

Health Data Cooperatives – Citizen Empowerment

E. Hafen¹; D. Kossmann²; A. Brand³

¹Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland;

²Institute for Information Systems, ETH Zurich, Zurich, Switzerland;

³Institute for Public Health Genomics, University of Maastricht, Maastricht, The Netherlands

Keywords

Health data, empowerment, ownership, cooperative, citizen centric, democracy

Summary

Introduction: This article is part of a Focus Theme of *Methods of Information in Medicine* on Health Record Banking.

Background: Healthcare is often ineffective and costs are steadily rising. This is in a large part due to the inaccessibility of medical and health data stored in multiple silos. Furthermore, in most cases molecular differences between individuals that result in different susceptibilities to drugs and diseases as well as targeted interventions cannot be taken into account. Technological advances in genome sequencing and the interaction of 'omics' data with environmental data on one hand and mobile health on the other, promise to generate the longitudinal health data that will form the basis for a more personalized, precision medicine.

Objectives: For this new medicine to become a reality, however, millions of personal health data sets have to be aggregated. The value of such aggregated personal data has been recognized as a new asset class and many commercial entities are competing for this new asset (e.g. Google, Facebook, 23andMe, PatientsLikeMe). The primary source and beneficiary of personal health data is the individual. As a collective, society should be the beneficiary of both the eco-

nomie and health value of these aggregated data and (health) information.

Methods: We posit that empowering citizens by providing them with a platform to safely store, manage and share their health-related data will be a necessary element in the transformation towards a more effective and efficient precision medicine. Such health data platforms should be organized as cooperatives that are solely owned and controlled by their members and not by shareholders. Members determine which data they want to share for example with doctors or to contribute to research for the benefit of their health and that of society. Members will also decide how the revenues generated by granting third parties access to the anonymized data that they agreed to share, should be invested in research, information or education.

Results: Currently no functional Health Data Cooperatives exist yet. The relative success of health data repositories such as 23andme and PatientsLikeMe indicates that citizens are willing to participate in research even if – and in contrast to the cooperative model – the commercial value of these data does not go back to the collective of users.

Conclusions: In the Health Data Cooperative model, the citizens with their data would take the center stage in the healthcare system and society would benefit from the health-related and financial benefits that aggregation of these data brings.

1. Ineffective Healthcare Systems

The Institute of Medicine of the National Academies of Science estimated that in 2009 \$ 765 billions (30% of the healthcare budget) has been spent in the US on ineffective and unnecessary treatments, and excessive administrative costs [1]. Much of these excessive costs are caused by ineffective and inefficient interventions as well as inaccessible medical information stored in various silos and incompatible data structures. Furthermore, data protection and privacy laws rightly prevent third parties (e.g. the research community) to access the fragmented data without the informed consent of the individual. Inaccessibility of data is not only a cause of inefficiency and increased costs in the healthcare system, it also prevents the citizens to control their data and thus engage more actively in their own health.

Effective treatment is also diminished by the one-size-fits-all medicine that is still largely the rule, in spite of the obvious differences between individuals in their response to drugs and their susceptibilities to diseases. For example, Clopidogrel (Plavix) is an anticoagulant that is taking by some 40 million people after heart or bypass surgery. 30 percent of people carry at least one copy of an inactive allele of CP2C19 that encodes the enzyme Cytochrome P450 required for the activation of Clopidogrel. Carriers of one or two copies of this allele have reduced or no active levels of Clopidogrel, thus increasing the risk of thrombosis [2]. Similarly, for eleven widely used drugs replicated genome wide association studies have revealed significant differences in efficacy and safety based on common genetic variants. However, these ob-

Correspondence to:

Ernst Hafen
Institute of Molecular Systems Biology
ETH Zurich
8093 Zurich
Switzerland
E-mail: ehafen@ethz.ch

Methods Inf Med 2014; 53: ■■■
doi: 10.3414/ME13-02-0051
received: November 4, 2013
accepted: January 4, 2014
prepublished: ■■■

servations have not yet found their way into clinical practice [3, 4].

