Life Sciences with Industry - 2021

VITO Health

Personal Genome Pods

The sequencing of the first human genome in 2001 initiated a revolution. Since then, more than a million human genomes have been sequenced at ever-decreasing costs. Truly personal genomics is now becoming a reality, thanks to the steadily increasing availability of affordable sequencing technology. Initially used for disease-specific research, genomics has now evolved into the area of personal health and wellbeing in the research community and the general population.

Despite this remarkable progress and the promise held in genomics for transforming biomedical research, precision medicine, and lifestyle decisions, the use of the tremendous amount of available data has been hampered. The main reason for this is the concern of potential unethical handling of the data. This is exasperated by the current practice of sharing genomic data i) in an 'all or none' fashion where either all genome data from an individual is shared or none and ii) often in an irreversible manner (once the permission is granted it is difficult to undo). Hence, the build-up of trust and technologies for sharing highly sensitive, identifiable personal genome information is critical – individually, nation-wide, and internationally.

Another problem inherent to genome data is that even if all individual genome data were freely and openly available, it would be computationally impossible to download, process and analyze them. A federated model of data access offers a solution where data storage and processing are independent and performed by separate entities. On the one hand, this requires standards & protocols for interoperability between these entities. On the other hand, this depends on a well-thought-out data storage system.

VITO is committed to building a Personal Data Platform (PDP) called WeAre (https://we-are-health.be/en) for health applications powered by SOLID technology, where dynamic consent is at the heart of the system. The PDP provides a secure environment/data vault for individuals to store, manage, and own their health data in a structured way. It further allows for dynamic and also partial sharing of resources.

By developing Personal Genome Pods in the WeAre platform, we propose a novel approach to **enhance sustainable personal genome data sharing by utilizing existing data formats and resources**, paving the way for secure, reliable personal genome analysis for citizens across European countries and **providing an ecosystem for academic and industry partners to store, manage, re-use and analyze genome data in a responsible manner**.

The SOLID technology implemented in PDP is pioneered by Prof. Tim Berners-Lee, inventor of the World Wide Web, as the extension of FAIR ideology to personal data. The goal of SOLID is to decentralize the web and give people ownership of their data while maintaining backward compatibility with the world wide web in its current form. The core idea behind this approach is to create personal PODs (personal online data store) where the data of individual users are stored in the form of linked data, instead of being housed in third party data silos. PODs are completely under the control of the individual, who determines who can access their data, when, and for what purpose.

The creation of Personal Genome Pods comes with various challenges.
Technological Challenges:

- Building a FAIR Personal Genome PODs hosting system from ground up: All the data in the system should be findable, accessible, interoperable and reusable.
    - Which genome data types to be supported taking into account their information content, size, as well as the availability of the tools and the processing time for conversion ?
    - What should be the data granularity for genome data sharing ? Sharing data per chromosome could be a start but is there a better way ?
    - What would be the *estimated* resources required to host one genome with chosen specifications ?
- Can we technologically ensure that the data shared is no longer in possession of the third party after the sharing token is expired ?
- What are the current state of the art Security mechanisms to safeguard highly sensitive information (such as genome data) and what is the cost of implementing those for each personal genome ?

Academic & Industrial Challenges:

- How can we ensure that once created, Personal Genome PODs can be safely used for Academic and Industrial purposes (Research Cohorts, Clinical Trials etc.) as well without needing of transfer of data ?
    - Certification of academic & industrial partners following the protocol is an option but not easily scalable.

Ethical & Societal Challenges:

- What are the potential ethical and societal challenges that comes up with sharing of Personal Genomes ?
    - Can/should an individual share his/her genome data without the consent of family relatives ?

References:

1. Byrd, J.B., Greene, A.C., Prasad, D.V. et al. Responsible, practical genomic data sharing that accelerates research. Nat Rev Genet 21, 615–629 (2020). https://doi.org/10.1038/s41576-020-0257-5
2. Corpas, M., Kovalevskaya, N. V., McMurray, A., & Nielsen, F. (2018). A FAIR guide for data providers to maximise sharing of human genomic data. PLoS computational biology, 14(3), e1005873. https://doi.org/10.1371/journal.pcbi.1005873
3. Boonen, K.; Hens, K.; Menschaert, G.; Baggerman, G.; Valkenborg, D.; Ertaylan, G. Beyond Genes: Re-Identifiability of Proteomic Data and Its Implications for Personalized Medicine. Genes 2019, 10, 682. https://doi.org/10.3390/genes10090682
4. Sequencing.com service for DNA Sequencing and DNA App store: https://sequencing.com/
5. Nebula Genomics service for DNA Sequencing and analysis platform: https://nebula.org/whole-genome-sequencing-dna-test/
6. The Roadmap of the 1+Million Genomes Initiative : https://digital-strategy.ec.europa.eu/en/library/roadmap-1million-genomes-initiative-now-clearly-illustrated-new-brochure

Contact Person:

Dr. Gokhan Ertaylan, gokhan.ertaylan@vito.be