

Chapter 22

Governance and IT Architecture



Serge Bignens, Murat Sariyar, and Ernst Hafen

Abstract Personalized medicine relies on the integration and analysis of diverse sets of health data. Many patients and healthy individuals are willing to play an active role in supporting research, provided there is a trust-promoting governance structure for data sharing as well as a return of information and knowledge. MIDATA.coop provides an IT platform that manages personal data under such a governance structure. As a not-for-profit citizen-owned cooperative, its vision is to allow citizens to collect, store, visualize, and share specific sets of their health-related data with friends and health professionals, and to make anonymized parts of these data accessible to medical research projects in areas that appeal to them. The value generated by this secondary use of personal data is managed collectively to operate and extend the platform and support further research projects. In this chapter, we describe central features of MIDATA.coop and insights gained since the operation of the platform. As an example for a novel patient engagement effort, MIDATA.coop has led to new forms of participation in research besides formal enrolment in clinical trials or epidemiological studies.

1 Introduction

Developing new kind of data-science-based products and services relies more and more on the analysis of personal data, especially in the medical domain (Dhar 2013; Jeff Leek 2013). A key question about personal data is by whom and how they are controlled. Medical data generated in the course of healthcare services are stored by the healthcare providers and cannot be used for research without the explicit informed consent of the individual. Health-related data generated by mobile apps, smartphones, and sensors are collected and stored by the service provider. By

S. Bignens (✉) · M. Sariyar
Institute for Medical Informatics, University of Applied Sciences Bern, Bern, Switzerland
e-mail: serge.bignens@bfh.ch

E. Hafen
Institute of Molecular Systems Biology, ETH Zürich, Zürich, Switzerland

accepting the general terms of use of the devices or apps, the users usually sign off the rights of their data reuse to the service provider, thereby ceding control over the secondary use of their data. Services providers and in general those who control big amounts of data have little incentive to abandon the siloed approach to data, which in turn hinders true integration of different personal data sets.

Effective integration of personal data for personal health and precision medicine will be facilitated if the individuals regain the control over the secondary use of their personal data (the individual can then decide to hand this control over to a third party). Article 15 of the new EU general data protection regulation (EU GDPR: <https://www.eugdpr.org>) introduces the right to get a copy of such data. In addition, Article 20 introduces the right to data portability, which enables individuals to transfer at least parts of their personal data from one service provider to another and obliges data controllers to have systems in place that allow individuals to realize their data portability right (Datenschutzbeauftragte 2018). In Switzerland, the federal council has accepted to evaluate a parliamentary motion for the “Right to a Digital Copy” of one’s personal data (Derder Fathi 2015). The Swiss Right to a Digital Copy, if approved, would grant individuals the right to control data reuse. Therefore, it can potentially empower citizens to actively control the secondary use of their personal data (Hürlimann and Zech 2016).

As the availability of comprehensive health data is more and more crucial for better health outcomes, there are worldwide initiatives targeting the patient-controlled integration of health data. A synonymous term to personal health record, which is used especially in the USA is “health record banking” (Yasnoff and Shortliffe 2014), used for example by the Health Record Banking Alliance (Health Record Banking Alliance 2018). The main goal of this alliance is to increase public awareness of personal health records and share lessons learned and best practices in personal health records. A systematic literature review on this topic is provided by A. Roehrs et al. (2017) that reviewed over 5000 articles. Especially the difference in governance structures (there are also personal health record initiatives, where the data management is not in the responsibility of the citizen) and in the goals (just recording or also manipulating and analyzing the data) are worth mentioning.

It is still a challenge make personal health record data available for research. The following activities are essential for implementing applied data science projects based on personal data:

- Definition of **benefits** resulting from new forms of data management and analytics
- Specification of the data to be **collected** and its source
- Consideration of legal issues relating to the usage of the data, particularly when personal data of **individual citizens** are involved
- Implementing a **governance** and business model for the collection and use of the data
- Operation of secure and scalable **IT platform** that can be augmented by third-party applications
- and finally, taking advantage of the present and future **opportunities** while managing the risks.

This chapter describes these different aspects and illustrates them with the help of the citizen science initiative named “MIDATA.coop”¹ [for a general introduction to citizen science, we refer to Irwin (2001) and Bonney et al. (2009)] that has addressed those challenges and is solving them in a technical, economical, and ethical way.

