance with the General Data Protection Regulation.

**Methods** We introduce a streamlined database concept, featuring an intuitive graphical interface built on Hypertext Markup Language revision 5 (HTML5)/Cascading Style Sheets (CSS) technology. The system can be effortlessly deployed within local networks, that is, in Microsoft Windows 10 environments. Our design incorporates FAIR principles for effective data management. Moreover, we have streamlined data interchange through established standards like HL7 Clinical Document Architecture (CDA). To ensure data integrity, we have integrated real-time validation mechanisms that cover data type, plausibility, and Clinical Quality Language logic during data import and entry.

**Results** We have developed and evaluated our concept with clinicians using a sample dataset of subjects who visited our memory clinic over a 3-year period and collected several multimodal clinical parameters. A notable advantage is the unified data matrix, which simplifies data aggregation, anonymization, and export. **This streamlines data**

## Keywords

- ▶ clinical data management
- ▶ research data management
- ▶ neurology
- ▶ database
- ▶ SQLite

received

July 14, 2023

accepted after revision

November 22, 2023

accepted manuscript online

February 1, 2024

DOI <https://doi.org/10.1055/a-2259-0008>.

ISSN 1869-0327.

© 2024. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

exchange and enhances database integration with platforms like Konstanz Information Miner (KNIME).

**Conclusion** Our approach offers a significant advancement for capturing and managing clinical research data, specifically tailored for small-scale initiatives operating within limited information technology (IT) infrastructures. It is designed for immediate, hassle-free deployment by clinicians and researchers.

The database template and precompiled versions of the user interface are available at: [https://github.com/stebro01/research\\_database\\_sqlite\\_i2b2.git](https://github.com/stebro01/research_database_sqlite_i2b2.git).

## Background and Significance

The volume of routinely collected data generated in clinical practice and trials presents multiple challenges for data management in terms of storage, processing, sharing, and protection.<sup>1–4</sup> Thus, efficient research data management (RDM) is essential in modern medicine and research,<sup>5</sup> and provides researchers with access to routinely collected data. It provides the basis for scientific research, promotes scientific data publication, and increases reproducibility.<sup>1,5,6</sup> In general, RDM describes the organization, storage, preservation, and sharing of scientific data.

Especially in neuroscience, RDM is indispensable. The complexity and volume of heterogeneous (multimodal) neuroscience data (e.g., laboratory, microscopy, imaging, clinical examination results and scores, electrophysiological data, etc.) require good documentation, processing, and standardization. Many neurological diseases, especially neurodegenerative diseases, need follow-up observations. In the most common neurodegenerative disease, dementia, the diagnosis is made from multimodal data including neuropsychological testing, imaging, functional imaging, cerebrospinal fluid levels, etc.<sup>7</sup> Patients are evaluated annually, and documenting the large amount of longitudinal and multimodal data generated is challenging. With optimal RDM, the collected data are prepared for further relevant analysis methods and can be efficiently used for retrospective studies.<sup>2</sup>

Unfortunately, there is a large gap in RDM in practice, especially in the neurosciences. A recent online survey by the National Research Data Infrastructure (NFDI-Neuro) on the state of RDM clearly shows the existing problems.<sup>1</sup> Routinely collected clinical and scientific data are usually incomplete or not retrievable.<sup>2</sup> Knowledge of and adherence to data and metadata standards are often limited.<sup>6</sup> Many researchers and clinicians lack the time to invest in standards-compliant data processing and management. In addition, there is a lack of secure data management in terms of privacy and secure data sharing.<sup>1</sup>

### How can the Gaps in Research Data Management be Closed to Ensure Optimal Data Storage, Processing, and Standardization?

Recognizing the urgent need for efficient and user-friendly data management systems tailored for clinicians working with multimodal and longitudinal data, our initial research aimed to assess the suitability of existing platforms. However, a gap immediately became apparent: the needs of clinical researchers were not being adequately met by

current solutions. For example, while platforms such as Longitudinal Online Research and Imaging System (LORIS) offer rich feature sets, their complexity and steep learning curve make them unsuitable for smaller projects or for clinicians with limited technical skills. This discrepancy is exacerbated by the unstructured and haphazard data management methods commonly used in clinical settings—often relying on rudimentary folder structures and Excel spreadsheets. Furthermore, existing platforms rarely address the specific challenges posed by GDPR regulations around sensitive data, or the ethical imperatives that require localized data storage and restricted access. Finally, the constraints of local IT infrastructures designed with data security in mind can make server-based solutions impractical for routine operations. These multiple, intertwined challenges served as the primary catalyst for our project, highlighting the urgent need for a system that reconciles usability with the complex ethical and regulatory landscape of clinical RDM.<sup>8</sup>

## Objectives

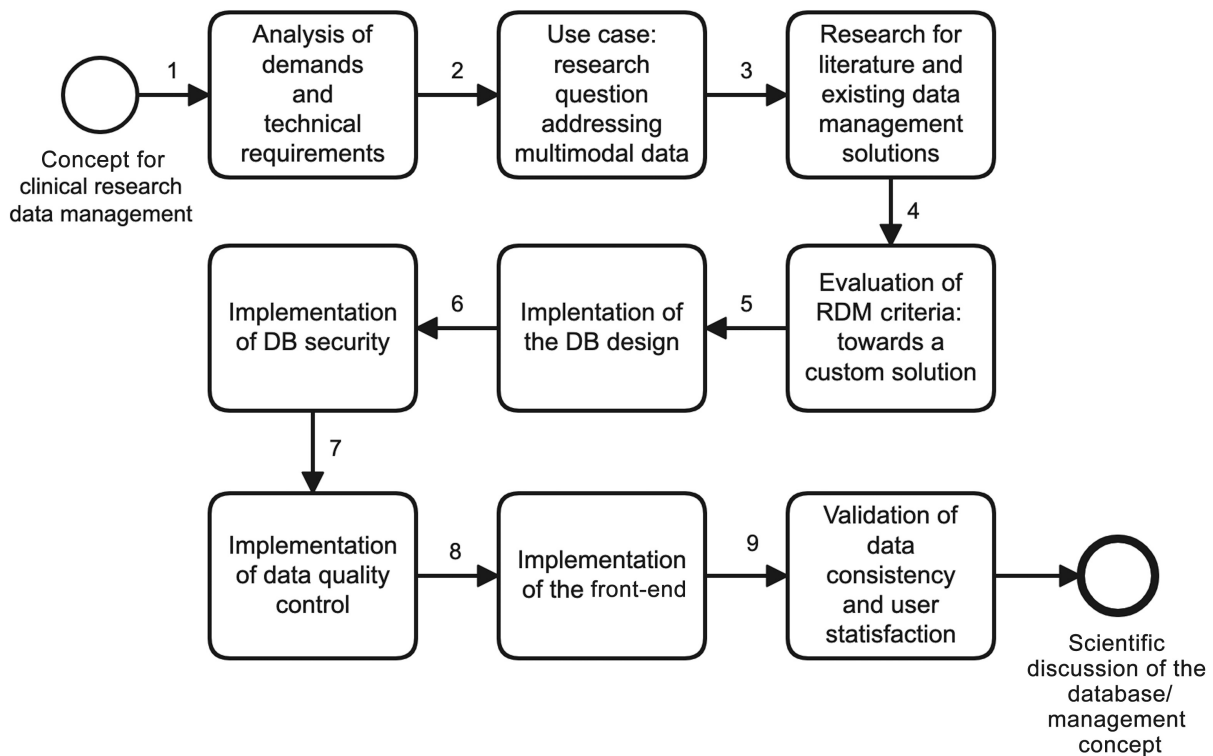
Initially, we researched the most appropriate RDM solution for clinical researchers, particularly in neuroscience, where the need to capture a variety of data types and manage their longitudinal nature is paramount. **However, we pinpointed a gap in existing solutions: they frequently fall short in catering to small-scale projects and offline capabilities, while also failing to balance ease of implementation with adherence to Findable, Accessible, Interoperable, and Reusable (FAIR) principles<sup>5</sup> and General Data Protection Regulation (GDPR) regulations.<sup>9</sup>**

This prompted us to develop a specialized solution aimed at

- storing diverse, longitudinal clinical data along with metadata;
- prioritizing user-friendly design and intuitive data capture;
- ensuring offline functionality and seamless IT integration;
- focusing on small-scale projects while maintaining data interchangeability with other applications.

## Methods

In the following section, we describe the steps for concept development, as shown in **–Fig. 1.**



**Fig. 1** Flow chart illustrating the steps from the assessment of user and technical requirements, the definition of a concrete use case, the literature research, the development, and implementation of a solution concept to the validation of the result. RDM, research data management.

