pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210584

Reinforcing Health Data Sharing Through Data Democratization

Yuhang WANG[1] and Bian YANG
Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjøvik, Norway

Abstract. In this paper, we propose a health data sharing infrastructure which aims to empower a democratic health data sharing ecosystem. Our project, named Health Democratization (HD), aims to enable seamless data mobility of health data across trust boundaries, through addressing structural and functional challenges of its underlying infrastructure with a throughout core concept of data democratization. A programmatic design of HD platform was elaborated, followed by an introduction about one of our exploratory designs —an "reverse onus" mechanism that aims to incentivize creditable data accessing behaviors. This scheme shows a promising prospect of enabling a democratic health data sharing platform.

Keywords. eHealth, data democratization, health data infrastructure, privacy enhancing

1. Introduction

Sharing health data creates value for clinical care, trials, and case studies, as well as improved knowledge base[1][2] for healthcare researchers and healthcare organizations. Health data has also immense commercial value [3] for other parties such as pharmaceutical industry, data analytics providers, insurers, data markets, business intelligence.

The huge value associated with health data can lead to data misuse, for example, targeted use of ransomware, participation in black market[4], and other cybercrimes. The conventional health data infrastructure was not designed for anticipating value-driven data mobility and the associated cyber threats. There is a structural deficiency in the conventional infrastructure on which patch-like remedies only add to the complexity of the challenge.

Related works such like the national eHealth infrastructure (e.g., Norsk Helsenett) [5] in Norway has been built since middle 1990s which emphasized localized data retention and confidentiality. The "one citizen – one journal" plan was proposed in 2012 with the laws regarding medical records and health registers updated in 2015 in order to facilitate data mobility. The national pilots Helseplattformen and Helseanalyseplattformen [6] were launched in recent years to technically implement the connectivity and coordination in data sharing. On the EU level, the effort has so far

[1] Corresponding Author: Yuhang Wang, Norwegian University of Science and Technology, Gjøvik, Norway; E-mail: yuhang.wang@ntnu.no

mainly focused on the technical (e.g., the epSOS project) and legal [7] interoperability towards the EU eHealth strategy 2020 [8].

Rather than the above-mentioned works which reinforcing the infrastructure from a traditional view of vulnerability identification, protection, detection, and response, our work aims to define, architect, implement, and evaluate a democratic health data infrastructure which is expected to incentivize all parties, including individuals, to prove, negotiate, and configure their rights associated with health data. The conflicts of interest among different parties can be reconciled through a set of automated mechanisms so that data can be mobilized across trust boundaries.

We dedicated to architecting and constructing a data transaction model by strikingly practice the concept of **Data Democratization** (or say, democratic data sharing). Formally, this indicates two kernel idea, which will be followed throughout the design of our HD platform: 1) All stakeholders are treated identically without discrimination, and 2) when facing the inequivalence reality among each party, to promote fairness as a complementary.

State-of-the-art research and ethical & legal efforts have pay extensive attention on the first, however, we argue that the fairness promotion is also critical with regard to the data democratization, due to the extremely unequal actuality exists between the individuals and the colossus entity. We gazed deep into the platform in a hierarchical perspective and proposed our **Conceptual Layered Architecture** that is promised to achieve our goal of data democratization.

The following context are organized as listed: We present the terms of the stakeholders defined in this paper and the formal conceptual architectures in Section 2. Section 3 will illustrate how the concept of reverse onus could be applied into the democratic-promoting designs. We then conclude our work in Section 4.

2. Conceptual Layered Architecture of HD Platform

2.1. Stakeholders Description

The prior task for our work is to distinct discrepant stakeholders with significant behavior characteristics and interest relationship. We first classify our HD platform-relevant stakeholders into 7 types, then we present a sample of matching between these types of roles and the roles defined in GDPR. HD platform will “circulate” among diverse stakeholders, e.g., some roles are tent to get the health data for their point of interest, while some others have the right of disposal of the health data. Some stakeholders may also tend to provider the data processing/storage/analyzing fundamentality. We classify these stakeholders into 7 different types as shown in Table 1.

2.2. HD Architecture

To fulfill the principle of data democratization and the promised capabilities, we gazed deep into the platform in a hierarchical perspective. The HD platform is responsible for developing and managing the democratic negotiation procedures during the healthcare data business, for use in and exchange of clinical and individual healthcare information between the potential DS/DM and the potential data consumer.