MIDATA.coop embodies an IT platform managing personal data and the governance needed to operate it. As a not-for-profit, citizen-owned cooperative, its vision is to allow citizens to collect, store, visualize, and share specific sets of their health-related data² with friends and health professionals, and to make anonymized parts of these data accessible to medical research projects in areas that appeal to them. The value generated by this secondary use of personal data is managed collectively to operate and extend the platform and support further research projects.

2 Utility and Benefits in Using Personal Health Data

The key to successful data science projects is addressing a specific need and producing direct benefits for all stakeholders involved in the process of generating, gathering, managing, using, and analyzing data.

Regarding personal data in general and health data in particular, the stakeholders are manifold (see Fig. 22.1). There are patients or just citizens (if we consider the healthy ones as well), health professionals as individuals or healthcare institutions, researchers, industry (in particular the pharmaceutical, medtech, and IT industries), and public health policy makers. In the following, we describe the kind of benefits that can be targeted for each stakeholder group when using MIDATA as a platform for applied data science.

2.1 Benefits for Citizens and Patients

Collecting health data and making them available to citizens and patients empowers and transforms them from rather passive objects in the healthcare system to active subjects that are empowered to influence the resource allocation in the healthcare system by governing the aggregation and distribution of their data (Woolley et al. 2016). The citizens can be supported in collecting and visualizing their data. As an extension of the quantified-self movement (Appelboom et al. 2014), in which fitness

¹<https://www.midata.coop>

²While in a first phase MIDATA.coop targets health-related data, its governance and IT architecture allow to extend its use to other personal data, for instance, education data, whose secondary use value is rapidly growing.

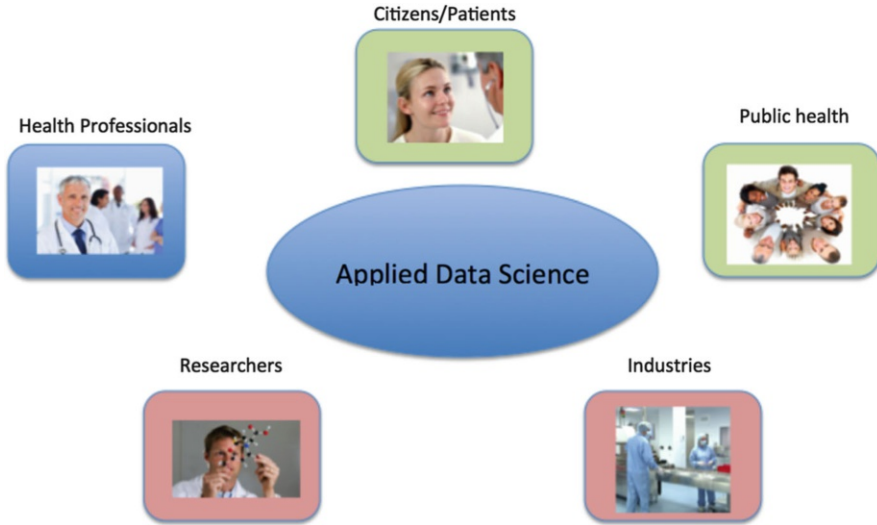


Fig. 22.1 Stakeholders involved in data science in healthcare

conscious people monitor their performance (steps, speed, altitude, heart rate, etc.), nowadays, thanks to the rapid spread of mobile sensors and tracking devices, people can also monitor several types of health data, including but not limited to blood pressure, glycemic index, pain level, and sleep quality.

Harnessing these new possibilities of broad data collection and combining them with (big) data analytics approaches, data science projects aim to generate new knowledge about risk factors, the development of new diagnostic tools, the impacts of new treatments or medications. Citizens and patients will benefit from and become active contributors to this new knowledge.

2.2 Benefits for Health Professionals

Health professionals benefit directly from good, accurate, and continuous health data from their patients. This allows them to take more informed and therefore better decisions, thereby increasing the quality of the treatment. They also benefit from the aforementioned new knowledge and new diagnostic tools and possible treatments that result from data science projects.

2.3 Benefits for the Researchers

Researchers are among the main actors in data science projects, which they mostly lead. They benefit from new ways of recruiting study participants, obtaining additional data and faster access to information, as well as new and more powerful data analytics tools. Furthermore, they benefit from the possibilities to aggregate data from multiple sources.

2.4 Benefits for the Pharmaceutical, Medtech, and IT Industries

Data are key for developing new drugs, devices, IT services as well as for achieving regulatory approval required to bring them on the market. The improved ability to obtain access to the necessary data, to link and aggregate the data from different sources, to gain and manage the consent of the data owners in a defined time frame are important assets.