Another attribute of the present health-care system is, that it is heavily focused on treatment and much less on prevention. In Switzerland, less than 3% of the entire healthcare spending goes to prevention [5]. For the individual, staying healthy is more beneficial than quickly recovering from illness. While neither the state nor the healthcare providers can force individuals to lead a healthy lifestyle, self-empowerment and personal feedback can help. Moreover, understanding what keeps individuals healthy is as important as finding out why people get ill. For example, it is most informative to find out what genetic or environmental factors keep some 85 year-old chain smokers from getting lung cancer. Per definition the current medical research largely focuses on the understanding of diseases and their treatment.

2. Technology – Two of Three Pillars of Precision Medicine and Personalized Health

Understanding the complex interaction between an individual's genome, the environment and its behavior in order to better prevent or treat diseases in a targeted fashion, is extremely challenging. Two technological revolutions aid the progress towards this goal, however. The costs of DNA sequencing have dropped more rapidly than Moore's law of decreasing cost for computing power. In the next 2–3 years, costs for a complete genome sequence will drop below \$1000 [6, 7]. Information about drug susceptibility (i.e. Clopidogrel, see above) will be readily available. This will generate large amounts of sensitive data about individuals. Discussions are ongoing how to handle, store and interpret these data [8]. The second technological revolution that is well underway is the ability of individuals to monitor an increasing number of health parameters via sensors and applications connected to smartphones. There are currently over 40,000 health-related Apps for smartphones available [9]. They range from monitoring

weight, blood pressure, blood oxygen and glucose levels, sleep patterns, food intake, electrocardiograms, and every possible fitness activity. The ability to monitor vital health parameters continually over time without a visit to the doctor's office, offers an extremely rich and dynamic source of personal health information that provides direct and personal feedback to the individual. Furthermore, in an aggregated form, such data from large numbers of individuals can be correlated with their individual genome sequences to learn more about the complex interaction between an individual's genotype and his/her phenotypes (i.e. health parameters). This is the promise of personalized prevention and treatment and has been coined precision medicine [10] or P4 (personalized, preventive, predictive and participatory) medicine [11]. However, these two technological advances alone are only two pillars of the personalized health platform. The data are still looked up in various data silos (e.g. apps, genome sequencing companies) and can rightfully not be accessed without the permission of the citizen from whom the data stem.

3. The Citizen – The Third Pillar in Precision Medicine and Personalized Health

When we talk about the need to integrate data in health, we talk exclusively about health data of individuals, from medical records, anamnesis, genome data, genome-based information [12] to mHealth data. These data are sensitive and in most countries the citizen has the right to access his data and control who can access it [13]. Aggregation of such data cannot occur without the explicit permission of the individual, particularly when the identity of the person can be inferred. Given that genome data are the ultimate identifier of a person, consent has to be obtained for each analysis. This important, yet difficult process hinders the access to vital information. Even if the patient is willing to provide access to his/her medical data for research, this consent is limited to the doctor or the institution in which the research is conducted. It is difficult for the patient – and

often of little interest to the doctor conducting the research – to make the data available to the research community at large in order to obtain the highest benefit of his/her data sharing. One way to change this is to empower the patients and citizens to control the access to their data themselves. Sharing of relevant health data with the research community generates knowledge that aids improved personalized prevention and treatment. A patient with a rare disease is interested in learning as much about treatments from other patients with the same disease, irrespective of whether they live in Switzerland, Nigeria or the US. While healthcare systems, data protection laws and research regulations are confined to national borders, health is a global issue and can only be solved by finding ways to efficiently share personal health information beyond borders.

Putting the decision who can access the personal health data into the hands of the citizens is an empowerment but also an additional burden. For the citizen to accept this burden and opportunity, a trusted and transparent interaction platform has to be established in addition to promoting health literacy [14]. Furthermore, the risks and the benefits for the individual and for society have to be explained clearly.