Table 1. Stakeholders in HD platform

Stakeholder	Description
computing resource manager (CRM)	Supporter participant which assists each player in managing computing, storage, and communication resources in facilitating data sharing with other players.
data consumer (DC)	Player participant which can access data directly, query a database, or receive data from DS, DG, or DSP in order to exploit the value of the shared data.
data generator (DG)	Player participant which directly generates data from a data subject or converts sensed signal to formatted data.
data manager (DM)	Supporter participant which assists each player in processing, managing, and exchanging the data with other players.
dataset provider (DSP)	Player participant which creates and maintains, under the consent given by DS and possibly the agreement with DG, one or several structured datasets sourced from DS or / and DG.
data rights manager (DRM)	Supporter participant which assists each player in managing their rights with other players, i.e., proving, negotiating, and recording the terms and conditions describing the rights and obligations about the data for sharing.
data analysis service provider (DASP)	Participant which provides data analysis as a service to DS, DG, DC, or DSP.

For each principle and the potential promised scenario, the executive process could be considered as a correlation between the data sharing participants and an affair-relevant data sharing function in different executive level.

Guiding by the eHealth standardization in the Nordics countries[9] with respect to the interoperability [10][11], the data sharing function ranges from the incipient data provenance to the rights and obligation tracking after the agreement. We stratify our platform into four conceptual layers, named “Computing Infrastructure Layer”, “Data Sharing Operation Layer”, “Data Sharing Logic” and “Healthcare Business Layer”. Our Architecture also obtains references from the peer work on a diverse eHealth networking and healthcare data sharing solutions [12][13]. Figure 1 expounds our conceptual layered Architecture in detail, it describes the layers of the data sharing hierarchical structure, the data sharing participants, and the data sharing function.

The main systematic-level functions required in our platform are listed in Table 2.

Table 2. Main systematic-level functions

Function	Description
Data provenance	To provide a backward traceability of medical device, personal device in homecare environment, etc., and the health data sourced from these devices need to be audited in a trusted way of their rights and operation status.
Risk Assessment	Enabling each data subject has different risk acceptance tolerance and incentive degrees when they are entitled to rights and benefits from data.
Computational negotiation	Negotiating agents can operate and negotiate decisions. The requirements will be developed in compliance with the GDPR, healthcare regulations, and other relevant regulations. When processing and exchanging personal data between the agents, the design of the infrastructure will address such key requirements of the GDPR as data protection by design and by default, accountability, pseudonymization, right of access and right to erasure.
Multi-lateral security policing	Enabling individuals be able to share and control access to health data without having to place extensive trust in entities, and institutions must also be able to share data responsibly for research, innovation, and quality across institutional boundaries.