2.5 Benefits for Public Health

Policy making, in particular in the health and public health domains, has to rely on evidence and data. That is not only true in industrialized countries, but also in low- and middle-income ones. These countries do also suffer from the burden of the rise in chronic diseases. Their scarce resources force them to optimize the actions they plan and can afford to take. Availability of data in sufficient quality allows such optimization.

3 Which Data?

We know that our health depends not only on our genetic constitution and the effectiveness of the treatments that our healthcare system can provide, but also on the environment we live and work in, and on our lifestyle. “Environment” encompasses many aspects including air quality, noise level, traffic intensity, pollen density, weather conditions, etc. For health data science initiatives, data concerning all those determinants are relevant and must be considered in specific combinations depending on the targeted benefits.

Generally, data about our **health status** and the treatments provided are generated mainly by healthcare professionals and their institutions (hospitals, home care, radiology institutes, etc.). They may include vital signs, blood analysis, diagnostics,

medications, radiology imaging, and omics data. These data are managed by a very large set of applications in structured and unstructured format.

In addition to that, **environmental data** are widely available such as weather data (UV light, air pressure, and temperature) and to some extent air quality data (small particles, NO_x, etc.), even though much less for noise, exposure to viruses, and other environmental data. Environmental data can be linked to an individual based on geolocation, but geolocation is not always known/collected and environmental data are not available for all localizations with the necessary precision.

The collection of **lifestyle data** has seen a rapid development, starting from the *quantified-self* movement mainly around sports activities and extending to the daily monitoring of physical activity, weight, blood pressure, and respiratory capacity. The emergence of low-cost sensors, the possibility to connect these sensors either to the Internet (through Internet of things, IOT) or to smartphones, and the availability of health (mobile) applications to visualize the values and evolution of these data have accelerated the adoption by the population (Takacs et al. 2014). This results in a large amount of data collected first by the users and then stored in a multitude of cloud services.

One important field that acquires and requires all these kinds of data is **precision medicine**, which aims at refining our understanding of disease onset and progression, treatment response, and health outcomes through the more precise measurement of genetic, treatment, and environmental and behavioral factors that contribute to health and disease (National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease 2011). Some forms of precision medicine have already been established, particularly in the field of cancer therapy, where certain drugs are given depending on individual tumor genetics. Including self-collected data on environmental and behavioral factors will foster the extension of individualized medicine (Swan 2012).

One of the challenges in data science projects is the ability to collect the necessary data from various sources (Töpel et al. 2008). The main challenge, however, lies in the correct interpretation of the data. Nowadays, most of the clinical data are represented in medical reports in free text form. Interpretation of these data necessitates either manual coding, which is very resource-consuming and therefore limited, or natural language interpretation through machine learning algorithms that need many well-annotated training data in order to generate good results (Savova et al. 2010; Uzuner et al. 2010). The availability of data in structured form is a large advantage for data science, but it is not sufficient. The format and the semantics used to structure these data are most important. In the healthcare domain, international standards for ontologies and classifications have emerged, like, for instance, *SNOMED* (SNOMED International 2018) for coding medical data or *LOINC* (LOINC 2018) for coding laboratory results. Unfortunately, some standards are not for free use and most of them are not widespread in the current information systems of the healthcare providers.

4 Trust-Promoting Frameworks for Data Science Projects

Data science initiatives are run today in the private and the public sectors with different governance and business models and promote different levels of trust. Both, the business and public sector models, relying on profit and the general public good, leave very little participation space for the citizens, the owners of the personal data. Therefore, these models do not allow a relationship with the citizens that would facilitate the extension of data science projects beyond the initial goal targeted by the company or the research institution. One important mechanism for a long-term commitment of citizens to data sharing is trust (Kaye et al. 2009).

To have citizen that engage with trust and active participation in data science initiatives and to develop a new and fair data economy, the following pillars are important from our experience-based point of view: democratic organization, not-for-profit model, transparency, and data security (Hafen et al. 2014):