4. The Economical Value of Personal Data

In a 2011 study, the World Economic Forum called personal data a new asset class [15]. The Boston Consulting Group estimates that the market value of personal data from Europeans alone could be € 1 trillion in 2020 [16]. These estimates include only the value of data for targeted marketing and loyalty programs. The potential value of health data is not included in this figure since the data are currently largely inaccessible. Thus, in addition to the obvious value of being able to securely access personal medical and health data for personal health and effective treatment, there is large economic value associated with personal data particularly in its aggregated form. In the realm of personal data used for targeted advertisement, large multinational companies compete for ac-

cess to these data and a fast growing market for selling such data has established itself. In the realm of health data the interest from pharmaceutical companies, research organizations and insurance companies is increasing. Currently, this access is restricted on one hand by the decentralized storage of these data and on the other hand by data protection laws that protect the privacy of the individual donor of the data.

Whether dealing with personal data in general or health data in particular, we find ourselves in the strange situation that the citizen as the source of these data has very little control over, let alone benefits from these data. Thus providing individuals with the possibility to safely store, control, manage and share their data would allow them, and society at large, to profit from the value of these data both financially and by the knowledge generated from its aggregation. In the financial world, currencies and banks have established themselves as essential tools and providers to store, manage and invest personal assets. With the rapidly increasing digitization of personal information, similar systems should finally also be established in the healthcare system.

5. Health Data Repository and the Cooperative Model

The value of personal data and health data in particular can best be fully valued to the benefit of the citizen as the source of the data and society at large when the data are controlled by the citizens themselves. For safe storage, management and sharing of the data citizens need a trusted entity such as a bank for financial assets. The primary goal of such a health data bank should be the benefit of the individual account holders and the collective of the holders, not shareholders. The cooperative model of a corporation mirrors this goal perfectly. A cooperative is fully and equally owned and controlled by all its members. Cooperatives have their origin in farming. Farmers formed cooperatives when they decided not to divide parts of their land but to manage it collectively. Cooperatives have established themselves in Switzerland and other

European countries as a successful and long lasting form of commercial entities. Their primary interest is to serve the interests of its members. Being owned by the members, cooperatives cannot be bought or taken-over by other companies easily. To the contrary, they contribute highly to the public good.

The cooperative model for a personal health data repository is particularly well suited since each citizen of the world, whether from a developing or from a Western country, has similar amounts of health data (e.g. all humans have 6 billion letters in their genome). Members of a cooperative purchase a membership certificate for a small unit price (i.e. € 100) and provide data into their account. As a member they have one vote and are entitled to decide on the articles of association and the nomination of the members of the board. Furthermore, they can decide how the revenues of the cooperatives are spent (e.g. for research, information, education).

6. The Health Data Cooperative (HDC) Business Model

Given the value of large amounts of aggregated health data, cooperatives should be profitable without charging users and members for data storage and data maintenance. Pharmaceutical and medtech companies will pay to query the data that users have agreed to share. They may also query the health data records to recruit patients with specific characteristics for clinical trials providing that account holders have indicated their willingness to participate in such trials. Furthermore, pharmaceutical companies could return the results of the clinical trials to the account of the participants. In this way, the decision, whether to make the clinical trial data open, rests with the data owner, i.e. the participant in the trial. Therefore, the problem that pharmaceutical companies must not make clinical trial data accessible because they are bound by restricted consent regulation with the participants can be resolved. Moreover, querying data for in life trials or phase IV studies would be particularly useful if citizens agree that not only their

medical data but also their mHealth data are queried.

Revenues obtained by the HDCs form companies and research organizations (which will pay merely cost-covering fees) will be reinvested into research programs, information platforms or continued education programs. Members who share data will be able to specify how the revenues are invested.

7. The HDC Structure

The HDC platform is a user-friendly and secure repository for different types of health data that permits citizen participation in research. It is composed of three main components: a) the “core” for data acquisition and secure data management, b) the “app store” to provide meaningful applications and visualizations for citizens of their data, and c) the “Big Data” analytics system.

The core is based on a semi-structured database system that allows health-related data to be stored in different formats in a secure way [17]. In addition, it provides the basic infrastructure to group data into different “spaces” (e.g. wellness, fitness, outcomes, medical or health records) and to share data within “circles” [18, 19]. For instance, a user can add a doctor to the user’s “medical” space, thereby allowing the doctor to see all of the user’s medical records.