### Step 1: Analysis of User Demands and Definition of Minimal Technical Requirements for Research Data Management

As a first step, we defined the requirements for the target group of clinical researchers (users) **working in patient care on a daily basis**, based on a set of minimum technical requirements as follows.

#### User Demands

To define the requirements for a clinical research data storage system from a user perspective, we conducted a standardized expert telephone interview with a total of 41 (open and closed) questions (duration: approximately 35–45 min). We conducted a total of 22 expert interviews, 20 of which were finalized and analyzed for our study. Of the final 20 participants, 11 were aged between 25 and 34 years, 8 were between 35 and 44 years, and 1 participant was over 45 years. In terms of their professions, 14 were physicians, 2 were psychologists, and 4 were researchers; the interview is available in **–Supplementary Table S1** (available in the online version only).

#### Minimal Technical Requirements

Following the German Research Foundation guidelines, we established minimum requirements for an optimal RDM system.<sup>10,11</sup>

In establishing specific criteria, we focused on elements such as FAIR principles, data security, validation, HL7 integration, and offline functionality. **To bolster data interoperability**

and sustainability, we mandate the incorporation of extensive metadata, persistent identifiers, and standardized data formats, such as HL7.<sup>12</sup> Additionally, we aligned with clinical classification systems like logical observation identifiers names and codes (LOINC) and systematized nomenclature of medicine and clinical terms (SNOMED-CT). For data security, we added role-based access control and potential pseudonymization techniques to the criteria. The specific criteria are listed in **–Table 1** under the category of “*Technical requirements*.”

### Step 2: Use Case: Clinical Research Question with Multimodal Neuroscientific Data

**Based on the defined user needs and technical requirements for RDM, we aimed to investigate how these criteria apply to a research question involving a large amount of multimodal data.** We decided to use the clinical course data (retrospective) of patients who visited our memory clinic at the Department of Neurology during the period 2014 to 2022 with a diagnosis of mild cognitive impairment (MCI). For illustration, please refer to **–Supplementary Fig. S1** (available in the online version only).

### Step 3: Literature Research and Search for Suitable Existing Research Data Management Solutions

In our study, we conducted a literature search following preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement,<sup>13</sup> focusing on “database solutions for RDM.” We included 25 tools, categorized into

**Table 1** User demands and technical requirements for the data management solution

Description	Details on how to implement it
<b>User demands</b>	
Clear storage structure and easy access	Intuitive interface, structured storage, and efficient search functionality
Defined standards for data storage and export	Adoption of community standards, configurable export options; “no more data munging”
Rights management and data security	Role-based access control, encryption, pseudonymization techniques
Backup strategies, lifecycle management, and data duplication avoidance	Scheduled backups, data retention policies, and duplicate prevention algorithms
Accessibility from clinical workplace and home office	Responsive design, secure remote access, and offline functionality
<b>Technical requirements</b>	
FAIR principles: Findable, Accessible, Interoperable, Reusable data principles	Comprehensive metadata, persistent identifiers, standardized data formats
Multiple data types and extensibility: Support for diverse data types and ease of adding new data types	Modular architecture, customizable data type handling, and seamless integration of new types
Data security: Protect sensitive data from unauthorized access, modification, or disclosure	Role-based access controls, encryption, pseudonymization techniques
Data validation: Ensure data integrity and accuracy by validating input data	Input validation techniques at database schema and user interface levels
Duplicate prevention: Maintain data consistency by preventing duplicate records	Duplicate prevention algorithms and database constraints
HL7 integration: Facilitate seamless data exchange with other health care systems	Support HL7 standards for importing and exporting data
Offline functionality: Allow users to access and manage research data from within a local network/file system	Incorporate offline functionality without the need of a server configuration etc.

To avoid confusion, some items have been left duplicated between the user demands and the basic technical requirements (e.g., offline functionality, rights management).

Electronic Data Capture (EDC), clinical trial data management systems (CTMS), specific data management systems, and Electronic Laboratory Notebooks (ELN). Categorization is challenging because there is no universally accepted nomenclature, leading to the use of different terminologies in the field. We also distinguished between well-known and lesser known EDC and CTMS. Our research included an extensive literature review to ensure that we covered prominent systems and presented a holistic view of these essential tools. →Table 2 lists the available tools with their technical requirements and a brief functional description.

#### Step 4: Applying Research Data Management Criteria and Evaluation: Towards a Custom Solution

We conducted an internal review of various existing solutions, to assess their suitability for our specific needs. While many of these platforms offer a broad range of features and applications, they frequently fell short in several key areas. These included data storage options, either cloud-based or requiring a local server, as well as ease of implementation and administration. Even seemingly straightforward solutions like DataLad<sup>14</sup> or research electronic data capture (REDCap)<sup>15</sup> come with their own challenges, such as the

need for administrative rights and complex setup processes. Additionally, cost considerations were a factor. We also noted that many of these tools lack support for clinical classification systems, which would enable the use of international standards for data capture, despite the feasibility of implementing such features. A detailed rationale for the selection of various tools under consideration is elaborated upon in the “Results” section.

This outcome prompted the development of our custom solution, tailored to meet our defined requirements. SQLite was chosen as a serverless and self-contained database management system. It is well-suited for small applications and research projects due to its lightweight and portable design. The entire database is stored in a single file, making it easy to manage, backup, and share. SQLite, compatible across platforms and adhering to the atomicity, consistency, isolation and durability (ACID) standard,<sup>16</sup> is a continuously evolving open-source project supported by the community.<sup>17</sup> Leveraging HTML5/CSS for front-end development, we utilized Electron to create a cross-platform desktop application, ensuring consistent user experience and simplifying development and maintenance due to the extensive use and documentation of these web technologies.

**Table 2** Overview of existing data management solutions based on literature research

Category	Name	Technical requirements	Functions
Renown EDC/CTMS	REDCap (Harris et al 2009 <sup>15</sup> ) (REMOVED PRINTDATE FIELD)	<ul style="list-style-type: none"> <li>Operating system: Windows, macOS, Linux, Unix</li> <li>Web server: Windows/IIS and Linux/Apache with PHP 7.2.5 and above</li> <li>MySQL installation</li> <li>SMTP email server: configuring PHP with an institutional SMTP server</li> </ul>	<ul style="list-style-type: none"> <li>Secure web application</li> <li>Clinical and translational research support</li> <li>Create and manage online surveys and databases</li> <li>Support for multiple field types</li> <li>Built-in reporting and data export tools</li> </ul>
	OpenClinica (Cavejaars et al 2015 <sup>20</sup> )	<ul style="list-style-type: none"> <li>Operating system: Linux, Unix, or Windows</li> <li>Application server: Apache Tomcat</li> <li>PostgreSQL installation</li> <li>Language: JDK 7</li> </ul>	<ul style="list-style-type: none"> <li>Open source software</li> <li>Data collection and management for research studies</li> <li>Built-in reporting and data export tools</li> <li>Integrates with visualization and analysis tools</li> <li>Allows you to extend their functionality</li> </ul>
	ClinCapture <sup>34</sup>	<ul style="list-style-type: none"> <li>Operating system: Windows, macOS, and Linux</li> <li>Internet connection</li> <li>Developed from OpenClinica (version 3.1.3) source code</li> <li>Modern web browsers</li> </ul>	<ul style="list-style-type: none"> <li>Open source and cloud-based EDC system</li> <li>Data collection and management for research studies</li> <li>Drag and drop form design</li> <li>Allows you to extend your functionality</li> <li>Risk-based monitoring</li> <li>Integrates with visualization and analytics tools</li> </ul>
	MACRO EDC <sup>35</sup>	<ul style="list-style-type: none"> <li>Cloud-based or on-site installation</li> </ul>	<ul style="list-style-type: none"> <li>Clinical data management software</li> <li>Data collection and management for research studies</li> <li>Drag and drop form design</li> <li>Built-in reporting and data export tools</li> <li>Integration with visualization and analysis tools</li> </ul>
	EZ-Entry (Gao et al 2008 <sup>36</sup> )	<ul style="list-style-type: none"> <li>Operating system: Microsoft Windows</li> <li>Microsoft SQL Server 7.0 installation</li> <li>Programming language: Microsoft Visual Basic 6.0</li> </ul>	<ul style="list-style-type: none"> <li>Clinical data management software platform</li> <li>Data entry and capture</li> <li>eCRF design and management</li> <li>Built-in reporting and data export tools</li> </ul>
	Oracle Clinical <sup>37</sup>	<p>Database:</p> <ul style="list-style-type: none"> <li>Oracle database</li> <li>Operating system: Oracle Linux 7.6, 7.7 and 8.1, Oracle Solaris SPARC 11.4, HP-UX Itanium 11.31, Microsoft Windows Server 2019</li> </ul> <p>Application:</p> <ul style="list-style-type: none"> <li>Operating system: Microsoft Windows Server 2016 and 2019 Standard</li> <li>Other Oracle-specific applications</li> </ul> <p>Client:</p> <ul style="list-style-type: none"> <li>Operating system: Microsoft Windows 10, iOS 15.0 (iPad), macOS Big Sur version 11.6</li> </ul>	<ul style="list-style-type: none"> <li>Clinical data management system</li> <li>Data collection and management for research studies</li> <li>eCRF generation</li> <li>Integration with visualization and analysis tools</li> </ul>
	TrialDB (Brandt et al <sup>38</sup> )	–	<ul style="list-style-type: none"> <li>Web-based database system for clinical trials</li> <li>Open source</li> <li>Support for web-based interfaces</li> <li>Supports multiple field types</li> <li>Clinical research form (eCRF) generation</li> </ul>

