

Adryana Hutchinson

5.8.2022

Executive Summary

As digital data is increasingly used in our everyday lives, the preservation of citizen data rights becomes paramount. To this end, data cooperatives, which facilitate the voluntary, collaborative pooling of individuals' data, present a solution. Such initiatives empower participants, who become the negotiators of their own data. Furthermore, they serve to mend biases in data and the inequality of representation owing to the data divide. Data collection apps that prioritize accessibility, security, and fairness present a way forward for data cooperatives and the collectivist mindset they foster. In this policy proposal, I argue that the combination of smartphone applications and cloud computing technologies present a viable way for data cooperatives, especially those that relate to biomedical data, to prosper in both research and healthcare access.

Problem Statement

Digital data has been described as the “new oil” that is used in all segments of society [1]. Data that is especially useful in massive quantities is biomedical data [2]. However, access to this data for research remains an increasing legislative challenge for policy makers. The challenge is due to the fact that control over users' data is often concentrated in large companies, rather than with individuals themselves. This has led to many suggested policy proposals, one of which being the adoption of a data cooperative; data cooperatives refer to the voluntary, collaborative

pooling of an individuals' data, usually with the aid of a third party [1]. Despite this legislative push, however, data is still widely privately owned, and disincentives stand in the way of data sharing, such as scientific competition, unwillingness to grant access to existing datasets, and regulatory burdens associated with releasing data to third parties [2]. As it stands, vast quantities of data is under exclusive control of the private sector, and while open-source repositories of data are available, aggregating individual data at the scale needed to enable data mining techniques, especially in the biomedical sector, have been lackluster [2].

In this policy proposal, I will propose legislation that will allow for access to biomedical data in the form of data cooperatives, though this proposal could pertain to other areas of human data as well with small modification. More specifically, I aim to provide a model of data cooperatives similar to that of trade unions and pre-existing health data cooperatives -- international trade unions, such as the Worker Info Exchange in the United Kingdom, have already started recognizing the need for more control over data decisions pertaining to workers [3]. By analyzing the policies and the protocols they use, we will be able to effectively implement legislation in the United States relating to data cooperatives and rights ownership.

In the following section, I will analyze how data cooperatives are useful tools for research, as well as explain why their usefulness especially pertains to the biomedical sector. I will also be analyzing the policies that current health-related data cooperatives use presently.

Current Proposals

Data cooperatives broadly refer to the voluntary collaborative pooling of an individual's personal data for the benefit of the membership of a community [1]; they aim to dismantle the power imbalances between an organization that collects data about subjects and the subjects

themselves. In this policy proposal, I will be looking at ways for data cooperatives to be implemented in the context of biomedical data. There are two distinct arguments for the implementation of biomedical cooperatives; one that relates to the improvement of healthcare access throughout the United States, and one that addresses the inherent exploitation and bias in healthcare. I will discuss both in this section.

Public Health & Medical Innovation

Having readily-available data can enable researchers to effectively find and study medicinal effects, whether they be in pharmaceutical, diagnostic, or hospital settings. Freely available biomedical data has the potential of improving healthcare by preventing disease, increasing the accuracy of diagnoses, providing safer medication prescriptions, and making treatments more effective [4]. Additionally, by investing in this technology, healthcare costs would go down, as less priority would be placed on individual tests and data aggregation.

Citizen Empowerment

A motivation for individuals to pool their data is the need to share common insights across data that would be otherwise inaccessible [1]. Members of data cooperatives are able to become an asset to research, as individuals become the negotiators of their data; this allows members of a community to engage with data-driven health research that can empower them. This possibility can address the underrepresentation of minorities relating to health research, which not only amplifies minority voices, but also negates the consequences of using biased datasets [2].