The HDC is an open platform that allows the integration of different kinds of data from multiple sources. This openness is important to support multiple health care standards (currently, there is no single internationally accepted standard accepted by all stakeholders, no “Health Information Markup Language”). Besides, this open architecture allows the HDC to integrate data that is not subject to standardization, but which is nevertheless important to maintain. A typical example is wellness and fitness data such as statistics about workouts and nutrition, and in the future heart rate and regularity. To integrate and visualize these data, each record is associated to an app. An app defines the structure of the data, its visualization, and mappings to records generated by other apps.

The most critical component for this project is the Big Data platform required to carry out large-scale analytics on all the data generated by apps and stored in the HDC core. Machine learning technology is employed to automatically validate health status and disease models that are extracted from large data sets. For this purpose, the data is cleansed of personal identification information and integrated using the app-specific mappings. The Big Data platform allows researchers to correlate the data in different ways, thereby providing new insights into the cause(s), progression, and treatment of complex diseases. In the medical domain, vast amounts of data are generated without expert annotations. Semi-supervised learning is employed to harvest this data and to guide the modeling process by precious expert information from medical practitioners to generate predictive models for diagnosis and prognosis of therapy.

The HDC would operate its digital data repository in a cloud solution provided by a trusted Swiss cloud-computing provider. In such an HDC, the accounts are managed exclusively by the account holders (members or users). Data is entered on the request of the account owner from healthcare providers and smartphone applications via specific apps that contain the API for the data provider (see above). Via an easy-to-use user interface (dashboard) the user has the ability to annotate the data (comments, corrections) and sort the data into user-defined spaces (i.e. fitness, check-up, knee injury etc.). Data in spaces can be shared with others (doctors, friends, research, industry, public health institutions etc.) in user-defined circles. In this way, the account owner decides which data to share and with whom. Moreover, each citizen who is a member or a user of the HDC is in full control over all his health data and citizens have access to the data from anywhere in the world. In contrast to the individual consent signed by the patient and the researcher, HDC acts as a trusted broker on behalf of its members for the data they make available for research. In this way, the data is made available to the research community at large according to the specification of the data owners. Hence maximal knowledge can be obtained from

this data. Since HDCs act solely in the interest of their members it must also take the utmost precaution to prevent misuse of the data.

8. A Federation of National Health Data Cooperatives

The global nature of health and the fact that each citizen in the world has similar amounts of valuable health data is best reflected in a federation of national citizen-owned health data cooperatives that share a common IT infrastructure and a common data storage instance. In this way, millions of data sets that users in different countries agreed to share can be queried. The revenues would flow back to the national cooperatives with their different funding needs. The goal that citizens across the world become pro-active participants in managing their own health and at the same time contribute to a better, more effective and efficient personalized healthcare would make a true revolution and would lay the basis for the creative destruction of medicine as postulated by Eric Topol [20]. In many ways, health data cooperatives would solve for healthcare what banks have solved for the financial sector during the last several hundred years. Moreover, with the important improvement of the citizen-centered cooperative structure of health data repositories, the striving for excessive gains for personal and shareholder satisfaction that has contributed to the recent financial crisis might have been partly prevented.

9. Conclusions

Rapid technological advances in genome sequencing and mobile health monitoring paired with the decreasing costs of data storage and computing power and the increased security capabilities of cloud-based storage and analysis systems, make the formation of health data cooperatives possible and timely. Their success depends on an entirely democratic, bottom-up process that is initiated by citizens being the major drivers of these innovations. A key factor for rapid adoption is trust and transparen-

cy. Even though no system can guarantee absolute data security, such health data repositories should be certified by independent government regulatory bodies and good governance structures. Flying in an airplane is not without risks but we trust that airlines take care of their airplanes and the FAA regulates which airplanes can transport passengers. Similarly the FDA controls drugs produced by pharmaceutical companies. Furthermore, the focus of HDCs should be on solidarity and harm mitigation. In this way, citizen-owned personal data cooperatives would efficiently unleash the huge value of personal data and contribute to a democratization of healthcare and the global economy and global access of health interventions at the same time.

