Analyze Decentralized Personal Health Data using Solid, Digital Consent, and Federated Learning

Chang Sun[1,2], Johan van Soest[3,4] and Michel Dumontier[1,2,3]

[1]*Institute of Data Science, Faculty of Science and Engineering, Maastricht University, Maastricht, The Netherlands*

[2]*Department of Advanced Computing Sciences, Faculty of Science and Engineering, Maastricht University*

[3]*Brightlands Institute of Smart Society, Faculty of Science and Engineering, Maastricht University*

[4]*Department of Radiation Oncology (Maastro), GROW School for Oncology and Reproduction, Maastricht University Medical Centre, Maastricht, The Netherlands*

Abstract

TIDAL is a Solid (SOcial LInked Data)-based, citizen-centric data platform that facilitates interactions by citizens and researchers for health research. In this demonstration, we will show how TIDAL 1) can store personal data in Solid pods as RDF with well-known health-related vocabularies (e.g., SNOMED CT), 2) controls access to query fine-grained subsets of personal data, 3) enables researchers to post a human- and machine-readable digital consent using Data Privacy Vocabulary, and 4) uses federated learning to analyze personal health data from multiple individuals using a Personal Health Train framework. TIDAL offers a new data paradigm of sharing and using personal data for research and ultimately increase the availability of personal data for societal relevant uses.

Introduction

Personal health data is largely collected and managed by hospitals, clinics, insurers and other healthcare providers. The European Commission has reported that many EU citizens have limited access to and control over their health data [1]. Although patients highly value the data rights granted by legislation such as the General Data Protection Regulations, patients experience difficulties to exercise their rights due to the lack of technology and standards [2].

Towards addressing these challenges, several initiatives are geared towards empowering individuals to manage their own data. Initiatives such as DEecentralised Citizen Owned Data Ecosystems [3] and MyHealthMyData [4], MIDATA [5] in Switzerland, and MedMij [6] in Netherlands have been pursued to store, access, and share personal data. However, existing solutions largely focus on one particular aspect such as data storage, access control, or data exchange among healthcare providers, rather than the entire data life cycle. Consequently, the quality and interoperability of the data is typically not well addressed and limits its re-usability. Lack of transparency and auditability of the data process (not knowing who is using/has used their data) in the existing tools raises concerns from individuals regarding trust of the tools.

Citizen-centric data platform (TIDAL) Demonstration

We will demonstrate TIDAL, a citizen-centric data platform for trustworthy health research. TIDAL allows researchers to post a request for participation, with clear indications of the

purpose of the research, the study design, and methods, as well as the specific data elements requested. The citizen will see the set of research requests, be able to review the request, examine the requested data elements, and ultimately decide whether to participate by formalizing their consent (which can be withdrawn at any time). TIDAL is implemented as a Solid application that stores, queries, and manages RDF data using health-related standards and vocabularies. Researchers create human- and machine-readable digital consent using the Data Privacy Vocabulary. Researchers specify the desired data elements through autocomplete support using the API from the Bioportal ontology repository. Selected data elements specified in the consent are transformed directly as the parameters for the data analysis (and hence no other elements may be requested in the query). TIDAL makes use of the Personal Health Train [7] to execute federated learning using data of consenting individuals.
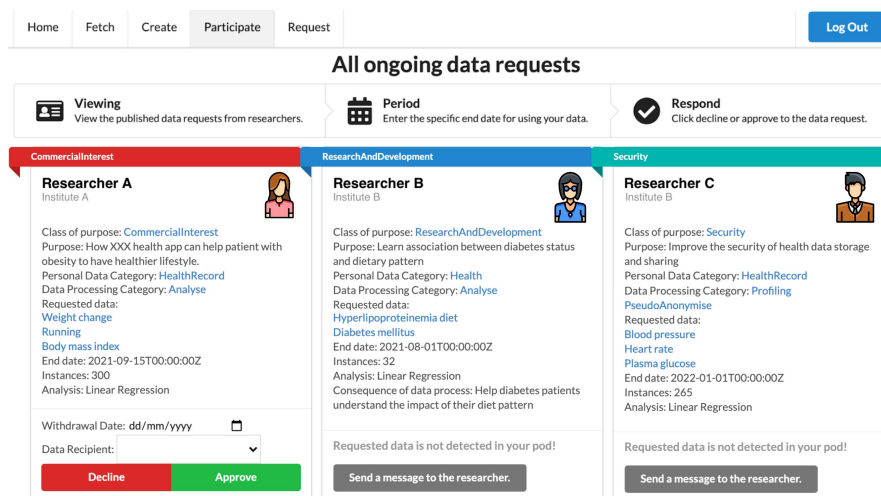

Figure 1: A screenshot of TIDAL displaying ongoing research requests.

References

[1] European Commission, White paper: on enabling the digital transformation of health and care in the Digital Single Market, Technical Report COM(2018) 233 final, 2018.

[2] European Parliament, Understanding EU data protection policy, Technical Report, 2020.

[3] T. Symons, T. Bass, Me, my data and i: The future of the personal data economy, 2017.

[4] M. Koscina, D. Manset, C. Negri, O. Perez, Enabling trust in healthcare data exchange with a federated blockchain-based architecture, in: International Conference on Web Intelligence, 2019, p. 231–237. doi:10.1145/3358695.3360897.

[5] MIDATA Cooperative, Midata: My data - our health, 2021. URL: https://www.midata.coop/en/home/, accessed on 03-12-2022.

[6] MedMij, Medmij: Personal health data in the palm of your hand, 2021. URL: https://www.medmij.nl/en, accessed on 03-12-2022.

[7] C. Sun, L. Ippel, J. Van Soest, B. Wouters, A. Malic, O. Adekunle, B. van den Berg, O. Mussmann, A. Koster, C. van der Kallen, et al., A privacy-preserving infrastructure for analyzing personal health data in a vertically partitioned scenario., in: MedInfo, 2019, pp. 373–377.