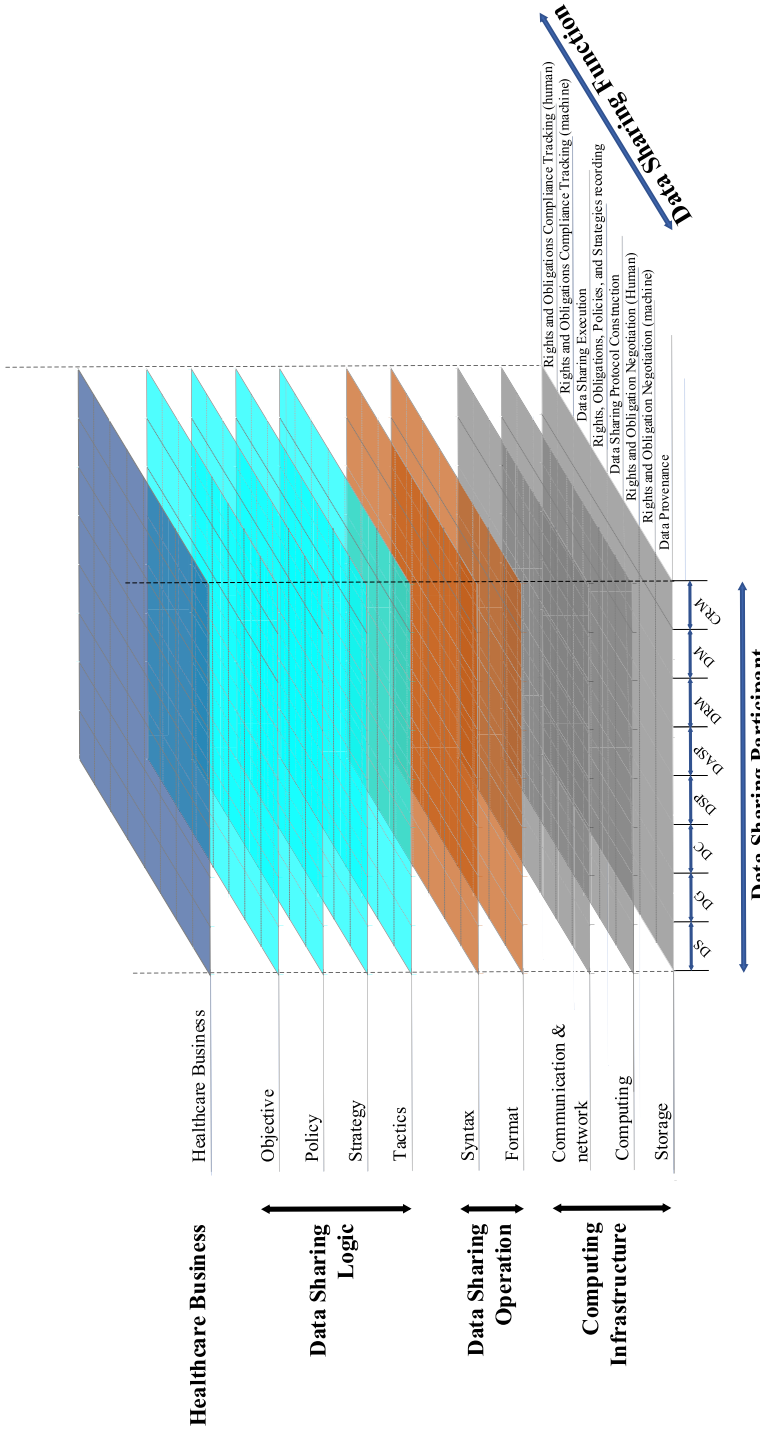

Figure 1. Conceptual Layered Architecture of HD platform.

3. "Reverse Onus" in Health data Negotiation

In most of the current ecosystem of digital market, an incommensurable inequality exits between the individuals and the so called "digital oligarchy" [14]. The fast growing of these consolidation power tries to gain monopoly in various of aspects, including the power of interpretation of the privacy data usage, and the health data sharing cannot be righteous alone.

In our HD platform, the DC could be played by such kind of roles. For example, DC is an influential giant company who wishes to constitute its global health big data warehouse, while the DM is just a small agent of DSs. In this design, we assume an inequitable situation between such two kinds of stakeholders and seek to resolve a potential unfair issue of knowledge asymmetry.

During the negotiation process in our platform, dominant DC may have much more right to argue 1) the (social or monetary)value of the health data, and 2) the scale and granularity of health data are demanded to perform a certain healthcare service. On the contrary, the DM may lack of knowledge to assess the opponent's proposal.

Our HD platform utilizes the concept which similar to "Reverse Onus" to mitigate this problem. Whenever a health data relevant negotiation happened between to stakeholders, say, a dominant DC and a regular DM, with great disparity, the platform shifts the burden of proof onto the DC specified to prove the necessity of the health data claim. When DM raises a discontent against the proposal with regard to data minimization (an essential privacy enhancing principle defined in GDPR), data value, ethical issue, etc., DC is in the position to provide convincible specification on his proposal.

Data Usage Approval:

The negotiation procedure is protected by requiring the DC to submit an application form (*appFm*) on the usage of the health data. Including: 1) usage purpose. 2) data precision upper limit in percentage. 3) data requesting schedule instant/time period/data manager triggered/etc. 4) requiring pattern in frequency distribution. 5) if necessary, reasonableness report.

The *appFm* will be assessed by the platform, based on the history usage log, will consider: 1) purpose to precision. 2) purpose to schedule. 3) purpose to pattern. 4) history comparison across entities. In this paper, we only assess the privacy-leakage risk and register the *appFm* in the following credit system proposed in the next subsection.

Credit Mechanism for Promoting Reverse Onus:

After reaching the mutual-agreement and the contract was built, the **credit mechanism** inherited from our previous work [15] will monitor the execution of the protocol to stimulate the DC to follow the terms. We applied a credit score mechanism upon the DC to encourage conformity and generate the virtual credit of DC based on his record, this credit will be further used to consult the future negotiations.