References

1. Yong PL, Saunders, RS, Olsen LA. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. IOM (Institute of Medicine). Washington, DC: The National Academies Press; 2010.
2. Scott SA, Sangkuhl K, Stein CM, Hulot JS, Mega JL, Roden DM, et al. Clinical pharmacogenetics implementation consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clin Pharmacol Ther* 2013; 94 (3): 317–323.
3. Harper AR, Topol EJ. Pharmacogenomics in clinical practice and drug development. *Nat Biotechnol* 2012; 30 (11): 1117–1124.
4. Frueh FW. On rat poison and human medicines: personalizing warfarin therapy. *Trends Mol Med* 2012; 18 (4): 201–205.
5. Kosten des Gesundheitswesens nach Leistungen [database on the Internet]. Neuchatel, Bundesamt für Statistik 2013 cited 2013 Nov 03]. Available from: <http://www.bfs.admin.ch/bfs/portal/de/index/themen/14/05/blank/key/leistungserbringer.html>.
6. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). 2013. Available from: <http://www.genome.gov/sequencingcosts>.
7. Harvey A, Brand A, Holgate ST, Kristiansen LV, Lehrach H, Palotie A, et al. The future of technologies for personalised medicine. *N Biotechnol* 2012; 29 (6): 625–633.
8. Horgan D, editor. Innovation and Patient Access to Personalised Medicine. EAPM Report from Irish Presidency Conference March 20/21, 2013. Available from <http://euapm.eu/wp-content/uploads/2012/07/EAPM-REPORT-on-Innovation-and-Patient-Access-to-Personalised-Medicine.pdf>
9. Pelletier SG. Explosive Growth in Health Care Apps Raises Oversight Questions. Association of American Medical Colleges; 2012. Available from

- <https://www.aamc.org/newsroom/reporter/october2012/308516/health-care-apps.html>
10. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. The National Academies Press; 2011. Available from http://www.nap.edu/openbook.php?record_id=13284
 11. Hood L. Systems biology and p4 medicine: past, present, and future. *Rambam Maimonides Med J* 2013; 4 (2): e0012.
 12. Brand A. Public Health Genomics and personalized healthcare: a pipeline from cell to society. *Drug Metabol Drug Interact [Editorial]* 2012; 27 (3): 121–123.
 13. EC Press Release: Better use of health data will transform the healthcare landscape, says expert report [database on the Internet] 2012 [Cited 2013 Nov 03]. Available from: http://europa.eu/rapid/press-release_IP-12-453_en.htm.
 14. Sorensen K, Brand H. Health literacy: the essential catalyst for the responsible and effective translation of genome-based information for the benefit of population health. *Public Health Genomics* 2011; 14 (4–5): 195–200.
 15. Personal Data: The Emergence of a New Asset Class. World Economic Forum; 2011. Available from <http://www.weforum.org/reports/personal-data-emergence-new-asset-class>
 16. Rose J, Rehse O, Röber B. The Value of Our Digital Identity. Boston Consulting Group; 2012. Available from https://www.bcgperspectives.com/content/articles/digital_economy_consumer_insight_value_of_our_digital_identity/
 17. Arasu A, Blanas S, Eguro K, Joglekar M, Kaushik R, Kossmann D, et al. Secure database-as-a-service with Cipherbase. Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data; New York, New York, USA. ACM 2013. pp 1033–1036.
 18. Sanamrad T, Braun L, Marfurt A, Kossmann D, Venkatesan R. POP: A New Encryption Scheme for Dynamic Databases. Technical Report Nr. 782, Systems Group, Department of Computer Science, ETH Zürich, Zürich. 2013. Available from <ftp://ftp.inf.ethz.ch/pub/publications/tech-reports/7xx/782.pdf>
 19. Bamford R, Borkar VR, Brantner M, Fischer PM, Florescu D, Graf DA, et al. XQuery Reloaded. *PVLDB* 2009; 2 (2): 1342–1353.
 20. Topol EJ. The creative destruction of medicine. New York: Basic Books; 2012.