- **Democratic organization:** The cooperative form is quite adequate for the entity in charge of collecting and managing personal data. The democratic principle “one member one vote” of cooperatives fits particularly well as every citizen has similar amount of personal data to be managed. In this way, the individual is able to make various sets of personal data (genome, nutrition, geolocation, medical, and fitness data) accessible for novel data services and research projects. The data cooperative as a representation of its users ensures that the data made accessible by the data subjects will only be used for the purpose of the service or data science project and there is a fair financial compensation for data use that goes to the not-for-profit cooperative, and can be used for actions the cooperative decides to be useful, for example, extending the information dissemination regarding scientific results.
- **Not-for-profit:** This model finds its justification in the fact that the economic value of the data is not in the individual data sets but in the whole data set resulting from the aggregation of the data of all persons. Therefore, the value of the data should not belong to a private organization or be paid to individuals but managed in a not-for-profit manner to the benefit of the society, represented by the cooperative. Financial incentives to share personal data should not be distributed. For instance, in the case of blood donation, numerous studies have shown that the quality of the blood is worse when the donor receives a direct payment (Mellström and Johannesson 2008). The large economic value resulting in the aggregation of data under the control of the citizen flows back into the cooperative, and its members can decide how to allocate those resources for the development of new services and for conducting further data science projects in the same not-for-profit framework.
- **Transparency:** The organization of the cooperative with its general assembly of its members is one of the elements of transparency. The cooperative should also nominate an ethics committee, which overlooks and validates the initiatives, projects, and new services that the cooperative targets to operate. On the

information technology side, the software should be open source, so that its quality and purpose can be verified at any time by any person.

- **Data security:** In order to protect privacy and increase trust the latest data security mechanisms, such as advanced encryption technologies, should be used and validated with regular security audits. Having a public-key infrastructure allows the user to have the data encryption key under his or her control. In this way, even database administrators have no access to the data.

The combination of transparent, democratic, not-for-profit governance and secured open-source infrastructure is the basis and condition for the emergence of a new sustainable data economy supporting fair data science projects. Not only traditional services and research projects can be run on such a framework but also innovative services from third parties can be offered.

The MIDATA.coop is an example of a cooperative that is based on these pillars. Its articles of association imply citizens as members, one cooperative share per member, and one vote per share. As any cooperative, it is managed by its board members. To assure that the goals are followed and the means used by the cooperative are aligned with its vision and ethical rules at any time, an ethics committee has the task to review all data science projects before they are allowed to be operated on the framework. In addition, an international and multidisciplinary advisory board has been constituted.

In the MIDATA ecosystem, the IT platform (data storage and access management governed by the member-owned cooperative) is separated from the data applications. This forms a novel type of personal data innovation framework. As data account holders, the citizens can choose from a variety of different data services and research activities in which they decide to participate. Startup companies, corporations, and research groups can offer services in the form of mobile applications that analyze and collect data (e.g. medical, activity, and nutrition data)—data that hitherto have resided in noncompatible data silos. Access authorization always resides with the data owner.

As mentioned earlier and illustrated in Fig. 22.2, the cooperative acts as the fiduciary of the account holders' data. It negotiates with research organizations, data service providers, and pharmaceutical companies the financial conditions for accessing the data that individuals have authorized for these specific use cases. Such use cases may include the recruitment of patients for clinical trials, providing mobile app-based health services or drug side effect reporting (outcomes-based medicine). The revenues generated will be allocated to the maintenance of the platform, security checks, upgrades, and on additional services for the account holders. Members of the cooperative will decide how additional profits will be invested in projects (e.g. research projects) that will profit society. Value is created collectively; there is no pressure on individuals to sell their individual data.

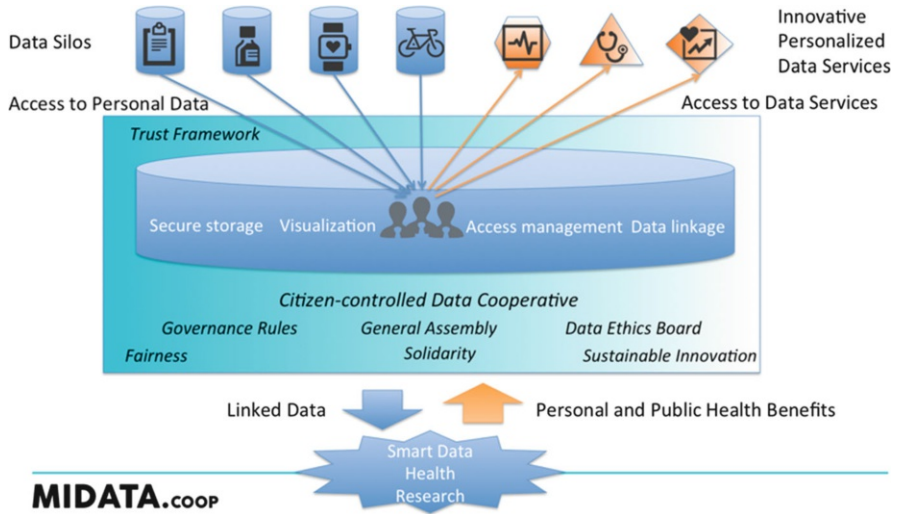


Fig. 22.2 MIDATA.coop governance and ecosystem

5 IT Platform

The ability to make use of a right to a digital copy of personal data requires a suitable and trusted IT infrastructure to securely store, manage, and control access to personal data. Such a trust-promoting framework for active data sharing for personal data services and medical research rests on data security, the individual’s control over their data, and a fair as well as transparent governance.

