Healthcare Transformation with Informatics and Artificial Intelligence
J. Mantas et al. (Eds.)

doi:10.3233/SHTI230515

Representing NIH Genetic Test Registry Data in the FHIR Genomic Study Resource

ALY KHALIFA[a,b,1] and ROBERT R. FREIMUTH[a,b]
[a] *Department of Artificial Intelligence and Informatics*
[b] *Center for Individualized Medicine*
Mayo Clinic, Rochester, Minnesota
ORCiD ID: Aly Khalifa https://orcid.org/0000-0002-7084-1345, Robert Freimuth https://orcid.org/0000-0002-9673-5612

Abstract. The National Institute of Health (NIH) Genetic Testing Registry (GTR) provides a variety of information about genetic tests such as relevant methods, conditions, and performing laboratories. This study mapped a subset of GTR data to the newly developed HL7®-FHIR® Genomic Study resource. Using open-source tools, a web application was developed to implement data mapping and provides many GTR test records as Genomic Study resources. The developed system demonstrates the feasibility of using open-source tools and the FHIR Genomic Study resource to represent publicly available genetic testing information. This study validates the overall design of the Genomic Study resource and proposes two enhancements to support additional data elements.

Keywords. Genetic Testing, Standards, FHIR, Genomic Study

1. Introduction

Precision medicine plays a key and growing role in personalizing health care [1]. Genetic testing, a fundamental component of the precision medicine, covers a wide range of diseases and is becoming increasingly common due to steady reductions in sequencing costs [2]. Interoperable and standards-based representation of genetic data will enable deeper utilization of genomic data for both clinical practice and research [3]. Many challenges still exist, however, including the representation of complex genomic data and competing business priorities that impact adoption of standards during data generation [3]. Additional challenges exist in the integration and storage of genomic data in Electronic Health Records (EHRs) [2, 4]. Heterogeneity in supporting information systems also makes it difficult to adopt common approaches for sharing genetic data [5].

Health Level Seven (HL7®) Fast Healthcare Interoperability Resources (FHIR®) standard provides specifications for the representation of clinical concepts and the relations between them, e.g., patient, condition, and observation. These concepts are structured as information models called resources. Where each resource has a set of predefined data elements and associated constraints, e.g., data type and cardinality. In FHIR R5, the new official release, a new resource is included to better support complex

[1] Corresponding Author: Aly Khalifa, Department of Artificial Intelligence and Informatics. Email: khalifa.aly@mayo.edu.

genomic testing. The Genomic Study resource aims to capture important metadata about analyses that comprise clinical and research testing and analytical pipelines [6]. For example, the Genomic Study resource includes structured data elements that describe patient subjects, the reasons for study, and the testing protocols that were performed. It also provides links to the FHIR Observation, Encounter, Plan Definition and other FHIR resources to create a web of knowledge about a specific genomic testing use case.

The National Institute of Health (NIH), National Center of Biotechnology Information (NCBI) maintains the Genetic Testing Registry (GTR), which includes self-reported records about the genetic tests performed by various labs [7, 8]. Each record includes structured information about the relevant conditions, testing methodology, related genes, and other data that are relevant to the test.

This study aimed to map a subset of NIH-GTR data to the newly developed Genomic Study resource, and to use those mappings and open-source tools to develop a web application that exposes GTR data as FHIR Genomic Study resource instances.

2. Methods

For the mapping, a subset of GTR data (including "Test version history" and "disease names" files) were downloaded from the GTR web site [9, 10]. Online GTR webpages and the field definition document were used to identify and confirm the meaning of each column in the retrieved data set. Relevant GTR fields were identified and mapped to semantically comparable elements of the FHIR Genomic Study resource. Two approaches to mapping were followed: the first approach modeled genomic tests as an instance of Genomic Study (base resource); and the second approach modeled genomic tests as instances of the analysis component within a parent Genomic Study.

A Java web application, backed by the open-source PostgreSQL database, was developed [11]. The application used the open-source HAPI FHIR client library to build the FHIR resource [12], and the OpenApi Specification 3.1.0 (Swagger) to briefly document the developed Application Programming Interfaces (APIs) and to provide an interactive web page for these APIs and how to use them [13].

3. Results

We successfully mapped many relevant data fields (i.e., columns) in the GTR records to corresponding data elements within the FHIR Genomic Study resource. The mapping process identified two gaps in the current structure of the Genomic Study resource related to the representation of genes studied and pharmacological responses. The conditions codes used by the GTR were sourced from Mondo Disease Ontology, OMIM, and Human Phenotype Ontology, all of which could be supported by the Genomic Study resource.

Figure 1 depicts the general architecture of the developed web application that serves various APIs that can be leveraged by healthcare application developers. In addition to generating instances of the Genomic Study resource, the system also serves the raw GTR data and provides a search interface based on test accession, lab test identifier, and gene names. The full mapping file and the application are available in the GitHub repository https://github.com/InformaticsGenomicMedicine/Genomic_Study_GTR_pilot.

4. Discussion

We successfully mapped the core data elements of the GTR and proposed two extensions to represent information that are not yet fully represented in the Genomic Study resource. These enhancements may be considered as part of the next version of the resource. The mappings developed here are a first step in the validation of the new Genomic Study resource and facilitate the incorporation of genetic testing data within various EHR systems that will implement FHIR R5 specification. The availability of genetic testing data and its linkage to other concepts represented by FHIR resources would facilitate the development of clinical decision support systems [14], and it may motivate research projects and drug discovery [15].

While this study demonstrates promising results, it has various limitations, including that only a subset of the GTR data was mapped and the application runs locally. Those limitations may be addressed in future work.

