

# Chapter 9

## Personal Data Cooperatives – A New Data Governance Framework for Data Donations and Precision Health



Ernst Hafen

**Abstract** Personalized health research depends on aggregated sets of personal data from millions of people. Given that personal data can be copied, individuals are entitled to copies of their data and individuals are the ultimate aggregators of all their personal data, citizens are elevated to new roles at the center of health research and a novel personal data economy. There, citizens, not some multinational company, control the use of and benefit from the intellectual and economic value of these data. Here, I show that democratically controlled nonprofit personal data cooperatives provide a governance and trust framework for data sharing and data donation. They also provide a means of attaining improved precision health and a digital society in which socio-economic asymmetries can be balanced.

**Keywords** Personal data · Electronic health records · EHR · Data portability · Data protection regulation · Empowerment · Cooperatives · Common good · Public health · Precision medicine · Artificial intelligence

### 9.1 The Unique Features of Personal Data

Medicine is transitioning from a heuristic science that is relying on relatively small numbers of cases to an increasingly data driven science involving large numbers of cases. Much of the data relevant to health research and efficient healthcare are personal data. These include data stored in medical records, registry databases and increasingly personal data recorded by patients/citizens themselves with the sensors in their smartphones. The challenges in using these data for research and better healthcare span from issues such as data silos, interoperability, data accessibility and obtaining the consent of individuals to access and use their data. Before

---

E. Hafen (✉)

Institute of Molecular Systems Biology, ETH Zürich, Zürich, Switzerland

MIDATA Cooperative, Zürich, Switzerland

e-mail: [ehafen@ethz.ch](mailto:ehafen@ethz.ch)

addressing these issues and offering at least partial solutions, I would like to highlight three special features of personal data: (1) Personal data, like other data, once generated, can easily be copied at near zero marginal costs. The copies can be used for different purposes. Data are a non-rivalrous good. Therefore, it is difficult to talk about data ownership. The doctor is bound by law to keep medical records for 15 years. Patients in most countries, however, have the right to obtain a copy of their medical records. (2) Personal data, or copies thereof, are a new asset class (World Economic Forum 2011). They comprise one of the few assets that are fairly equally distributed among people. All humans are billionaires in genome data since their genomes contain a unique set of six billion base pairs of DNA. Likewise, people have similar numbers of heart beats, steps and they consume a similar number of meals and liters of water. It is clear that access to resources like food and data is currently not equally distributed around the world. However, the rapid spread of smartphones and broadband internet in Low and Middle Income Countries (LMICs) has the potential to change this in the near future. (3) Individuals are the ultimate aggregators of their data. Only they will be in a position to aggregate data from their medical records, their genome data, their shopping data and their smartphone data. This is of course under the provision that they can obtain a copy of all these data. A right that, as we will see, has been strengthened greatly by the European General Data Protection Regulation (GDPR) (De Hert et al. 2017). The fact that data can be copied, data are equally distributed amongst individuals and that each individual is the ultimate aggregator of the data, forms a basis for a new role for patients and healthy individuals at the center of health research, prevention and care.

## 9.2 The Need – Aggregated Datasets on Millions of People

The human genome can be compared with the operating system of a smartphone. A smartphone app interacts with the operating system and produces a certain effect (e.g. recording and visualizing one's daily steps). A drug interferes with the human operating system to elicit an effect (e.g. relief of pain). We may ask why it takes two weeks to develop a smartphone app at a cost of \$ 20,000, whereas it takes 15 years and \$ 2bn to develop a drug. Moreover, we can expect the app to work on all smartphones using the same operating system, while drugs only work less than 50 percent of the time (Chhibber et al. 2014). Of course, the reasons for this difference are manifold. Above all is the fact that the smartphone operating systems were engineered and are thus completely understood, whereas the human operating system is a product of evolution (i.e. random mutations and natural selection evolution). Another difference, however, is that the copies of the operating system on smartphones are identical, whereas the human genomes of individuals differ in 1/1000 base pairs. The effectiveness of drugs in pairs of identical twins is significantly higher than in fraternal twins or unrelated people (Chhibber et al. 2014). To improve the odds of medicine we need to learn how differences in genomes in combination with differences in environmental and behavioral factors affect health, disease and the

treatment of disease. In other words, we need to make correlations between genotypes (genome) and phenotypes (health status) for millions of people. This is only possible with the active participation of the individuals as the rightful data aggregators and data access controllers themselves.