A suitable IT infrastructure that meets these criteria builds on a cloud-based data storage platform. Individuals open their own data account in which individual personal data records are securely stored. Only the data owners control to whom they grant access, to which data sets, and for what purpose.

To illustrate this, Fig. 22.3 shows the component architecture design of the MIDATA IT platform. This architecture has a clear separation of data acquisition, data management, and data analytics.

The data acquisition can be done interactively through mobile applications or a web portal or indirectly by importing data from external data sources. Sensors can either be integrated with mobile applications, which are directly connected to the MIDATA server or communicate with third-party servers, like the platforms provided by large activity tracker vendors.

The interface to the MIDATA server is provided through an application programming interface (API) that complies with the new standard FHIR (FHIR v3.0.1 2018) (Fast Healthcare Interoperability Resources framework). This standard, developed by the standardization organization HL7.org finds a rapidly growing acceptance for the connection of mobile health applications (mHealth) with back-end systems like the secured cloud services provided by MIDATA. The FHIR standard not only

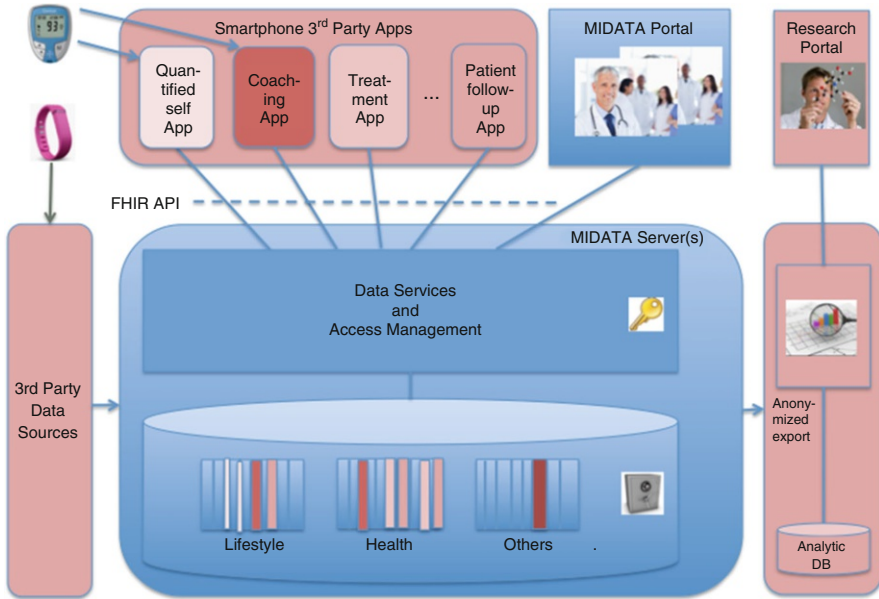


Fig. 22.3 Architecture of the MIDATA IT platform and third-party integration

defines the syntax of the messages exchanged between the mHealth Apps and the server but also the syntax and most importantly the semantics of the information to be stored and managed by the server. FHIR does not define “yet another semantics” but allows to reference well-known and widespread coding systems from the healthcare domain like *LOINC* or *SNOMED*.

The core of the MIDATA IT platform is the data management part. The data management is offered and implemented as “secured cloud services” and interacts, on the one hand, with the different data acquisition channels and, on the other hand, it can export data for analysis in anonymous or nominative form depending on the requirements and on the consent type. The MIDATA server is composed of a first layer, handling the identification and authentication of the users and also handling the access management and the cryptography to protect the data. Past this layer each data element is stored as a single object (in JSON format). Within the JSON object, the syntax and semantic of FHIR is used. These JSON objects are stored in a NoSQL database, thus allowing great flexibility in handling the large diversity of data types that are typically encountered in the health domain.

The data analytics part needs dedicated tools for the researchers who work on the data that the data owners have shared with them and which is therefore implemented in separate components.