There are biomedical data cooperatives that exist presently that aim to incorporate both of these points. Switzerland's MIDATA, for example, is a non-profit data cooperative founded in 2015 that is devoted to the management of health data. The cooperative fosters an open, data-driven ecosystem via a public infrastructure composed of mobile services and connections; users that opt to provide MIDATA with their personal data can directly contribute to the furtherment of clinical studies, lending power to the individual. Furthermore, the cooperative's commitment to transparency involves members directly in matters of governance, with each person being able to vote on how the cooperative functions. MIDATA's initiative is citizen-owned, and is structured in such a way that ensures users remain in control of their data while furthering an important cause [5]. Another pre-existing biomedical cooperative is Spain's Salus Coop. Salus Coop is a non-profit data cooperative focused on health research and data -- like MIDATA, it has a citizen-driven collaborative governance model, with each user being able to manage where their data goes. Salus Coop legitimizes citizens' rights to control their health records while facilitating data sharing to promote public research innovation in healthcare [6].

To conclude, biomedical data cooperatives are important because large quantities of data have the potential to accelerate the rate in which medicine is developed, researched, and tested. They operate by giving users control of their data, allowing the user to dictate where they want their data to go; this greater flexibility empowers users to determine which research projects they want to contribute to, and which ones they do not. This, along with decentralizing the governance structure of the data cooperative, gives users total control over how their data is managed, used, and seen. In the following section, I will propose potential legislative solutions as to how to adopt biomedical data cooperatives into current US law.

Proposed Solution & Contextualization

An integral aspect of the push against inequality of representation in data-driven technologies is the notion of community ownership. The cooperative ownership model seeks to safeguard data rights and place emphasis on social benefit via a paradigm shift in favor of the individual [6]. Crucially, granting control to the community fosters a collectivist mindset; in addition to preserving individual rights, it makes certain that data collection is carried out in the interest of the populace. This assumes the voluntary, autonomous nature of data cooperatives, as is necessary for citizens to maintain control over their data and how it is put to use [6]. Despite this, a collectivist model serves as a practical baseline for how data cooperatives should proceed with protecting the interests of their members. To promote the full-integration of users, a transparent interaction platform has to be established [8].

With the popularization of interpersonal technology in the 21st century, the ability for individuals to monitor an increasing number of health parameters via sensors and applications connected to smartphones has grown significantly. Currently, there are over 40,000 health-related apps available for smartphones; this ability to monitor health outcomes continuously without the need of a doctor offers a dynamic and effective way to gain feedback of personal health information [8]. This growth in health-related apps shows the viability of creating a generalized system, such as a cloud computing provider, that maps health records to one another. Such a system would allow for large-scale analytics on data generated by apps within the system, with machine learning technology being able to automatically validate health status and disease models extracted from data sets [8]. In order to keep privacy and explicit bias in check, data would be cleansed of personal information. Potential apps would include an easy-to-use user interface, the ability to annotate data, and sort data into user-defined spaces (i.e. fitness, checkup,

knee injury etc.) [8]. As mentioned, the user would have full control over their health data once it is collected, which could be shared with whomever (such as a public health organization, another data cooperative, etc.) the user sees fit; in this way, the system acts as a trusted broker on behalf of its members [8]. By having a user's information flow facilitated within an application, they are able to effectively control their data with the tap of a button.

Potential applications could be developed on both a federal and private level, and the user could choose whichever app they believe suits their specific needs. When private companies develop data-storing apps, however, they must follow specific guidelines that would be determined by research and technical experts, such as the Health Insurance Portability and Accountability Act (HIPAA), the US Federation of Worker Cooperatives, and the Institute of Electrical and Electronics Engineers (IEEE). One way to incentivize individuals to use these apps is by providing professional consensus at a quicker rate through telehealth services. Another way could be the possibility of providing community-based forums for specific illnesses, as well as rate comparisons between local providers. Overall app function does not need to be relegated entirely to data-holding services, it can also be a means to facilitate a wide variety of health-related tasks and conversation.