## **9.3 The Opportunity – The Legal Right of Citizens to Obtain Copies of Their Personal Data and Their Willingness to Contribute These Data to Research**

### ***9.3.1 The European General Data Protection Regulation (GDPR) – Data Portability***

The new European Data Protection Regulation (GDPR) specifies in Article 20 the right of citizens to data portability permitting them to obtain machine readable copies of all their personal data (De Hert et al. 2017). This includes data on social networks, shopping data, medical, education data and other data. Initially planned as a way to create competition among data collectors in much the same way mobile phone providers guarantee that one's smartphone number is portable when changing providers, it results in a true empowerment of citizens in a data-driven environment. While individuals have little secondary use for their mobile phone numbers, they can aggregate and use copies of their medical, shopping, genome and fitness data in a variety of different ways. For example, they can actively participate in research projects by making their aggregated data sets accessible or they can profit from personalized data analysis services via apps provided by companies.

### ***9.3.2 The Willingness and the Right to Citizen Science***

There is a great willingness of people to actively participate and contribute to scientific research. This is evident from the millions of people who contribute their time, knowledge and data to citizen science platforms such as [Zooniverse.org](https://www.zooniverse.org). On this platform, people help to annotate galaxies or wild-life in webcam images from the Serengeti, transcribe weather reports from old ship log books for oceanic climate information and annotate histological sections for the existence of cancer cells. Increasingly, people also contribute to science not only as data scientists as in the above examples but also by directly collecting and contributing their own data to scientific projects. For example, the large majority of patients consent to using their medical data for medical research. At the university hospital in Lausanne a recent study showed that over 80 percent of the cancer patients provide a general consent for using their data including genome data for research (Mooser and Currat 2014). The willingness to contribute personal data to research is by no means limited to

patients. Surveys with university students or elderly people who attend university programs for the elderly showed that roughly 60 percent would order a direct to consumer genetic test. The ability to support scientific research was mentioned by more people than to find out more about one's own genetic health risks (Vayena et al. 2014). Millions of citizens have even given their genome information to commercial companies such as [23andme.com](http://23andme.com) or [ancestry.com](http://ancestry.com), even though they had to agree that their data is used commercially for the benefit of the companies' shareholders. Giving citizens the right to aggregate and control sharing of their personal data empowers them to become active participants in science (Vayena and Tasioulas 2016). In doing so, health relevant data collected continuously with smartphone sensors can be combined with medical and other data. The active participation in research projects also provides an excellent means to improve the scientific literacy of citizens and patients.

## 9.4 The Challenge

Smartphones have been available for more than 10 years and medical records also have existed, at least in some countries, in digital forms for even longer. Why then have citizens and patients not taken a more active role in managing their personal data and contributing to research? The reasons lie at least in part in the dynamics of the nascent personal data economy. Internet and social media companies, as well as app providers, provide their services for free in exchange for personal data. Because of the ease of use we have gotten used to paying for digital services with our personal data. These data fuel a nascent and rapidly growing personal data economy that is largely controlled by large multinational companies and data brokers. The value lies in the aggregation of these different data types for the purpose of digital and personalized advertising (Lanier 2010). As the recent case of Cambridge Analytica and Facebook shows, these data are increasingly also used for more subtle psychological manipulation including influencing political voting behavior (Dehaye 2017). Moreover, the companies controlling the largest amounts of data will have the best resource to train their artificial intelligence algorithms, thereby further increasing the socio-economic asymmetry between individuals all over the world and a few multinational companies (Haynes and Nguyen 2013; Lee 2017).