The clear separation between mobile applications and data management allows third parties to develop and distribute mobile applications, which can be for-profit. Thus, an ecosystem can emerge where profit can be generated from value-added services for collecting and visualizing data, while the management of the data would

always remain a not-for-profit activity, with citizens retaining control about sharing and using their data. To allow the ecosystem to grow with development done by third-parties, MIDATA focuses on the development, operation, and maintenance of the components painted in blue in Fig. 22.3. Moreover, MIDATA enables third-party actors to develop, operate, and maintain the pink components.

6 Data Protection and Security

Data protection is key to any data science project. Data protection is not only required by legislation but is also mandatory to build and maintain trust with the data owners. Data protection means in particular allowing access to data only to parties that have received explicit informed consent by the data owners. In addition to that, data security implies management of the data based on security aspects, regular verification that security is continually assured, and protection against data loss.

The MIDATA IT platform enforces data protection by:

- Allowing a citizen to register to the platform and become an account owner
- Authenticating each data owner using the platform
- Securely managing the data of each data owner
- Allowing a data owner to share data with another user or with a third party conducting a data science project
- Managing the access to the data of each data owner
- Allowing a data owner to delete his/her data
- Allowing a data owner to withdraw from the platform and have all data optionally exported and then deleted
- Identifying each researcher using the platform
- Managing descriptions provided by researchers of each of their data science project as a basis for receiving explicit informed consent
- Managing the consent of each data owner willing to participate in the data science project and sharing part of his or her data in nominative, coded, or anonymized form
- Allowing each participant to withdraw consent to MIDATA-related aspects of a project

In addition to the services provided by the MIDATA IT platform, additional organizational measures have been taken, such as:

- Identifying users as real persons in order to prohibit fake users
- Managing the register of the researchers using the MIDATA IT platform
- Managing and vetting the MIDATA administrators of the MIDATA IT platform
- Review of the ethical quality of services by a dedicated ethics committee

On the MIDATA IT platform, each data item is stored and managed as a single record. Each record is encrypted with a first key, which is stored with other similar

keys in an access permission set. This access permission set is encrypted with a second key. In a third step, this second key is encrypted with the public key of the data owner.

A data owner willing to access his/her data will use their primary key to decrypt the second key that allows them to decrypt and read the access permission containing the keys to finally decrypt, access, and read the data. All those operations are triggered by the user but executed by the MIDATA IT platform, thus hiding this complexity to the user.

For a data owner giving consent to share data (referenced in one of his or her access permission sets) with a researcher or with another user, the second key that had been used to encrypt that access permission set will then be encrypted with the public key of the researcher or of the other user.

In this way, the researcher or the other user uses his/her primary key to decrypt the second key that allows him/her to decrypt and read the access permission containing the keys to finally decrypt, access, and read the data.

Security audits are run by external independent and recognized security expert organizations on an annual basis. These audits check that no unauthorized access to the platform and the managed data is possible. Some of those intrusion tests are run with no user login available to attempt access to any data; other tests are run with a user login with the intent to access more data than allowed.

7 Example of a Data Science Project Running on MIDATA

The MIDATA IT platform is used for one data science project in which patients with Multiple Sclerosis (MS) are monitored at home. It is a collaboration between MIDATA, University of Applied Sciences Bern and Andreas Lutterotti from the Neurology Clinic at the University Hospital in Zürich. MS is a prototypic autoimmune disease characterized by recurrent areas of inflammation in the brain and spinal cord (Sospedra and Martin 2005). With more than 2 million patients worldwide and more than 10,000 in Switzerland, MS is one of the leading causes of neurological disability in early adulthood. Like many other diseases, MS is a complex disease, where both etiology and disease course are strongly influenced by an interplay of genetic, environmental, and lifestyle factors (Olsson et al. 2017).

The whole project is set to measure both, (1) objective neurological function by implementing different tests for cognition, hand and arm function, as well as gait; (2) patient reported outcome measures using standardized questionnaires or visual analog scales to assess neurologic function, cognition, fatigue, and mood disorders; and (3) gather information on lifestyle and environmental factors that can influence the course of the disease. A smartphone app has been developed that allows continuous monitoring of patients with Multiple Sclerosis (MitrendS App; see Fig. 22.4). This app has two tests: one assesses the speed and correctness in reproducing a symbol-sign assignment, and the other the ability to memorize a path through a graph. It will be used by MS patients in order to monitor their motoric