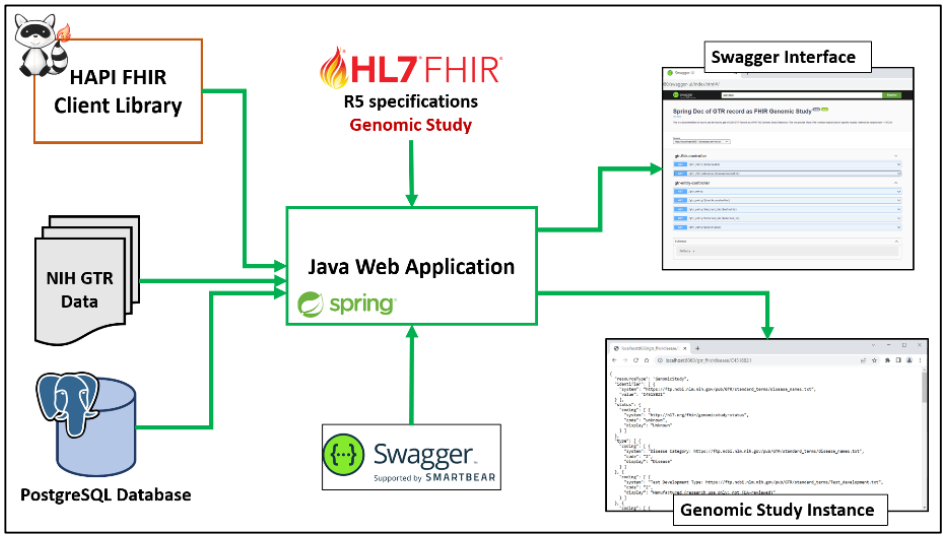

Figure 1. General application architecture

5. Conclusions

This study demonstrates that the FHIR Genomic Study resource can represent genomic testing data. This study successfully mapped a subset of GTR data to the FHIR Genomic Study resource and developed a web application that implemented the mapping. Two gaps were identified in the Genomic Study resource, which were proposed as enhancements.

This study shows how standard specifications, open-source tools, and public knowledge bases can be used to efficiently develop standards-based tooling that exposes public data sets. The continued development of standards and systems to exchange genomic data for clinical and research purposes will help advance precision medicine.

References

[1] Denny JC, Collins FS. Precision medicine in 2030-seven ways to transform healthcare. Cell. 2021;184(6):1415-9. Epub 2021/03/20. doi: 10.1016/j.cell.2021.01.015.

[2] Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. Nat Rev Genet. 2013;14(6):415-26. Epub 2013/05/18. doi: 10.1038/nrg3493.

[3] Khalifa A, Mason CC, Garvin JH, Williams MS, Del Fiol G, Jackson BR, et al. A qualitative investigation of biomedical informatics interoperability standards for genetic test reporting: benefits, challenges, and motivations from the testing laboratory's perspective. Genet Med. 2021;23(11):2178-85. Epub 2021/08/26. doi: 10.1038/s41436-021-01301-y. PubMed PMID: 34429527.

[4] Carter AB, Abruzzo LV, Hirschhorn JW, Jones D, Jordan DC, Nassiri M, et al. Electronic Health Records and Genomics: Perspectives from the Association for Molecular Pathology Electronic Health Record (EHR) Interoperability for Clinical Genomics Data Working Group. J Mol Diagn. 2022;24(1):1-17. Epub 2021/10/18. doi: 10.1016/j.jmoldx.2021.09.009. PubMed PMID: 34656760.

[5] Khalifa A, Mason CC, Garvin JH, Williams MS, Del Fiol G, Jackson BR, et al. A qualitative study of prevalent laboratory information systems and data communication patterns for genetic test reporting. Genet Med. 2021;23(11):2171-7. Epub 2021/07/08. doi: 10.1038/s41436-021-01251-5. PubMed PMID: 34230635.

[6] Health Level Seven International. FHIR R5 Genomic Study Resource: HL7; 26 March 2023 [cited 27 March 2023]. Available from: https://hl7.org/fhir/R5/genomicstudy.html.

[7] National Institute of Health. Genetic Testing Registry: NIH; 2023 [cited 18 March 2023]. Available from: https://www.ncbi.nlm.nih.gov/gtr/.

[8] Rubinstein WS, Maglott DR, Lee JM, Kattman BL, Malheiro AJ, Ovetsky M, et al. The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency. Nucleic Acids Res. 2013;41(Database issue):D925-35. Epub 2012/11/30. doi: 10.1093/nar/gks1173. PubMed PMID: 23193275.

[9] National Institute of Health. Genetic Test Registry - Test version history. In: NIH, editor. 2023.

[10] National Institute of Health. Genetic Test Registry - Disease Names. In: NIH, editor. 2023.

[11] The PostgreSQL Global Development Group. PostgreSQL: PostgreSQL; 2023 [18 March 2023]. Available from: https://www.postgresql.org/.

[12] Smile Digitial Health. HAPI FHIR: Smile CDR; 2023 [cited 18 March 2023]. Available from: https://hapifhir.io/.

[13] Swagger. OpenAPI Specification: Swagger; 2023 [cited 18 March 2023]. Available from: https://swagger.io/specification/.

[14] Dolin RH, Boxwala A, Shalaby J. A Pharmacogenomics Clinical Decision Support Service Based on FHIR and CDS Hooks. Methods Inf Med. 2018;57(S 02):e115-e23. Epub 2019/01/04. doi: 10.1055/s-0038-1676466. PubMed PMID: 30605914.

[15] Hernandez S, Fairchild K, Pemberton M, Dahmer J, Zhang W, Palchuk MB, et al. Applying FHIR Genomics for Research - From Sequencing to Database. AMIA Annu Symp Proc. 2022;2022:236-43. Epub 2022/07/21. PubMed PMID: 35854733.