We set a credit score for each DC, denoted as α ($\alpha \in [0, 1]$) and with 1 means DC is with the highest credit score. In view of the HD platform, one observation of DC could lead to a downgrade of its credit score, which is an excessive access to the DM's health data, this is possible happening when DC misuses his interpretation clauses and collects health data exceeds the defined amount, granularity, etc. The DC with lower credit score will face a more arduous negotiating process than usual, and hence loss the potential health data application value. Guaranteed by this credit mechanism, a rational would DC tends to behave responsible and honest to the reverse onus scheme, and therefore the fairness of HD platform will be strengthened.

4. Conclusion

In this paper, we raised the concept of data democratization which will reinforce the health data sharing with respect to privacy enhancement and benefit insurance. An overall conceptual layered architecture was proposed which aims to enable such vision.

We further introduced an advanced concept of data democratization, which emphasized the fairness promotion in HD platform. A credit-mechanism-powered incentive scheme for promoting "reverse onus" on data usage was proposed. This mechanism rebalances the inequitable situation among all the stakeholders.

The future work will keep on implementing and integrating the proposed conceptual designs, several landing cases studies will be put into effort to improve the practicability of our work. Specially, the concept of reverse onus and the corresponding credit mechanism should be verified prudently by applying it onto the health data ecosystem.

References

[1] Olson S, Downey AS. Sharing clinical research data: Workshop summary. Washington, District of Columbia: The National Academies Press; 2013.
[2] Walport M, Brest P. Sharing research data to improve public health. The Lancet. 2011;377(9765):537-539.
[3] Czeschik C. Black Market Value of Patient Data. Digital Marketplaces Unleashed. 2017;:883-893.
[4] G. Hartvigsen. Lessons learned from 25 years with telemedicine in Northern Norway. Tromsø: NST-rapportserie; 2015.
[5] Helseanalyseplattformen Project: https://ehelse.no/helsedataprogrammet/helseanalyseplattformen
[6] Overview - Public Health - European Commission [Internet]. Public Health - European Commission. 2021 [cited 6 September 2021]. Available from: https://ec.europa.eu/health/cross_border_care/policy_en
[7] Topic 1: EU eHealth strategy towards 2020 – Public Health – European Commission [Internet]. Public Health - European Commission. 2016 [cited 4 November 2016]. Available from: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20161121_co01_en.pdf
[8] Herlof N, Christine B. Nasjonal e-helsestrategi og mål 2017 – 2022 [Internet], Direktoratet for E-helse, 2017. Available from: https://ehelse.no/publikasjoner/nasjonal-e-helsestrategi-og-mal-2017-2022-oppdatert-2019
[9] Vincent van Pelt, Michiel Sprenger. Adoption and take up of standards and profiles for eHealth Interoperability [Internet]. Antilope. 2015 [cited 2018 Nov 19]. Available from: https://www.antilope-project.eu/wp-content/uploads/2013/05/D1.1-Refinement_of_Antilope_Use_Cases_v1.2.pdf
[10] Kouroubali A, Katehakis DG. The new European interoperability framework as a facilitator of digital transformation for citizen empowerment. Journal of biomedical informatics. 2019 Jun 1;94:103166.
[11] Cole AM, Stephens KA, Keppel GA, Lin CP, Baldwin LM. Implementation of a health data-sharing infrastructure across diverse primary care organizations. The Journal of ambulatory care management. 2014 Apr;37(2):164.
[12] Mandl KD, Simons WW, Crawford WC, Abbett JM. Indivo: a personally controlled health record for health information exchange and communication. BMC medical informatics and decision making. 2007 Dec;7(1):1-0.
[13] Shen M, Zhu L, Xu K. Layered Data Sharing Architecture with Blockchain. InBlockchain: Empowering Secure Data Sharing 2020 (pp. 29-37). Springer, Singapore.
[14] Digital oligarchy. Knowledge for policy [Internet]. Knowledge4policy.ec.europa.eu. 2021 [cited 6 September 2021]. Available from: https://knowledge4policy.ec.europa.eu/foresight/topic/diversifying-inequalities/new-digital-oligarchy_en
[15] Wang Y, Tian Z, Sun Y, Du X, Guizani N. LocJury: an IBN-based location privacy preserving scheme for IoCV. IEEE Transactions on Intelligent Transportation Systems. 2020 Feb 10.