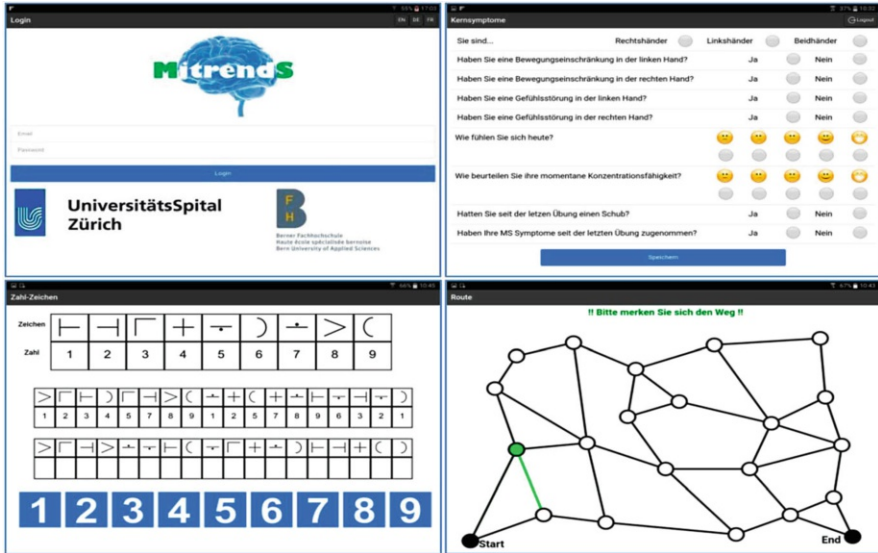


Fig. 22.4 Screenshots of the tests implemented in the MitrendS App

and cognitive capabilities and compare the results of the tests throughout time. Up to now, the app has been successfully tested with healthy persons (results are reproducible) and the recruitment of patients has started.

Tools to predict clinical evolution of MS for individual patients at the time of diagnosis or to identify the most effective treatment are major unmet medical needs for MS patients and their treating physicians. All currently available clinical/imaging measures or biomarkers used for risk assessment perform relatively well at the group level, whereas individual predictions for single patients are not yet available. Concerning the analysis of disease progression, the project intends to study the correlation between external factors such as sun exposure (there is no need for the patients to enter such information with geolocalization) and MS symptoms. Particularly, climatic and environmental variables in combination with lifestyle factors are the most easily accessible external factors that could influence the progression of MS and therefore those are the ones that will be used in the analysis.

8 Conclusion and Lessons Learned

MIDATA.coop is an example for a novel patient engagement effort. Organized as a citizen-owned not-for-profit cooperative, it allows patients and citizens to securely store and manage personal data. This leads to new forms of participation in research (by patient empowerment) besides formal enrolment in clinical trials or

epidemiological studies. In addition to that, much more data can be considered by the switch from institution-controlled data to citizen-controlled data, leading to new opportunities for precision medicine.

The following lessons have been learned since the MIDATA platform has been operational (2016):

- The widespread smartphone use in the population is a major opportunity for the collection of new kind of data and new business models.
- Many citizens are willing to contribute to science, especially when being affected by diseases or having relatives affected.
- Researchers are more and more using real data such as provided by MIDATA because of the need for a better understanding of all determinants of health.
- Not only should the data be at the center of business models, but also applications using such data.
- Structured data are often lacking; hence, the need for more NLP and text mining applications for unstructured data.
- Interoperability is still an issue due to the lack of agreement on semantics. We suggest FHIR as a promising collection of standards in the health domain.
- Data security such as cryptographic technologies is increasingly important when citizens store an increasing amount of data on a platform.
- Transparency concerning the MIDATA platform and its governance structure is highly important, both for those storing data and those using such data.

References

- Appelboom, G., LoPresti, M., Reginster, J.-Y., Sander Connolly, E., & Dumont, E. P. L. (2014). The quantified patient: A patient participatory culture. *Current Medical Research and Opinion*, 30(12), 2585–2587. <https://doi.org/10.1185/03007995.2014.954032>.
- Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., et al. (2009). Citizen science: A developing tool for expanding science knowledge and scientific literacy. *Bioscience*, 59(11), 977–984. <https://doi.org/10.1525/bio.2009.59.11.9>.
- Datenschutzbeauftragte, daschug G., externe. (2018). *Inhalte und Hinweise zur DSGVO/EU-Datenschutz-Grundverordnung*. Retrieved February 26, 2018, from <https://www.datenschutz-grundverordnung.eu/>
- Derder Fathi. (2015). *Recht auf Nutzung der persönlichen Daten. Recht auf Kopie*. Retrieved February 26, 2018, from <https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20154045>
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64–73. <https://doi.org/10.1145/2500499>.
- FHIR v3.0.1. (2018). Retrieved February 26, 2018, from <https://www.hl7.org/fhir/>
- Hafen, E., Kossmann, D., & Brand, A. (2014). Health data cooperatives - citizen empowerment. *Methods of Information in Medicine*, 53(2), 82–86. <https://doi.org/10.3414/ME13-02-0051>.
- Health Record Banking Alliance. (2018). *HRBA overview*. Retrieved February 26, 2018, from <http://www.healthbanking.org/hrba-overview.html>
- Hürlimann, D., & Zech, H. (2016). Rechte an Daten. *sui generis*. Retrieved from <http://sui-generis.ch/article/view/sg.27>