Table 2 (Continued)

Category	Name	Technical requirements	Functions
Less-known EDC and Less-known CTMS	LORIS (Longitudinal Online Research and Imaging System; Das et al 2012 <sup>38</sup> )	<ul style="list-style-type: none"> <li>Linux-/Ubuntu-based operating system, PHP support</li> <li>Web server running NGINX or Apache</li> <li>MySQL installation</li> <li>Network access and firewall configurations</li> </ul>	<ul style="list-style-type: none"> <li>Web-based platform for easy access from anywhere</li> <li>Open source</li> <li>Built-in data entry and validation tools</li> <li>Longitudinal data tracking and robust user management</li> <li>Integrates with popular neuroimaging processing tools</li> </ul>
	FIMED (Hurtado et al 2021 <sup>39</sup> )	<ul style="list-style-type: none"> <li>Operating system: Windows, Linux, and macOS</li> <li>Web server: Apache or Nginx</li> <li>Database: developed in MongoDB (JSON Code)</li> </ul>	<ul style="list-style-type: none"> <li>Open source clinical data collection software</li> <li>Collect and manage multidimensional clinical research data</li> <li>eCRF generation</li> <li>Integrated analysis and visualization tools</li> <li>Ability to extend functionality</li> </ul>
	DADOS-Prospective (Nguyen et al 2006 <sup>40</sup> )	<ul style="list-style-type: none"> <li>Operating systems: Linux and Windows</li> <li>Programming language: Java</li> <li>Other requirements: Tomcat 5.x</li> </ul>	<ul style="list-style-type: none"> <li>Open source web-based prospective data collection application</li> <li>Prospective data collection and management</li> <li>eCRF creation</li> <li>Secure patient information</li> </ul>
	PhOsCo (Venizeas et al <sup>41</sup> )	<ul style="list-style-type: none"> <li>Operating system: Platform-independent (Linux, Windows, AS400)</li> <li>Implementation: client-server operation</li> <li>Language: Java</li> </ul>	<ul style="list-style-type: none"> <li>Open source software</li> <li>Data collection and management for research studies</li> <li>Allows you to extend its functionality</li> <li>Hybrid communication with local database synchronization when online</li> </ul>
	Database management system (Lee et al 2015 <sup>42</sup> )	<ul style="list-style-type: none"> <li>MySQL and phpMyAdmin Installation</li> <li>Central machine server</li> <li>Client GUI: Microsoft Access</li> </ul>	<ul style="list-style-type: none"> <li>Database management system for interventional oncology clinical research</li> <li>Data collection, organization and reporting</li> <li>Built-in reporting and data export tools</li> <li>Costing tools integration</li> </ul>
	Phynx: an open-source software solution (Egbring et al 2010 <sup>43</sup> )	<ul style="list-style-type: none"> <li>Linux Virtual Amazon Machine Image (AMI; Setup: Amazon Elastic Compute Cloud)</li> <li>MySQL Installation</li> <li>Server running Apache Tomcat</li> </ul>	<ul style="list-style-type: none"> <li>Open source software to support data management and web-based patient-level data</li> <li>Creation and verification of electronic patient profiles from medical databases</li> <li>Verification and validation process</li> </ul>
Specialized data management tools	SEEK (Wolstencroft et al 2015 <sup>44</sup> )	<ul style="list-style-type: none"> <li>Operating systems: Linux, Mac OS X, VirtualBox-based virtual machine</li> <li>Other requirements: Ruby on Rails environment - version 3.2.x</li> <li>Programming language: Ruby 2.1.x, Java 6 or 7 (including OpenJDK)</li> </ul>	<ul style="list-style-type: none"> <li>Open source web platform</li> <li>Supports management, sharing and exploration of data and models in systems biology</li> <li>Integration of model simulations, representation of experimental data</li> <li>Supports data annotation and standardization</li> </ul>
	openBIS (Bauch et al 2011 <sup>45</sup> )	<ul style="list-style-type: none"> <li>Operating system: Unix-like operating system such as Linux</li> <li>Java Runtime Environment 1.6, PostgreSQL 9.0 or 9.1 installed</li> <li>License: Apache Software License 2.0</li> <li>Programming language: Java, Jython</li> </ul>	<ul style="list-style-type: none"> <li>Open source framework for managing and analyzing complex biological research data</li> <li>Integrates with public databases</li> <li>For collecting, integrating, sharing, and publishing data</li> <li>Integrates with visualization and analysis tools</li> </ul>

(Continued)

**Table 2** (Continued)

Category	Name	Technical requirements	Functions
	ISA (Sansone et al 2012 <sup>46</sup> )	<ul style="list-style-type: none"> <li>• Various tools</li> </ul>	<ul style="list-style-type: none"> <li>• Open source frameworks and tools</li> <li>• Manage life science, environmental, and biomedical experiments</li> <li>• Provides rich description of experiment metadata</li> <li>• Store and manage experiment metadata</li> <li>• Integrates with analysis tools</li> <li>• Sharing and publishing tools</li> </ul>
	BioBankWarden (Ferretti et al 2017 <sup>47</sup> )	<ul style="list-style-type: none"> <li>• Server Apache HTTP in Ubuntu Server version 14.04</li> <li>• PostgreSQL (version 9.0) installation</li> </ul>	<ul style="list-style-type: none"> <li>• Web-based system for translational cancer research</li> <li>• Manage, store, and integrate clinical, biomolecular, and biomaterial cancer data</li> <li>• Manages disease research groups and research projects</li> <li>• Integrates with analysis tools</li> </ul>
	WebBioBank (Rossi et al 2014 <sup>48</sup> )	<ul style="list-style-type: none"> <li>• Server:Microsoft SQL Server</li> </ul>	<ul style="list-style-type: none"> <li>• Web-based systems for research data collection and biosignal analysis (Parkinson's disease)</li> <li>• Used in multicenter studies to integrate clinical and physiological data</li> <li>• Integration of analysis tools</li> <li>• Data export tools</li> </ul>
	DataLad <sup>14</sup>	<ul style="list-style-type: none"> <li>• Operating Systems: Windows, Linux, and macOS</li> <li>• Python, git-annex, Git</li> </ul>	<ul style="list-style-type: none"> <li>• Open source and free data management system</li> <li>• Data tracking and structuring</li> <li>• Collaboration support</li> </ul>
ELN	LabArchives <sup>49</sup>	<ul style="list-style-type: none"> <li>• Cloud-based software</li> <li>• Operating system: iOS Mobile App version 3.0.2: iOS 11 or higher, Android Mobile App version 3.0.0b104: Android OS 5 or higher</li> <li>• Browser: Chrome, Firefox, Safari, Microsoft Edge</li> </ul>	<ul style="list-style-type: none"> <li>• Create and save electronic lab notes</li> <li>• Categorize and search research data</li> <li>• Free plan for small teams</li> <li>• Supports document versioning and revision control</li> <li>• Integrates with visualization tools</li> <li>• Integrates with laboratory instruments and devices</li> </ul>
	Labguru <sup>50</sup>	<ul style="list-style-type: none"> <li>• Cloud-based software</li> <li>• Operating System: macOS 10 and upper versions, and Windows 10</li> <li>• Browsers: Apple® Safari®, Mozilla® Firefox®, Google Chrome™, and Microsoft Edge®</li> </ul>	<ul style="list-style-type: none"> <li>• Web-based ELN for experiment data and observations</li> <li>• Inventory management of lab supplies and reagents</li> <li>• Experiment and protocol management</li> <li>• Collaboration tools for team work</li> <li>• Integration with scientific software and instruments</li> <li>• Compliance and security features</li> </ul>
	SciNote <sup>51</sup>	<ul style="list-style-type: none"> <li>• Cloud-based software</li> <li>• Operating System: Android 5.0 (API level 21) and later, iOS version 11 and later</li> <li>• Browser: Chrome, Firefox, Safari</li> <li>• Leverages Docker technology</li> <li>• Runs on PostgreSQL and MS SQL</li> </ul>	<ul style="list-style-type: none"> <li>• ELN for managing laboratory data and processes</li> <li>• Integration with scientific software and instrumentation</li> <li>• Collaboration tools for teamwork</li> <li>• FDA 21 CFR Part 11 compliance</li> <li>• Integration with visualization tools</li> </ul>
	Elabjournal <sup>52</sup>	<ul style="list-style-type: none"> <li>• Cloud-based software</li> <li>• Operating system: iOS und Android</li> <li>Local installation: <ul style="list-style-type: none"> <li>• Application server: Microsoft Windows Server 2019</li> <li>• Database server: Ubuntu Server 18.04 LTS</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• ELN for managing laboratory data and processes</li> <li>• Integrates with data export tools</li> <li>• Integrates with visualization tools</li> <li>• Integrates with other laboratory tools and software</li> </ul>