In the case of medical records, these are often locked in incompatible data silos in hospitals and the private practices of physicians. Even in countries where electronic health records (EHR) have been established, patients can access the records and make them accessible to other healthcare providers only for the primary use of the data, i.e. healthcare. There are few options to actively decide on the secondary use of these personal data for medical research or other data services. In order to aggregate medical and other health-related data under the control of the individual, a new governance framework that provides trust and empowerment for citizens/patients to aggregate and actively manage the access to their data is needed.

## 9.5 The Solution – Data Cooperatives and Personal Data, a Perfect Match

We posit that citizen-owned nonprofit data cooperatives provide the basis for a democratically controlled and fair personal data ecosystem from which society at large will profit. Democratic governance and self-help are part of the DNA of cooperatives and this sets them apart from other organizational forms including foundation and shareholder-controlled companies. The match between cooperative governance and personal data control rests on the three unique features of personal data.

First, data can be copied and individuals possess the right to obtain a digital copy of all their personal data according to the data portability article of the EU GDPR. This offers the possibility for a new parallel personal data ecosystem under the control of the citizens without directly interfering with the current personal data economy. Second, the fact that all people have similar amounts of personal data aligns well with the democratic one-member-one-vote governance of cooperatives. Third, the fact that individuals are the ultimate aggregators of their personal data offers the opportunity for entirely new data services, artificial intelligence and research on different data types that hitherto have been locked in different silos or whose access without the consent of the data subject is protected.

Data cooperatives operate a secure IT platform on which individuals can store, manage and control access to all their data copies. Much like the case of financial bank accounts, individual data account holders are in complete control over the use of their data. In contrast to most of the current banking software where administrators have access to customer data, however, in the personal data platform each record is encrypted and only the account holder has the key. Therefore, neither the administrator of the IT platform nor the management of the cooperative has access to the data. Account holders can become members of the cooperative and in this way also participate in the democratic governance of the cooperative. Their duties include the election of the board of the cooperative and to vote on how proceeds are invested (Hafen et al. 2014).

### 9.5.1 *Data Cooperatives, Business Model, Non-profit, Financial Incentives*

Data cooperatives act as the fiduciary of their respective account holders' data. When individuals make part or all their data accessible for academic or pharmaceutical research or for data services from other companies, the management of the cooperative negotiates access to these data with the researchers, pharmaceutical or data companies. It ensures that data access is fair. For example, data accessed by a data company providing a mobile app must not be sold to third parties, as is the case currently with all the “free” apps in the app stores. For research projects and clinical trials the cooperative ensures that the results generated with these personal data will

be published irrespective of whether the results are positive or negative, and that copies of the data generated during the project will be returned to the individuals' accounts. It will also negotiate a fair price for data access by third parties. This can be a fee on mobile apps running on its platform, or for the recruitment of patients for clinical trials.

Will this cooperative model be profitable? The personal data economy, also called the digital identity economy, with the currently freely accessible data via social media, free apps and internet access is projected to reach a market volume of € 1 billion in 2020 in Europe (The Boston Consulting Group 2012). This figure does not include medical data, access to which is restricted by data protection laws. Combining these different data types under the control of the individuals will generate a vastly larger market value of which data cooperatives will obtain a significant share. By acting as the fiduciary of people's data and their decisions as to who and for what purpose their data can be accessed, data cooperatives negotiate the terms for data access with industrial partners, including pharmaceutical companies and data companies. The revenues generated from data access and also from the recruitment of consenting participants for clinical trials will be used to maintain and develop the platform, regular security checks and data services on the platform. Moreover, members of the cooperative will be able to vote on which research funds or data services revenues will be objects for investment. The nonprofit character of the data cooperative model specifies that these financial benefits will go back to society and not to the individuals who make their data accessible. There are two arguments for this statement, which at first glance may appear somewhat counterintuitive. First, although the aggregated data set of an individual is more valuable than the partial aggregates that companies and institutions possess of the same individual, the societal, intellectual and economic value lies not in the data set of a single individual but in those of all the participating individuals. We argue that this value should be returned to society at large. A distribution of dividends to members of the cooperative would be unfair, since many account holders may not want to become members and would therefore not profit from these dividends. Second, offering financial incentives for data sharing corrupts the motivation for sharing in much the same way blood donations work better without financial incentives (Sandel 2012).