- Irwin, A. (2001). Constructing the scientific citizen: Science and democracy in the biosciences. *Public Understanding of Science*, 10(1), 1–18. <https://doi.org/10.3109/a036852>.
- Jeff Leek. (2013). *The key word in “Data Science” is not data, it is science*. Retrieved February 26, 2018, from <https://simplystatistics.org/2013/12/12/the-key-word-in-data-science-is-not-data-it-is-science/>
- Kaye, J., Heeney, C., Hawkins, N., de Vries, J., & Boddington, P. (2009). Data sharing in genomics – Re-shaping scientific practice. *Nature Reviews. Genetics*, 10(5), 331–335. <https://doi.org/10.1038/nrg2573>.
- LOINC. (2018). *The freely available standard for identifying health measurements, observations, and documents*. Retrieved February 26, 2018, from <https://loinc.org/>
- Mellström, C., & Johannesson, M. (2008). Crowding out in blood donation: Was Titmuss right? *Journal of the European Economic Association*, 6(4), 845–863. <https://doi.org/10.1162/JEEA.2008.6.4.845>.
- National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. (2011). *Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease*. Washington, DC: National Academies Press. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK91503/>
- Olsson, T., Barcellos, L. F., & Alfredsson, L. (2017). Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nature Reviews. Neurology*, 13(1), 25–36. <https://doi.org/10.1038/nrneurol.2016.187>.
- Roehrs, A., da Costa, C. A., Righi, R. d. R., & de Oliveira, K. S. F. (2017). Personal health records: A systematic literature review. *Journal of Medical Internet Research*, 19(1), e13. <https://doi.org/10.2196/jmir.5876>.
- Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., et al. (2010). Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications. *Journal of the American Medical Informatics Association: JAMIA*, 17(5), 507–513. <https://doi.org/10.1136/jamia.2009.001560>.
- SNOMED International. (2018). Retrieved February 26, 2018, from <https://www.snomed.org/>
- Sospedra, M., & Martin, R. (2005). Immunology of multiple sclerosis. *Annual Review of Immunology*, 23, 683–747. <https://doi.org/10.1146/annurev.immunol.23.021704.115707>.
- Swan, M. (2012). Health 2050: The realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen. *Journal of Personalized Medicine*, 2(3), 93–118. <https://doi.org/10.3390/jpm2030093>.
- Takacs, J., Pollock, C. L., Guenther, J. R., Bahar, M., Napier, C., & Hunt, M. A. (2014). Validation of the Fitbit one activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport*, 17(5), 496–500. <https://doi.org/10.1016/j.jsams.2013.10.241>.
- Töpel, T., Kormeier, B., Klassen, A., & Hofestädt, R. (2008). BioDWH: A data warehouse kit for life science data integration. *Journal of Integrative Bioinformatics*, 5(2), 49–57. <https://doi.org/10.2390/biecoll-jib-2008-93>.
- Uzuner, O., Solti, I., & Cadag, E. (2010). Extracting medication information from clinical text. *Journal of the American Medical Informatics Association*, 17(5), 514–518. <https://doi.org/10.1136/jamia.2010.003947>.
- Woolley, J. P., McGowan, M. L., Teare, H. J. A., Coathup, V., Fishman, J. R., Settersten, R. A., & Juengst, E. T. (2016). Citizen science or scientific citizenship? Disentangling the uses of public engagement rhetoric in national research initiatives. *BMC Medical Ethics*, 17, 33. <https://doi.org/10.1186/s12910-016-0117-1>.
- Yasnoff, W. A., & Shortliffe, E. H. (2014). Lessons learned from a health record bank start-up. *Methods of Information in Medicine*, 53(2), 66–72. <https://doi.org/10.3414/ME13-02-0030>.