**Table 2** (Continued)

Category	Name	Technical requirements	Functions
		<ul style="list-style-type: none"> <li>• Microsoft Office Online Server (OOS): Microsoft Windows Server 2019</li> <li>• Reverse proxy/load balancer and firewall</li> </ul>	
	RSpace <sup>53</sup>	RSpace Components <ul style="list-style-type: none"> <li>• Java Web Application</li> <li>• Apache Tomcat or similar web application</li> <li>• MariaDB database server</li> <li>• Database initialization script</li> <li>• Web Server</li> <li>• Ubuntu 20.04 LTS or 22.04 LTS or Debian 12</li> <li>• Java SDK version 17</li> <li>• Apache Web Server v2.4.</li> <li>• Apache Tomcat 9.x</li> <li>• MariaDB 10.3.</li> <li>• Hardware</li> <li>• Linux-based server with at least 8 GB RAM</li> <li>• End user browser requirements</li> <li>• Safari, Firefox and Chrome</li> </ul>	<ul style="list-style-type: none"> <li>• ELN for researchers to create and manage lab notes and research data</li> <li>• Integration with other laboratory tools and software</li> <li>• Collaboration tools for working with teams</li> <li>• Built-in reporting and data export tools</li> </ul>
	eLabFTW <sup>54</sup>	<ul style="list-style-type: none"> <li>• Cloud-based available</li> <li>• Operating System: Windows, Mac OS X, Linux, BSD, Solaris, etc</li> <li>• Must be installed on a server</li> <li>• Containerization technology like Docker or Podman</li> <li>• MySQL database</li> </ul>	<ul style="list-style-type: none"> <li>• Open-source, web-based lab notebook for recording and managing experiment data</li> <li>• Built-in reporting and data export tools</li> </ul>

Abbreviations: CTMS, clinical trial data management systems; eCRF, electronic case report form; EDC, electronic data capture; ELN, electronic laboratory notebooks; GUI, graphical user interface.

### Step 5: Implementation of the Database Design

We have based our design on the i2b2 Common Data Model (CDM)<sup>18</sup> star schema, prevalent for its efficient querying and analysis of clinical research data. This model enhances our database by offering scalability to accommodate growing clinical research datasets. Despite relying on join operations, its normalized structure ensures data consistency, efficient storage, and easier updates, thereby preventing data redundancy. Crucially, its global, unified representation accelerates data access and expands analytical capabilities across multiple subject areas by eliminating the need for separate tables.

### Step 6: Implementation of Data Security

**GDPR compliance was a key focus during development, necessitating data security at two levels.** Access to the SQLite database (DB) is controlled by assigned read/write permissions within the local or network drive. Second, the user interface (UI) utilizes a JavaScript class to dynamically generate user-specific views and handle “Create, Read, Update, Delete” operations, thereby ensuring secure management of database entries. **To comply with GDPR regulations, all data and data operations are tagged with a specific date, ensuring that research data are only retained for as long as is necessary to fulfill its intended purpose. At the**

**end of this period, or at the data subject’s request, the data can be securely deleted, in-line with the GDPR’s “right to be forgotten” provision.**

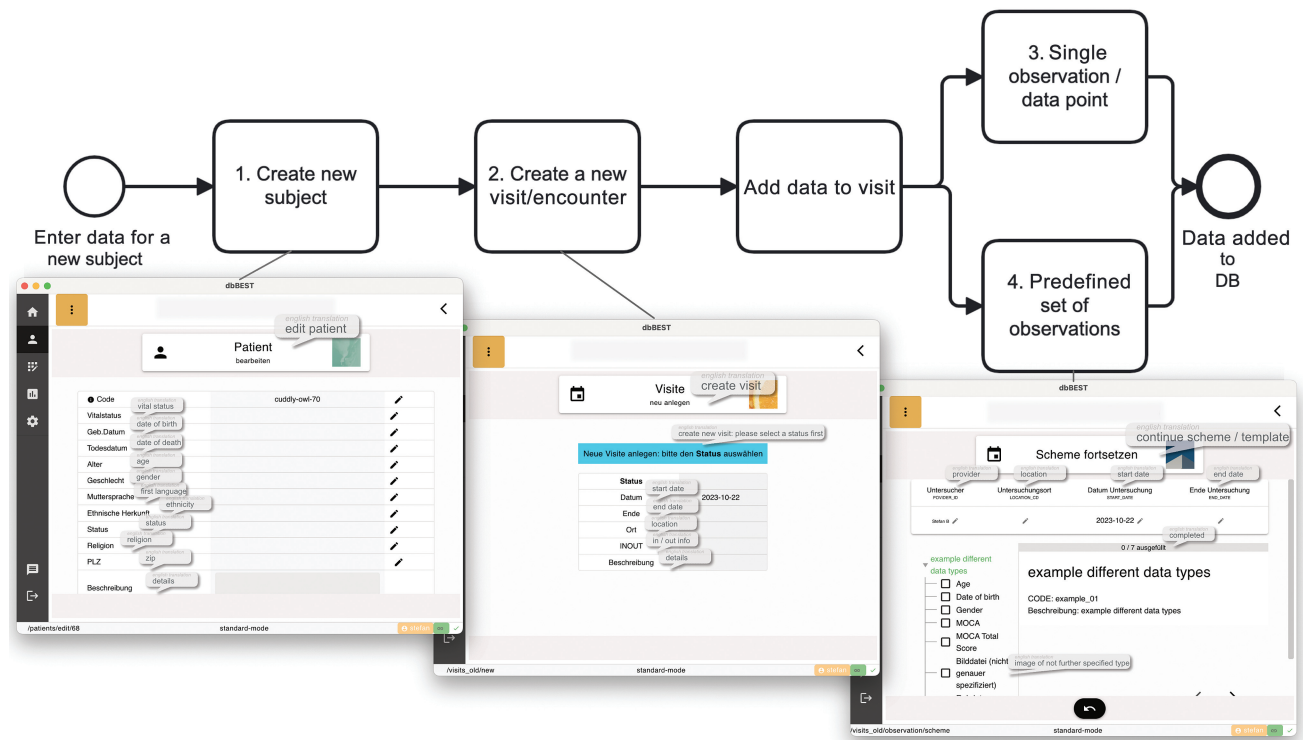
### Step 7: Implementation of Data Quality Control

To reduce erroneous data entry and improve data quality, we implemented type validation and duplicate detection methods directly in the input and import functions within the UI. In addition, we implemented logic-based rules based on Clinical Quality Language (CQL)<sup>19</sup> that provide direct user feedback on any rule violations during data entry and import. Integrating CQL using the CQL Framework (GitHub repository: <https://github.com/cqframework>) into our research database involves a two-step process. Firstly, CQL statements are transformed into JSON ELM (JSON expression logical model) representations. Secondly, the converted rule is interpreted and executed through JavaScript code. **→ Supplementary Fig. S2** (available in the online version only) **provides a graphical illustration of the process.**

### Step 8: Designing the Standard Views for the Graphical User Interface (Front-end)

During development, we delineated a data pathway comprising (1) new subject entry, (2) new visit creation, (3) individual





**Fig. 2** Illustration of the main views for entering clinical research data into the database using the UI. UI, user interface.

visit observations, or (4) fixed observations within a visit, resembling a Clinical Research Form (CRF). **Fig. 2** provides an illustration of this data path alongside the final UI.

### Step 9: Validation of Data Consistency and User Satisfaction Survey

For data validation during import and manual data entry, we utilized the data collected as part of our use case (Step 2). We progressively compared the representation of the data in the database with the collected data in terms of concept, data type, values, and temporal alignment, and made step-by-step adjustments. A detailed description of this can be found in the “Results” section and in **Fig. 3**.

To evaluate user experience, we conducted a streamlined questionnaire-based survey, focusing on our DB front-end (see **Supplementary Table S2** [available in the online version only]). After a 30-minute introduction and an hour of independent work with UI, 10 participants responded to 18 questions on a 1 to 5 agreement scale.