### ***9.5.2 Challenges to the Cooperative Model***

Even though in general financial cooperatives fared better in the financial crisis than shareholder-controlled companies, there are plenty of reports about the failure of cooperatives (Birchall 2013). The establishment of personal data cooperatives faces two main challenges. First, establishing a truly participatory democratic governance with large numbers of members requires new tools such as liquid democracy, also called delegative or proxy democracy (Rutt 2018). Second, a major challenge for cooperatives is the initial financing. Even though data cooperatives will be hugely

profitable with millions of account holders actively sharing data, establishing the platform, the initial services that provide a benefit for the users and the legal framework for the cooperative and data governance are challenging. In contrast to shareholder-controlled companies, cooperatives cannot give equity to financial investors, since they are owned by their respective members. Thus, cooperatives require initial financial support from foundations, from philanthropists and through research grants. Crowd financing is an obvious option but that would also require some tangible short term benefits for it to succeed.

### ***9.5.3 Example: MIDATA Data Cooperative***

The MIDATA cooperative is a first example of a data cooperative that shows how data can be used for the common good, while at the same time ensuring the citizens' sovereignty over their personal data. Founded in 2015, the non-profit cooperative operates a data platform, acts as a trustee for data collection and guarantees the sovereignty of citizens over the use of their data. The citizens actively contribute to research as users of the platform by providing access to data sets and as cooperative members to control and develop the cooperative. Currently, the Swiss MIDATA cooperative accepts only residents of Switzerland. With partners in Germany, Belgium, the Netherlands and the United Kingdom, we are setting up MIDATA cooperatives in these countries.

The articles of association of the cooperative define its nature as a non-profit organization and enshrine the sovereignty of users over their data and its use (including use in anonymized form). An internal cooperative data ethics review board, whose members are elected by the general assembly, controls the ethical quality of the services and related projects. Members and non-members of the cooperative can open an account and possess the same rights on the platform. Members participate also in the governance of the platform.

### ***9.5.4 The Data Platform and Governance Form the Core of a New Innovation Ecosystem***

The data platform used by MIDATA is being developed by ETH Zürich and the Bern University of Applied Sciences. It allows citizens to collect their health data and to freely decide on the use of that data in research projects. They can thus play an active role in medical research as “citizen scientists”.

The platform model allows the separation of the IT platform (data storage, access and consent management) from the data applications (mobile applications) and thus enables an open innovation ecosystem. Users will have access to various data

services and can decide to participate in research projects. Start-ups, IT service providers and research groups can offer mobile apps on the platform, for example, health apps or apps for the management of chronic diseases.

The IT platform is operational and is currently being used in several data science projects. In one project, patients after a gastric bypass operation record their condition, fitness and weight at home and share the data with the attending physician at the Bern University Hospital. In another project at the Zürich University Hospital, patients suffering from multiple sclerosis are testing the effect of treatments using a tablet app that assesses their cognitive and motor status. In the allyscience.ch project, people suffering from hay fever record their symptoms and contribute to a Swiss-wide allergy map in relation to pollen data provided by Meteo Swiss. More than 8,000 people have downloaded the AllyScience App and actively contribute as citizen scientists to this project.

## 9.6 Data Cooperatives and Data Donations

In the Middle Ages, feudal lords argued that they could not pay their serfs money for their work since the serfs could not handle the responsibility and would immediately spend it on drink and women. Today, most people in high income countries have a bank account and decide how to invest or spend their money. This forms the basis of our current economy and the globally improved standards of living. Inheritance of financial assets is also clearly regulated. In 5–10 years, people will possess their own data bank account and decide to whom and for what purpose they will provide access to their data. As in the case of banks today, there will be different business models for such data banks. Some may offer citizens financial incentives and maximize the profits for shareholders, others will be non-profit cooperatives whose revenues will be invested for the common good. Personal data is much more personal than money. Trust will be the most essential factor in generating a citizen-controlled personal data economy. In such a trusted environment people will become active contributors and actors in the digital society. By learning the benefit of data sharing for public health and as well as their own health and wellbeing they will be prepared do donate their data post mortem for the common good.

