Abstract

RareLink - An interoperability framework in REDCap for Rare Diseases linking Registry Use, FHIR, and Phenopackets

Adam SL Graefe[1, 2], Filip Rehburg[1], Samer Alkarkoukly[3], Alexander Bartschke[1], Daniel Danis[1], Ana Grönke[3], Miriam R Hübner[1], Steffen Sander[1], Sophie AI Klopfenstein[1], Daniel R Korn[4], Jana Zschüntzsch[5], Elisabeth F Nyoungui[6], Tatiana Kalashnikova[7], Beata Derfalvi[8], Nicola Wright[9], Susanna Wiegand[10, 11], Peter Kühnen[10, 12, 13], Melissa A Haendel[4], Sylvia Thun[1], Peter N Robinson[1, 14], Oya Beyan[2]

Affiliations

1. Berlin Institute of Health at Charité, Universitätsmedizin Berlin, Germany
2. Institute for Biomedical Informatics, University Hospital Cologne, Germany
3. Data Integration Center, Institute for Biomedical Informatics, University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne, Germany
4. Department of Genetics, University of North Carolina Chapel Hill, USA
5. Department of Neurology, University Medical Center Goettingen, Germany
6. Department of Medical Informatics, University Medical Center Goettingen, Germany
7. Clinical Research Fellow in Clinical Immunology/University of Calgary, Canada
8. Professor/IWK Health Centre/Dalhousie University, Canada
9. Clinical Associate Professor/University of Calgary, Canada
10. Department of Paediatric Endocrinology and Diabetology, Charité Universitätsmedizin, Berlin, Germany
11. Center for Chronically Sick Children, Charité Universitätsmedizin Berlin, Germany
12. Berlin Center for Rare Diseases - Charité University Hospital, Berlin, Germany
13. Deutsches Zentrum für Kinder- und Jugendgesundheit (DZKJ), Partner Site Berlin, Germany
14. The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA

corresponding author: Adam SL Graefe (adam.graefe@charite.de)

While rare diseases (RDs) affect over 260 million individuals worldwide, low data quality and scarcity significantly challenge effective care and research. REDCap is an open-source software clinicians use for RD registries, studies, and research. However, REDCap's largely unconstrained format for defining projects, variables, and data dictionaries means that each REDCap installation may have its own data schema, which in practice limits data exchange and reuse, hampering the development of standard software for downstream analysis. *RareLink* is a novel framework in REDCap for RD data management and processing with comprehensive documentation for clinicians to use the software. It pre-defines a local project structure, including its data collection instruments, data dictionaries, and variables. The embedding of ontology codes and the connection to the BioPortal Ontology Server enables international data standards for streamlined bioinformatic analysis and data export via HL7 FHIR and the Global Alliance for Genomics and Health (GA4GH) Phenopacket Schema.

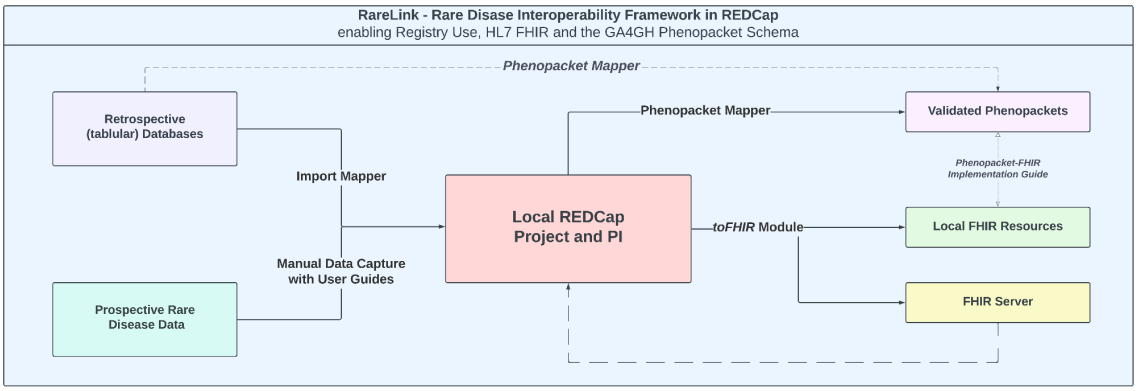
Based on our Rare Disease Common Data Model 2.0[1], we developed an entire project structure that can be uploaded to a local REDCap server. We implemented our framework in multiple German university hospitals and the Canadian Inborn Errors of Immunity National Registry (CIEINR)[2]. This helped improve our framework and its documentation, fix implementation-specific errors, and ensure correct medical semantics. To account for disease-specific extensions beyond our common

[1] https://github.com/BIH-CEI/rd-cdm / RD CDM Docs: https://rd-cdm.readthedocs.io/en/latest/
[2] https://www.sciencedirect.com/science/article/pii/S1521661624001268

data model, the extensions on rare metabolic diseases in Germany and the CEIENR helped to provide guidelines for developing further data capture forms that can be processed by our framework.

Prospective data can be manually captured using our data capture sheets, with detailed documentation to ensure accurate and consistent data entry. A semi-automated import process is available by adapting the Phenopacket Mapper that maintains data locally for retrospective data stored in tabular RD databases, such as Excel sheets. However, local semantic encoding is still required during this process. Once the local patient data is prepared, pipelines can be executed on-site to generate HL7 FHIR resources with the toFHIR module, and GA4GH Phenopackets with the Phenopacket Mapper. Adhering to these standards facilitates the linkage of registry data with other data centres, studies, and more. Additionally, exporting data to GA4GH Phenopackets supports the reusability of analysis pipelines, such as the GPSEA (genotype-phenotype: statistical evaluation of associations) tool (Rekerle, Danis et al., manuscript in preparation), enhancing the efficiency and breadth of data analysis.

In conclusion, this framework provides a comprehensive system for RD specialists and clinicians, covering the entire process from data capturing and processing to exporting data in compliance with international standards. Although some coding experience is still necessary in some instances, the setup of the REDCap project and forms can be managed with the assistance of a local REDCap administrator. Utilising RareLink for RD data management enhances the reusability of analysis pipelines, facilitates connections between registries and healthcare information systems, and supports various future applications. Our framework is open-source to accommodate the limited resources often available in RD care and research, making it accessible for clinicians without financial barriers.



- **GitHub Repository:** https://github.com/BIH-CEI/RareLink
- **RareLink Documentation:** https://rarelink.readthedocs.io/en/latest/
- **GPSEA Documentation**: https://monarch-initiative.github.io/gpsea/stable/index.html
- **Phenopacket Mapper:**
    - GitHub: https://github.com/BIH-CEI/phenopacket_mapper
    - Documentation: https://bih-cei.github.io/phenopacket_mapper/latest/
    - Abstract: https://www.overleaf.com/read/jmjqjqsxjhtg#bd0288
- Why digital medicine depends on interoperability (S. Thun):
    - https://www.nature.com/articles/s41746-019-0158-1
- The GA4GH Phenopacket Schema (PN Robinson):
    - https://www.nature.com/articles/s41587-022-01357-4