In order to have effective biomedical data cooperatives that can increase the throughput of integral research, it is paramount to develop safe, effective, and low-cost tools. One of the most effective ways to do this is through the development of health-related apps that can pool data collectively. Accessibility, security, and fairness are all concerns that these apps need to address to ensure the most effectiveness of its software. In the following section, I will evaluate the risks and concerns associated with biomedical data cooperatives.

Risks & Conclusion

One of the biggest concerns relating to data cooperatives is the security in a cloud computing environment. A critical characteristic of cloud computing storage is that the data is kept in one place; in this storage environment, the privacy and security of data can never be fully assured [7]. Unpredictable factors may pose a threat to the confidentiality of the data stored in the cloud, such as hackers, public management interface issues, and privilege abuse [7]. Because of this, health data repositories should be certified by independent government regulatory bodies and good governance structures, similar to how the FDA controls drugs produced by pharmaceutical companies [4].

Accessibility and discrimination based on socio-economical factors is another concern; some individuals, such as those who are unable to access health-related apps due to disability, not having access to a smartphone, etc., may not be able to reap the benefits of using digitally provided cooperatives. In order to make these applications available to a wide variety of audiences, apps should be required to have a default set of accessibility options, as well as accessibility on a large variety of smartphones -- both newer models and otherwise.

Another important issue relates to the preexisting precautions of biomedical information. HIPAA imposes tight legislation on electronic health information in the United States, and such restrictions may hinder the growth of biomedical data applications due to the high liability of the context. This particular problem would need intensive research that is outside the scope of this policy proposal; however, I suggest that biomedical data used in data cooperatives should be legislatively unique, such that any biomedical data in a data cooperative requires intensive “scrubbing” of any identifiable information before it can be viewed by anyone outside the owner of said information (the user). This could potentially be done by trained machine learning

algorithms that can be used to identify such information -- the particular mechanisms for this scrubbing would likely be regulated by federal bodies, security specialists, and research experts.

This policy proposal aims to offer guidance in the implementation of fair, accessible, and secure data solutions in regards to biomedical data. The collection and sorting of digital health data offers great social benefit; however, we must ensure that the mechanisms that collect said data are fairly used and fully controlled by the user base. This can be done by a variety of ways, some of which include transparency, open-governance, and data protection measures. In order to meet the demand of public health outreach, a human-centered approach is required.

Sources

- [1] Hardjono, Thomas, and Alex Pentland. "Data Cooperatives: Towards a Foundation for Decentralized Personal Data Management." *ArXiv.org*, 21 May 2019, <https://arxiv.org/abs/1905.08819>.
- [2] Blasimme, Alessandro, et al. "Democratizing Health Research Through Data Cooperatives." *Philosophy & Technology*, vol. 31, no. 3, 2018, pp. 473–479., <https://doi.org/10.1007/s13347-018-0320-8>.
- [3] Kapoor, Astha. "Collective Bargaining on Digital Platforms and Data Stewardship." *Friedrich-Ebert-Stiftung*, Friedrich-Ebert-Stiftung E.V., Feb. 2021, <https://library.fes.de/pdf-files/bueros/singapur/17381.pdf>.
- [4] Van Roessel, Ilse, et al. "Potentials and Challenges of the Health Data Cooperative Model." *Public Health Genomics*, vol. 20, no. 6, 22 June 2018, pp. 321–331., <https://doi.org/10.1159/000489994>.
- [5] "Cooperative." *MIDATA*, 5 May 2021, <https://www.midata.coop/en/cooperative/>.

[6] Scholz, Trebor, and Igor Calzada. “Data Cooperatives for Pandemic Times.” *Public Seminar*,

The Editorial Board of Public Seminar, 19 Apr. 2021,

<https://publicseminar.org/essays/data-cooperatives-for-pandemic-times/>.

[7] Kossmann, D., et al. “Health Data Cooperatives – Citizen Empowerment.” *Methods of*

Information in Medicine, vol. 53, no. 02, 11 Feb. 2014, pp. 82–86.,

<https://doi.org/10.3414/me13-02-0051>.