Even though the digital revolution threatens to further increase the global digital dependence of individuals from data and AI service providing companies and augment the socio-economic asymmetries, it also offers a new avenue out of this dependency. With their unique personal data assets, individuals can help to democratize the personal data economy and contribute to John Rawls' vision of the most just form of society, a property-owning democracy in which people possess not only a political but also an economic vote (Rawls 2009). With their personal data, citizens all over the world now possess an equally distributed economic value that in combination with a cooperative governance as outlined above could contribute to such just societies globally.

## References

- Birchall, J. 2013. *Resilience in a Downturn: The Power of Financial Cooperatives*. Geneva: ILO.
- Chhibber, A., D.L. Kroetz, K.G. Tantisira, et al. 2014. Genomic architecture of pharmacological efficacy and adverse events. *Pharmacogenomics* 15 (16): 2025–2048. <https://doi.org/10.2217/pgs.14.144>.
- De Hert, P., V. Papakonstantinou, G. Malgieri, L. Beslay, and I. Sanchez. 2017. The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review: The International Journal of Technology Law and Practice* 34 (2): 1–11. <https://doi.org/10.1016/j.clsr.2017.10.003>.
- Dehaye, P.O. 2017. *Cambridge Analytica and Facebook data*. Medium. Retrieved May 23, 2017.
- Hafen, E., D. Kossmann, and A. Brand. 2014. Health data cooperatives – Citizen empowerment. *Methods of Information in Medicine* 53 (2): 82–86. <https://doi.org/10.3414/ME13-02-0051>.
- Haynes, P., and C.M.-H. Nguyen. 2013. Rebalancing socioeconomic asymmetry in a data-driven economy. In *The global information technology report*, ed. B. Bilbao-Osorio, S. Dutta, and B. Lanvin, 67–72. Geneva: World Economic Forum.
- Lanier, J. 2010. *You are not a gadget*. Vintage.
- Lee, K. Opinion/The real threat of artificial intelligence. June 2017. <https://www.nytimes.com/2017/06/24/opinion/sunday/artificial-intelligence-economic-inequality.html>.
- Mooser, V., and C. Currat. 2014. The Lausanne Institutional Biobank: A new resource to catalyse research in personalised medicine and pharmaceutical sciences. *Swiss Medical Weekly* 144: w14033. <https://doi.org/10.4414/smw.2014.14033>.
- Rawls, J.A. 2009. *Theory of justice*. Cambridge, MA: Harvard University Press.
- Rutt, J. 2018. *An introduction into liquid democracy*. Medium. <https://medium.com/@memetic007/liquid-democracy-9cf7a4cb7f>. Accessed 25 Aug 2018.
- Sandel, M.J. 2012. *Sandel: What money can't buy: The moral limits of markets – Google scholar*. The Boston Consulting Group. 2012. *The value of our digital identity*, 1–65. <http://www.liberty-global.com/PDF/public-policy/The-Value-of-Our-Digital-Identity.pdf>.
- Vayena, E., and J. Tasioulas. 2016. The dynamics of big data and human rights: The case of scientific research. *Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences* 374 (2083): 20160129. <https://doi.org/10.1098/rsta.2016.0129>.
- Vayena, E., C. Ineichen, E. Stoupka, and E. Hafen. 2014. Playing a part in research? University students' attitudes to direct-to-consumer genomics. *Public Health Genomics* 17 (3): 158–168. <https://doi.org/10.1159/000360257>.
- World Economic Forum. 2011. *Personal data: The emergence of a new asset class*, 1–40. World Economic Forum.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

