Personalized Medicine Through Personal Data Pods

Elias Crum[1,2,*], Ruben Taelman[1], Bart Buelens[2], Gokhan Ertaylan[2] and Ruben Verborgh[1]

[1]*IDLab, Department of Electronics and Information Systems, Ghent University – imec, Belgium*
[2]*Flemish institute for Technological Research (VITO) Mol, Belgium*

Abstract

Medical care is in the process of becoming increasingly personalized through the use of patient genetic information. Concurrently, privacy concerns regarding collection and storage of sensitive personal genome sequence data have necessitated public debate and legal regulation. Here we identify two fundamental challenges associated with privacy and shareability of genomic data storage and propose the use of Solid pods to address these challenges. We establish that personal data pods using Solid specifications can enable decentralized storage, increased patient control over their data, and support of Linked Data formats, which when combined, could offer solutions to challenges currently restricting personalized medicine in practice.

Keywords

Genomic Data Privacy, Solid, Solid Pods, Linked Data

1. Problem Identification

Human genome sequencing advancements have allowed for early personalized medicine applications, such as in drug development and prescription [1]. With the cost of human genome sequencing continuing to decrease in recent years, personal genome sequences (PGS) are becoming an increasingly accessible and powerful tool within healthcare [2]. Importantly, PGS data is exceedingly private, thus being classified as "sensitive data" in the 2018 European Union's General Data Protection Regulation (GDPR) legislation.

Two main perspectives on the issue of PGS data storage and usage are exceedingly important moving forward – the patient and the clinician.

From the patient perspective, PGS data collection and usage will necessitate scalable, secure storage and sharing. Currently, significant amounts of PGS data are stored in centralized databases that are inaccessible to patients [3]. Thus, patients have little control over their genomic data or transparency about how it is used. In addition, centralized databases are consistently cited in data leaks where malicious actors are able to gain access and even publicize stored sensitive data.

SWAT4HCLS 2024: Bridging Life Sciences and Technology, February 26-29, Leiden, The Netherlands
*Corresponding author.
elias.crum@ugent.be (E. Crum); ruben.taelman@ugent.be (R. Taelman); bart.buelens@vito.be (B. Buelens); gokhan.ertaylan@vito.be (G. Ertaylan); ruben.verborgh@ugent.be (R. Verborgh)
0009-0005-3991-754X (E. Crum); 0000-0001-5118-256X (R. Taelman); 0000-0001-7734-3747 (B. Buelens); 0000-0001-5602-6435 (G. Ertaylan); 0000-0002-8596-222X (R. Verborgh)

From the clinician perspective, centrally stored PGS data is often stored in data silos where patient data is not discoverable. Further, patient consent to data usage and legal considerations concerning consent introduce added complexity. For a clinician today, assessing links between various types of data for a single patient or similar data between multiple patients, is nearly impossible due to current data privacy and storage efficiency strategies in centralized databases.

While the problem is complex, a largely unexplored storage strategy, that addresses many challenges posed, is using personal data vaults implemented using the Solid protocol [4].

2. Implementation Considerations

(i) Solid protocol for storing PGS data in decentralized data pods. A Solid data pod is a standardized personal, permissioned web server that allows the user to dictate with whom which data stored on the server is shared and accessible. Solid protocol provides a standardized read/write interface as well as a framework for obtaining consent for data access. We aim to populate (simulated) patient Solid pods with PGS data, which can then be accessed by an authorized application for use in a medical context.

(ii) A standardized PGS data model. The overarching goal of such a PGS data storage system involves embedding the PGS storage and accessing system within the Belgian/European Data Sharing Landscape. To achieve this scalability, an ontology and data partitioning strategy will be established for RDF conversion that follows existing standardization initiatives such as FHIR, GA4GH (beacons), or others, thereby, improving the shareability of such data with and between clinicians and researchers alike. We aim to link PGS data to other patient data stored within the same pod, publicly available external databases containing pertinent data, and/or other patient data within separate Solid pods, discoverable to those with authorization.

This project endeavors to connect PGS data storage to personalized healthcare in a scalable way that fosters improved patient control and practitioner access to PGS data.

Acknowledgments

Funding provided from VITO NV (`UG_PhD_2303_contract`). Ghent University acknowledges funding from the Research Foundation – Flanders (FWO).

References

[1] Y. K. Ko, J.-A. Gim, New drug development and clinical trial design by applying genomic information management, Pharmaceutics (2022).

[2] S. Mathur, J. Sutton, Personalized medicine could transform healthcare, Biomedical Reports (2017).

[3] T. Tanjo, Y. Kawai, K. Tokunaga, O. Ogasawara, M. Nagasaki, Practical guide for managing large-scale human genome data in research, J Hum Genet (2021).

[4] R. Verborgh, Re-decentralizing the Web, for good this time, in: Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web, ACM, 2023, pp. 215–230.