## Results

This project involved developing an SQLite-based research database template and a user-friendly front-end for data entry. The database template, along with precompiled front-ends for Windows, macOS, and Linux, are available at the given GitHub repository: [https://github.com/stebro01/research\\_database\\_sqlite\\_i2b2.git](https://github.com/stebro01/research_database_sqlite_i2b2.git).

In this section, we present (1) expert interview findings on on-demand analysis, (2) DB structure, (3) implementation of data validation tools using real-time clinical data, (4) user feedback on our solution, and finally, (5) specific use cases for our solution.

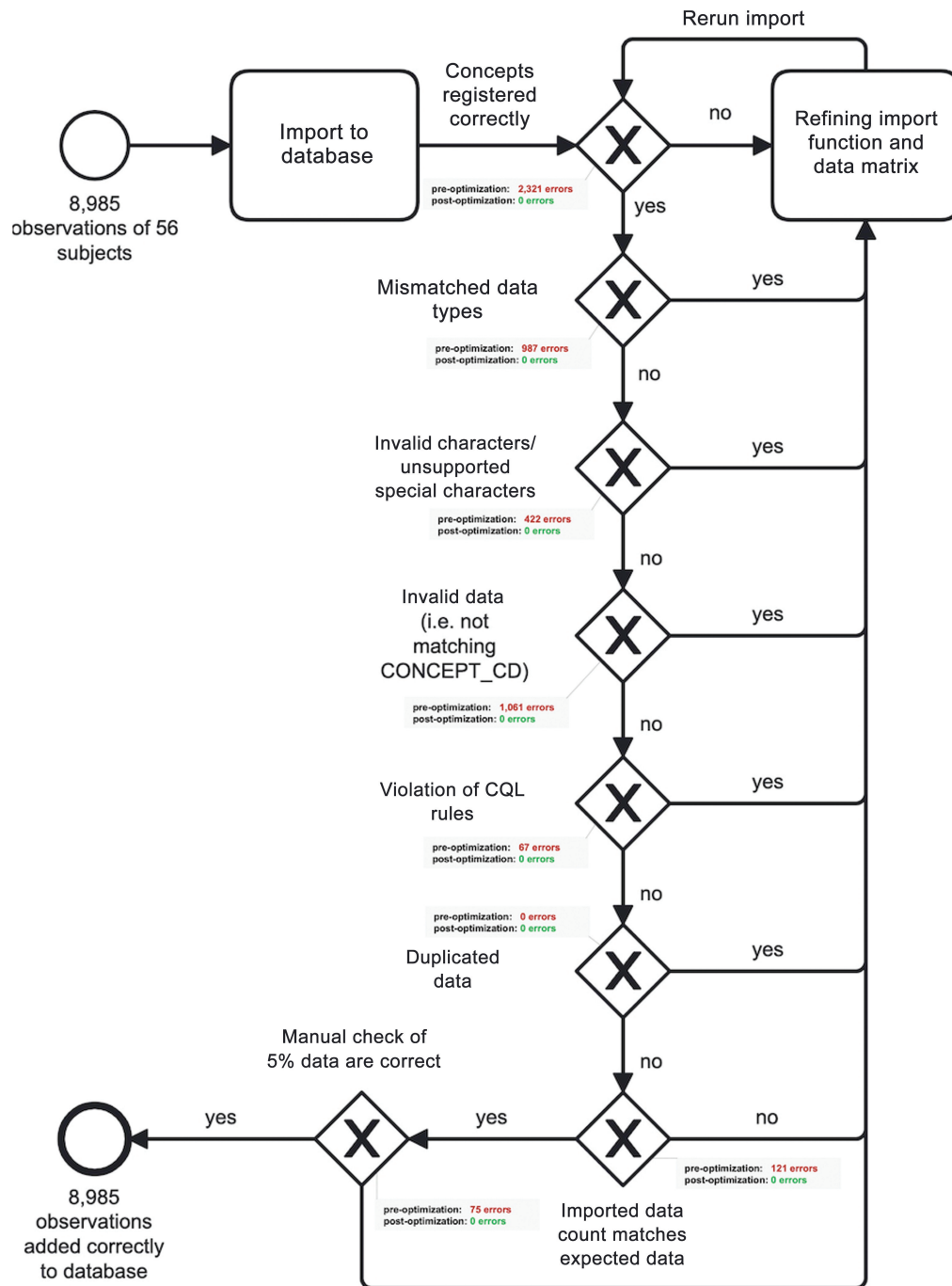
### Expert Interview Findings on On-demand Analysis

All respondents were directly involved in the design and conduct of experimental and clinical trials. Their level of experience was subjectively considered to be at least intermediate. The essential requirements identified by the respondents are presented in **Table 1**. A key finding from the interviews was that none of the respondents used a standardized method or system for managing clinical research data. All participants expressed a strong interest in a software solution easily implemented within their existing IT infrastructure.

### Evaluation of Existing Tools and Rationale for Selection

To assess existing solutions, we conducted comprehensive research, primarily focusing on EDC tools; the findings are summarized in **Table 2**. While we identified several potential candidates, the following outlines our reasoning against their selection.

REDCap,<sup>15</sup> for instance, is a renowned tool for capturing and managing data, particularly in large clinical studies. However, it falls short on several of our criteria. Firstly, REDCap operates as a server-based solution with cloud storage, which contradicts our requirement for offline capabilities. Secondly, any changes to the data structure require approval from the REDCap team, limiting flexibility. Lastly, it lacks built-in support for clinical classification systems like SNOMED-CT and LOINC and has a steep learning curve. DataLad,<sup>14</sup> on the other hand, is a free, open-source platform allowing users to manage data on their local machines. While it offers advanced features like data versioning and structured storage, its technical demands and command line-based features make it less user-friendly. Moreover, it lacks specialized support for clinical



**Fig. 3** Step-by-step import process demonstrating the reduction in error rates throughout each stage of the data import procedure. CQL, Clinical Quality Language.

classification systems. OpenClinica<sup>20</sup> is another notable tool that offers both free and paid versions. We ruled it out primarily due to the lack of explicit support for clinical classification systems and their IT requirements.

## DB Structure and Features

### DB Structure

Built upon the i2b2 CDM<sup>18</sup> star schema, our database structure, detailed in [Supplementary Table S3](#) [available in the online version only], integrates specific tables, views, and

triggers. **A comprehensive technical description is available in both the [Supplementary Materials](#) and the associated [GitHub repository](#).**

Utilizing the i2b2 schema, our data structure centers around the OBSERVATION\_FACT table, with extensions for auditing purposes, such as IMPORT\_DATE, DOWNLOAD\_DATE, UPDATE\_DATE, and UPLOAD\_ID. **In the current release, auditing primarily involves timestamping new data (via IMPORT\_DATE) and tagging it with the creator's ID (UPLOAD\_ID), as well as logging any data modifications (UPDATE\_DATE).** We have designed a table called "NOTE\_FACT"

**intended for more comprehensive auditing, which would capture differential data changes, data deletions and restorations, and database query access. This feature is currently inactive in the UI but is on the roadmap for future releases.**

#### *Data Concepts*

Data points are standardized in the CONCEPT\_DIMENSION table, utilizing globally recognized classifications like international statistical classification of diseases and related health problems (ICD-10), LOINC, and SNOMED-CT for interoperability, promoting data consistency and exchangeability across health care systems.<sup>21</sup>

#### *User Management*

We manage user access via the USER\_MANAGEMENT and USER\_PATIENT\_LOOKUP tables, controlling data access and permissions, ensuring data are only accessible to authorized users.

#### *Clinical Quality Language*

To facilitate quality management, a dedicated CQL\_FACT table stores CQL<sup>19</sup> rules and their JSON ELM representations. The CONCEPT\_CQL\_LOOKUP table links these rules to clinical concepts in the CONCEPT\_DIMENSION.

#### *User Feedback*

Finally, the NOTE\_FACT table collects user feedback and allows note creation for individual subjects, paving the way for future enhancements like reminder or appointment management systems or a more extensive auditing system.

#### **Front-end (Graphical User Interface)**

We have developed a UI that simplifies data input through comma-separated-values (CSV) imports or a CRF-like interface, with the graphical user interface (GUI) designed to focus on individual clinical visits, as demonstrated in **→Fig. 2**. Further, the GUI grants administrative control over user, provider, and location tables, in addition to managing CONCEPTS, inclusive of a SNOMED application programming interfaces (API) link. Our UI also streamlines data exchange processes by facilitating the import/export of all nonobservational data in JSON format.

Observational data can be imported and exported via a standardized CSV file. **Additionally, we have incorporated HL7 JSON support for interchange. In the current version, we specifically focus on the HL7-CDA (version 2.0.1) standard, limiting our support to the “Composition” resource type. Within this resource, we employ properties such as “subject,” “event,” and “section” to encapsulate relevant patient and observational data.**

By default, observational data are exported in a pseudonymized manner (creating a UID for each exported object).

#### **Versatile Data Capture for Clinical and Neuroscience Research**

Our RDM system has been designed with a strong focus on versatility, particularly for the collection of diverse clinical and neuroscientific data types. The database accepts standard text,

numeric, and date formats, providing a foundation for the collection of behavioral data, psychometric assessments, and clinical findings. It is also equipped to handle raw data types, facilitating the storage of images, PDF documents, and specialized reports. Our data schema covers a wide range of research needs, including laboratory values and subject-specific documents such as privacy statements. Using SQLite's capabilities, the database is theoretically capable of storing more complex datasets such as neuroimaging and genetic information. However, it is important to note that we have not yet validated the system's performance with these larger datasets.

#### **Implementation and Application of Data Validation to Real-world Clinical Data (Use Case)**

We have introduced real-time data validation, incorporating type checking, CQL rule enforcement, and duplicate checks to ensure data integrity directly into the data entry and import routines. **In the current version, the class responsible for adding data to the database performs double-entry checks. If a given subject already has a specific concept with the same value, the user is notified within the UI. At this point, the user has the option to either skip the entry or proceed with adding the data. However, since the data are stored in an SQLite database, double entries can also be managed directly through structured query language (SQL) statements.**

We implemented these features incrementally using real clinical data from our use case. In total, we used a data matrix with 8,985 data entries from 56 subjects, consisting of 82 different types (concepts). The iterative development of the import function, illustrated in **→Fig. 3**, systematically addressed data types, concept-conforming data, and erroneous data like invalid character sequences or unsupported special characters.

We logged errors and their frequency at each optimization stage, only modifying the initial data matrix when necessary (e.g., incorrect data type or significant misspelling of concepts/answers). Alterations to the original Excel spreadsheet were successively stored and documented via the KNIME analytics platform.<sup>22</sup>

#### **User Feedback on Our Solution**

Overall, the survey (**→Supplementary Table S4**, available in the online version only) results indicated a high level of user satisfaction. Most respondents found the system easy to use and the majority expressed confidence in their ability to use the system without technical support. The integration of different functions within the system was well received and users reported that they could learn to use the system quickly. Users found the process of entering new subjects, visits, and observations simple and straightforward. The layout was considered user-friendly and provided a good overview of relevant data. Users also found it easy to export data from the system.

#### **Specific Use Cases for Our Solution**

Herein, we delineate potential application scenarios of our database concept for clinical research queries, with the

decision-making process illustrated in **►Supplementary Fig. S1** (available in the online version only). First, a data scheme (like a CRF) is created akin to designing an Excel sheet. More concepts can be added to the database, currently comprising over 800 in the CONCEPT\_DIMENSIONS table, if required. The researcher can then store data in the database through the front-end, with three primary constellations:

1. Single user: Ideal for clinical researchers with limited subjects, where the user creates a CRF, store data in their local SQLite database, and directly manage the data with our solution's database structure and front-end.
2. Multiple users: A shared central data repository stored on a network drive, where users only access their created subjects, while an administrator can collate and evaluate all data. Role-specific rights control data access, and our solution also facilitates data merging from multiple users for joint analysis.
3. Multiple users/separate DB: Each user processes different subjects in a local DB version, with data later combined into a main database using the HL7 export and import function. Suitable for users at disparate locations needing later database consolidation, our solution provides an HL7 export function for this purpose.

Challenges in scenarios 2 and 3 may involve ensuring data consistency and integrity when merging data from multiple users and databases, necessitating clear data management and validation rules. **In scenarios where multiple users are working on the same database, there is the potential for data conflicts when modifying, adding, or deleting the same data. A possible solution to this problem is to implement a locking mechanism for individual subjects or visits. When a user is editing, the data would be set to "read-only" for other users; this feature is currently under development. In addition, the current version lacks an "undo" function, which will be included in future updates. Merging data from different sources also presents challenges, particularly when dealing with new data, matching data with identical identifiers, or managing different versions of the same data. Currently, the import process is straightforward: new data are added if no matching subject exists, and existing data are appended if it does. If conflicting data with the same creation date are detected, the version with the most recent update is used. However, this approach may not be suitable for all cases, so users are encouraged to implement custom SQL statements or code as needed.**

Finally, as the database can be stored on a local or network drive, it can be accessed remotely by various means, including remote desktop or network drive access, making it suitable for home office setups.

## Discussion

Our work addresses the pressing need for a comprehensive data platform that simplifies the management of clinical research data, with a special emphasis on the needs of clinicians involved in research. Our SQLite-based database

solution emerges as a user-friendly, secure, and efficient platform that particularly shines in scenarios where sophisticated server architectures or complex infrastructures are not readily available.

At the inception of our project, we embarked on a comprehensive needs analysis and rigorous internet research to identify the specific requirements and shortcomings of the existing alternatives. This inquiry revealed a pressing demand for a straightforward and intuitive system for data collection and management. The prevalent solutions often suffered from complex installation processes and operational challenges, making them ill-suited for small- and medium-sized projects. They frequently turned out to be excessively complicated and cumbersome, thereby diminishing their appeal for projects of smaller scale.

Acknowledging this gap, our aim was to develop a simpler, user-friendly alternative that offered ease of installation and manageability. Our concept, tested and validated in a clinical use-case scenario, was designed to fulfill the exigencies of real-world practice. To fine-tune our system and cater to user needs effectively, we conducted a user satisfaction survey at the culmination of the project. The invaluable feedback thus gathered will aid us in further refining and optimizing our system.

## What do Researchers Need?

Scientific researchers often play multiple roles including data collector, manager, and analyst.<sup>23</sup> However, researchers usually lack proficiency in data management, even though it is crucial in the latter part of the research workflow.<sup>4</sup> Although researchers are experts in their respective fields, their grasp of RDM is usually subpar, as revealed by studies and interviews. Many are unaware of standardized RDM tools and methods. Our survey and another study<sup>1</sup> revealed a need for defined data and metadata standards, set workflows, and a reliable RDM infrastructure. The lack of standardization and automated data exporting and anonymizing often hinders public data sharing. For instance, less than 40% of functional neuroimaging data are openly shared.<sup>23</sup> Implementing RDM standards can enhance the quality and volume of scientific publications.<sup>24</sup>

Nonetheless, many researchers see themselves as key to improving RDM and implementing open science practices.<sup>25</sup> Efforts are underway in Germany, like the establishment of an NFDI, targeting fields such as clinical neuroscience to ensure GDPR-compliant research.<sup>2</sup> The complex process of adopting an RDM culture integrates technological, economic, and political aspects. The increasing relevance of the topic is evidenced by a surge in scientific publications since 2016.<sup>26</sup> Our local project aims to connect to evolving data infrastructure interfaces, emphasizing the use of standard classifications and established representation models for integration.

Our expert interviews with clinical scientists engaged in daily patient care highlighted a persistent need for research tools that seamlessly integrate into existing workplace infrastructures. These tools must effectively compete with ubiquitous applications like Excel. For many researchers,

transitioning between their professional and research workspace is not as straightforward as one might assume. Consequently, there is a high demand for a bespoke database solution that functions as a “one-click” alternative to Excel, simplifying the user experience while facilitating scientific inquiry.

### Research Data Management in Scientific Practice

In a recent review from 2016, Perrier and colleagues<sup>27</sup> analyzed a total of 301 articles that examined RDM procedures in academic institutions. They were able to show that most of the work deals with the creation and initial storage of data (creating data and processing data as described, for example, in the Research Data Lifecycle Framework of the UK Data Archive). It is precisely this aspect of the scientific workflow that we are trying to improve here.

When it comes to managing clinical research data, researchers have a range of options at their disposal. These options include simple spreadsheet-based solutions to more complex database systems that demand advanced technical expertise to operate. Excel spreadsheets are one of the most used solutions by clinical researchers.<sup>28</sup> However, despite being a popular tool for data analysis, Excel lacks the required structure and functionality to manage complex clinical research data. Large datasets can quickly render Excel sheets unwieldy and error-prone, with error rates reported to be as high as 7 to 80%.<sup>29,30</sup> Moreover, spreadsheets do not provide the necessary security and backup functionality needed for sensitive clinical data.

To address these limitations, several popular web-based solutions are available, such as REDCap,<sup>15</sup> OpenClinica,<sup>20</sup> and LORIS.<sup>8</sup> **Table 2** in the methodological section provides a systematic overview of the most common solutions.

### Implementing Findable, Accessible, Interoperable, and Reusable Principles and Data Sharing

When collecting clinical research data, researchers must adhere to well-defined standards and rules. The use of different International System of Units (SI) units alone presents a common nontrivial challenge. Only when a high standard is established during data collection can general concepts such as the FAIR principles<sup>5</sup> be effectively implemented.<sup>31</sup>

Our SQLite-based database solution uniquely addresses the challenges of clinical RDM by integrating user needs, technical requirements, and industry standards. It prioritizes clear storage structure, ease of access, data security, and rights management, ensuring an uncomplicated and efficient RDM process.

The solution aligns with the FAIR principles, employing standardized data and metadata formats, facilitating data sharing and integration, and securing role-based access control. Each data observation is linked directly to a clinical concept and a visit, enriching it with valuable metadata, thus aligning with the researchers’ “minimum requirement” approach for metadata entry.<sup>27</sup>

Our choice to integrate HL7, particularly HL7/JSON, into our SQLite database improves functionality and adherence to

FAIR principles. This globally recognized standard ensures seamless data exchange with other systems, facilitating efficient data sharing and collaboration.<sup>32</sup> The HL7/JSON integration also enhances data accessibility, allowing data to be easily parsed across platforms and applications due to its human-readable format.

Our SQLite database system offers a structured storage system for diverse clinical data types, enabling efficient search and filtering. This minimizes errors due to manual data manipulation. The user-friendly interface allows easy data management without demanding extensive technical knowledge, while the provision for CSV export supports further analysis.

Security and backup features safeguard sensitive clinical data against unauthorized access and loss. The database is stored on a network drive with defined user access, enabling secure, easy access from researchers’ clinical workplaces. This is especially beneficial for research in resource-limited settings where advanced server infrastructure may not be available.

The system is also designed for straightforward adoption. By simply customizing the SQLite DB template and installing the front-end, researchers can start data collection and always maintain direct control over their data, making our solution an effective tool in current scientific practice.<sup>2</sup> **Moreover, the system’s flexible design can easily adapt to future requirements and be repurposed for subsequent studies. Its standardized data template is not only conducive to data exchange but also effectively addresses the crucial need for interoperability.**

It is important to note that the current version of the database solution is a small, local project. The current iteration can be seen as a feasibility study designed to directly investigate its implementation, technical architecture, and user satisfaction. While it is possible to use this solution in larger clinical trials with a multicenter approach, such as shared network drives, it is not currently designed for this purpose. If the project is accepted and actively utilized by our local research groups, we would like to expand it further and implement new features. However, it is crucial to acknowledge that in the context of regulated clinical trials, an extensive certification process for the application would be required.

### Meeting a Real-world Scenario: Use Case with Longitudinal Data from Mild Cognitive Impairment Patients

To demonstrate the applicability of our database solution, we selected a use case involving the longitudinal tracking of data from patients with MCI. This real-world scenario includes a variety of data types, including sociodemographic, clinical, laboratory, and neuropsychological data, as well as scores and parameterized neuroimaging data from MRI, CT, and electrophysiological studies such as EEG. We aimed to accommodate multiple classification standards, such as ICD-10, LOINC, and SNOMED-CT, and manage data collected at different points in time.

While the SQLite database is theoretically capable of storing large files, our initial implementation focuses on storing structured reports of processed neuroimaging data. Specifically, we are focusing on VBM, SBM, or Functional

magnetic resonance imaging (fMRI) analyses presented in JavaScript Object Notation (JSON) or Extensible Markup Language (XML) formats. These structured reports can be fully integrated into the database using custom concept definitions.

### Future Directions and Improvements

Looking to the future, there are several potential enhancements and improvements that could be made to the proposed database solution. One possible avenue for development is the direct integration with open science data hubs, allowing for seamless data sharing and collaboration across multiple institutions. This would enable researchers to easily contribute and access data from a centralized repository, promoting greater transparency and reproducibility in research.

In addition to open science integration, another potential area for improvement is the export functionality of the database solution. Currently, the solution supports CSV and HL7 JSON export formats, but there is room for expansion to other popular analysis tools such as SPSS, Python, and R. By expanding the range of supported export formats, users will have greater flexibility in conducting further analysis on the collected data.

Exploring the implementation of a Fast Healthcare Interoperability Resources (FHIR) interface could be very beneficial. As an established standard for electronic health information exchange, FHIR could significantly improve the interoperability of our database. This would streamline the exchange and integration of data across different health care systems and platforms. In this context, Translational Research Informatics and Data-management grid's focus on service-oriented architecture, and proven interoperability strategies offer valuable insights.<sup>33</sup>

Another potential area for improvement is the implementation of automatic lifecycle management and backup routines. Currently, these tasks are performed manually by the database administrator, which may be time-consuming and prone to human error. By automating these tasks, the database solution can ensure greater data consistency and reliability.

One significant limitation of our database solution is that our primary consultation was with clinical scientists who have limited experience in RDM. However, the inclusion of many senior scientists in our consultations lends credence to the expressed need for a solution that is easy to implement. This would enable them to integrate RDM practices seamlessly into their existing workflows. Additionally, the transparency of data storage in a local database file may alleviate concerns around data privacy and ownership, potentially making our system more readily acceptable than preexisting, highly integrated alternatives.

### Conclusion

In this study, we have developed and evaluated a user-friendly SQLite database with a front-end for streamlined clinical RDM.

While we do not consider our database solution to be superior to existing systems that address FAIR principles, it

does offer distinct advantages in certain contexts. Our system is designed as a “one-click” implementation with a locally stored SQLite database, offering a straightforward setup for clinical scientists lacking database management expertise. By lowering the entry barriers in this way, our solution serves as a catalyst for establishing RDM practices in labs that might otherwise be hindered by technical complexity or budget constraints. Consequently, our system adds a layer of FAIR compliance to research environments that may currently lack it, enhancing the overall FAIR landscape.

A user-satisfaction survey confirmed high acceptance among our target group. However, further refinement and evaluation are needed to optimize performance, usability, and data security across varied research applications.

### Clinical Relevance Statement

- Our SQLite-based database delivers a user-friendly, effortlessly installable solution for research data storage, processing, standardization, and sharing.
- This database proficiently manages multimodal and longitudinal data, and seamlessly integrates clinical classification systems like SNOMED-CT and LOINC.
- By adhering to FAIR principles and implementing standards for data storage and exchange like HL7-CDA JSON, our database not only streamlines RDM but also enhances the reproducibility and publication of scientific research.

### Multiple Choice Questions

**Question 1:** Which statement accurately reflects the use and limitations of Excel spreadsheets in managing clinical research data?

- a. Excel spreadsheets are ideal for managing complex clinical research data due to their advanced functionality and structure.
- b. Large datasets can make Excel sheets prone to high error rates, ranging from 7 to 80%.
- c. Spreadsheet-based solutions provide the necessary security and backup functionality for sensitive clinical data.
- d. Clinical researchers seldom use Excel spreadsheets for data management.

Answer: b.

**Question 2:** Which of the following statements best represent one of the FAIR principles for data management?

- a. The “Findable” principle emphasizes the use of persistent identifiers (e.g., DOIs) and rich metadata to ensure that datasets are uniquely identifiable and easily discoverable by both humans and machines.
- b. The “Secure” principle focuses on safeguarding data integrity, confidentiality, and availability.
- c. The “Exclusive” principle emphasizes controlled access to data, allowing only a select group of individuals or organizations to access and use the data.

- d. The “Closed” principle refers to restricting data access to a specific group or organization, thereby limiting its availability, and preventing broader reuse.

Answer: a.

#### Protection of Human and Animal Subjects

The study adhered to the World Medical Association Declaration of Helsinki’s ethical guidelines for research involving human subjects and received approval from our local Ethics board (Reference: 2022-2658-Daten).

#### Funding

This work was supported by funding from the Foundation “Else Kröner-Fresenius-Stiftung” within the Else Kröner Graduate School for Medical Students “Jena School for Ageing Medicine (JSAM)” and Else Kröner Anti Age.

#### Conflict of Interest

None declared.

#### References

- Klingner CM, Denker M, Grün S, et al. Research data management and data sharing for reproducible research—results of a community survey of the German National Research Data Infrastructure Initiative Neuroscience. *eNeuro* 2023;10(2):ENEURO.0215-22.2023
- Wachtler T, Bauer P, Denker M, et al. NFDI-Neuro: building a community for neuroscience research data management in Germany. *Neuroforum* 2021;27(01):3–15
- Denker M, Grün S, Wachtler T, Scherberger H. Reproducibility and efficiency in handling complex neurophysiological data. *Neuroforum* 2021;27(01):27–34
- Amorim RC, Castro JA, Rocha da Silva J, Ribeiro C. A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Univers Access Inf Soc* 2017;16(04):851–862
- Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018
- Roche DG, Kruuk LEB, Lanfear R, Binning SA. Public data archiving in ecology and evolution: how well are we doing? *PLoS Biol* 2015;13(11):e1002295
- Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer’s disease: a systematic review. *Front Aging Neurosci* 2017;9:329
- Das S, Zijdenbos AP, Harlap J, Vins D, Evans AC. LORIS: a web-based data management system for multi-center studies. *Front Neuroinform* 2012;5:37
- Hussein R, Wurhofer D, Strumegger EM, et al. General Data Protection Regulation (GDPR) toolkit for digital health. *Stud Health Technol Inform* 2022;290:222–226
- Deutsche Forschungsgemeinschaft. Handling of Research Data. Accessed July 7, 2023 at: [https://www.dfg.de/en/research\\_funding/principles\\_dfg\\_funding/research\\_data/index.html](https://www.dfg.de/en/research_funding/principles_dfg_funding/research_data/index.html)
- Schelter JS, Abiteboul S, Howe B, Jagadish HV, Sebastian S. Responsible Data Management. Accessed July 7, 2023 at: <https://cacm.acm.org/magazines/2022/6/261173-responsible-data-management/fulltext>
- Duda SN, Kennedy N, Conway D, et al. HL7 FHIR-based tools and initiatives to support clinical research: a scoping review. *J Am Med Inform Assoc* 2022;29(09):1642–1653
- Moher D, Liberati A, Tetzlaff J, Altman DGPRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6(07):e1000097
- Halchenko YO, Meyer K, Poldrack B, et al. DataLad: distributed system for joint management of code, data, and their relationship. *J Open Source Softw* 2021;6(63):3262
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42(02):377–381
- Database Trends and Applications. The Role of ACID in the Integrity of Your Database Data. Published November 1, 2017. Accessed July 7, 2023 at: <https://www.dbta.com/Authors/Craig-S.-Mullins-3535.aspx>; <https://www.dbta.com/Columns/DBA-Corner/The-Role-of-ACID-in-the-Integrity-of-Your-Database-Data-121440.aspx>
- Bhosale S, Patil P. SQLite: Light Database System. *Int J Comput Sci Mobile Comput* 2015;44:882–885
- i2b2 Community Wiki. 2. Quick Start Guide - Bundles and CDM. Accessed March 30, 2023 at: <https://community.i2b2.org/wiki/display/BUN/2.+Quick+Start+Guide>
- HL7 Consortium. CQL - Clinical Quality Language Release 1. Accessed April 2, 2023 at: <https://cql.hl7.org/01-introduction.html>
- Cavelaars M, Rousseau J, Parlayan C, et al. OpenClinica. *J Clin Bioinforma* 2015;5(01):S2
- Bodenreider O, Cornet R, Vreeman DJ. Recent developments in clinical terminologies - SNOMED CT, LOINC, and RxNorm. *Yearb Med Inform* 2018;27(01):129–139
- Berthold MR, Cebron N, Dill F, et al. KNIME - the Konstanz information miner: version 2.0 and beyond. *SIGKDD Explor* 2009;11(01):26–31
- Lyon L. Dealing with data: roles, rights, responsibilities and relationships consultancy report. Published online January 1, 2007
- Hruby GW, McKiernan J, Bakken S, Weng C. A centralized research data repository enhances retrospective outcomes research capacity: a case report. *J Am Med Inform Assoc* 2013;20(03):563–567
- Checklists work to improve science. *Nature* 2018;556(7701):273–274
- Donner EK. Research data management systems and the organization of universities and research institutes: a systematic literature review. *J Librarian Inform Sci* 2022;09610006211070282
- Perrier L, Blondal E, Ayala AP, et al. Research data management in academic institutions: a scoping review. *PLoS ONE* 2017;12(05):e0178261
- Springer Publishing Connect. Introduction to Healthcare Data and the Role of Excel®. 2022. Accessed April 30, 2023 at: <https://connect.springerpub.com/content/book/978-0-8261-5028-8/chapter/ch01>
- Dobell E, Herold S, Buckley J. Spreadsheet error types and their prevalence in a healthcare context. *J Organ End User Comput* 2018;30(02):20–42
- Rajalingham K, Chadwick DR, Knight B. Classification of spreadsheet errors. 2008; eprint no.0805.4224:2-5
- Kush R, Goldman M. Fostering responsible data sharing through standards. *N Engl J Med* 2014;370(23):2163–2165
- Solbrig HR, Prud’hommeaux E, Grieve G, et al. Modeling and validating HL7 FHIR profiles using semantic web Shape Expressions (ShEx). *J Biomed Inform* 2017;67:90–100
- Payne P, Ervin D, Dhaval R, Borlawsky T, Lai A. TRIAD: The Translational Research Informatics and Data Management Grid. *Appl Clin Inform* 2011;2(03):331–344
- ClinCapture. Captivate® Electronic Data Capture (EDC). Accessed October 23, 2023 at: <https://www.clincapture.com/captivate-electronic-data-capture-edc/>

- 35 eNOV Software for Life. MACRO Accessed October 27, 2023 at: <https://en.ennov.com/solutions/clinical/macro/>
- 36 Gao QB, Kong Y, Fu Z, et al. EZ-Entry: a clinical data management system. *Comput Biol Med* 2008;38(09):1042–1044
- 37 Oracle Help Center. Oracle Clinical and Oracle Clinical Remote Data Capture Release 5.4. Accessed October 23, 2023 at: <https://docs.oracle.com/en/industries/health-sciences/clinical-remote-capture/5.4/index.html>
- 38 Brandt CA, Deshpande AM, Lu C, et al; Center for Medical Informatics. TrialDB: a web-based clinical study data management system. *AMIA Annu Symp Proc* 2003;2003:794
- 39 Hurtado S, García-Nieto J, Navas-Delgado I, Aldana-Montes JF. FIMED: Flexible management of biomedical data. *Comput Methods Programs Biomed* 2021;212:106496
- 40 Nguyen L, Shah A, Harker M, et al. DADOS-Prospertive: an open source application for web-based prospective data collection. *Source Code Biol Med* 2006;1(01):7
- 41 Ärzteblatt DÄG Redaktion Deutsches. PhOSCo (Pharma Open Source Community): Open Source für klinische Studien. *Deutsches Ärzteblatt* Published May 7, 2004. Accessed October 22, 2023 at: [https://www.aerzteblatt.de/archiv/41737/PhOSCo-\(Pharma-Open-Source-Community\)-Open-Source-fuer-klinische-Studien](https://www.aerzteblatt.de/archiv/41737/PhOSCo-(Pharma-Open-Source-Community)-Open-Source-fuer-klinische-Studien)
- 42 Lee H, Chapiro J, Scherthaner R, et al. How I do it: a practical database management system to assist clinical research teams with data collection, organization, and reporting. *Acad Radiol* 2015;22(04):527–533
- 43 Egbring M, Kullak-Ublick GA, Russmann S. Phynx: an open source software solution supporting data management and web-based patient-level data review for drug safety studies in the general practice research database and other health care databases. *Pharmacoepidemiol Drug Saf* 2010;19(01):38–44
- 44 Wolstencroft K, Owen S, Krebs O, et al. SEEK: a systems biology data and model management platform. *BMC Syst Biol* 2015; 9(01):33
- 45 Bauch A, Adamczyk I, Buczek P, et al. openBIS: a flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics* 2011;12(01):468
- 46 Sansone SA, Rocca-Serra P, Field D, et al. Toward interoperable bioscience data. *Nat Genet* 2012;44(02):121–126
- 47 Ferretti Y, Miyoshi NSB, Silva WA Jr, Felipe JC. BioBankWarden: a web-based system to support translational cancer research by managing clinical and biomaterial data. *Comput Biol Med* 2017; 84:254–261
- 48 Rossi E, Rosa M, Rossi L, Priori A, Marceglia S. WebBioBank: a new platform for integrating clinical forms and shared neurosignal analyses to support multi-centre studies in Parkinson's Disease. *J Biomed Inform* 2014;52:92–104
- 49 LabArchives. All Your Research in One Place. Accessed October 27, 2023 at: <https://www.labarchives.com/>
- 50 Labguru. Lab Management Software | Laboratory System. Accessed October 27, 2023 at: <https://www.labguru.com>
- 51 SciNote. Electronic Lab Notebook (ELN Software). Published July 18, 2023. Accessed October 27, 2023 at: <https://www.scinote.net/>
- 52 eLabNext. eLabJournal. Accessed October 27, 2023 at: <https://www.elabnext.com/products/elabjournal/>
- 53 RSpace. ELN & Inventory. Accessed October 27, 2023 at: <https://www.researchspace.com/>
- 54 eLabFTW. Open Source Laboratory Notebook. Accessed October 27, 2023 at: <https://www.elabftw.net>