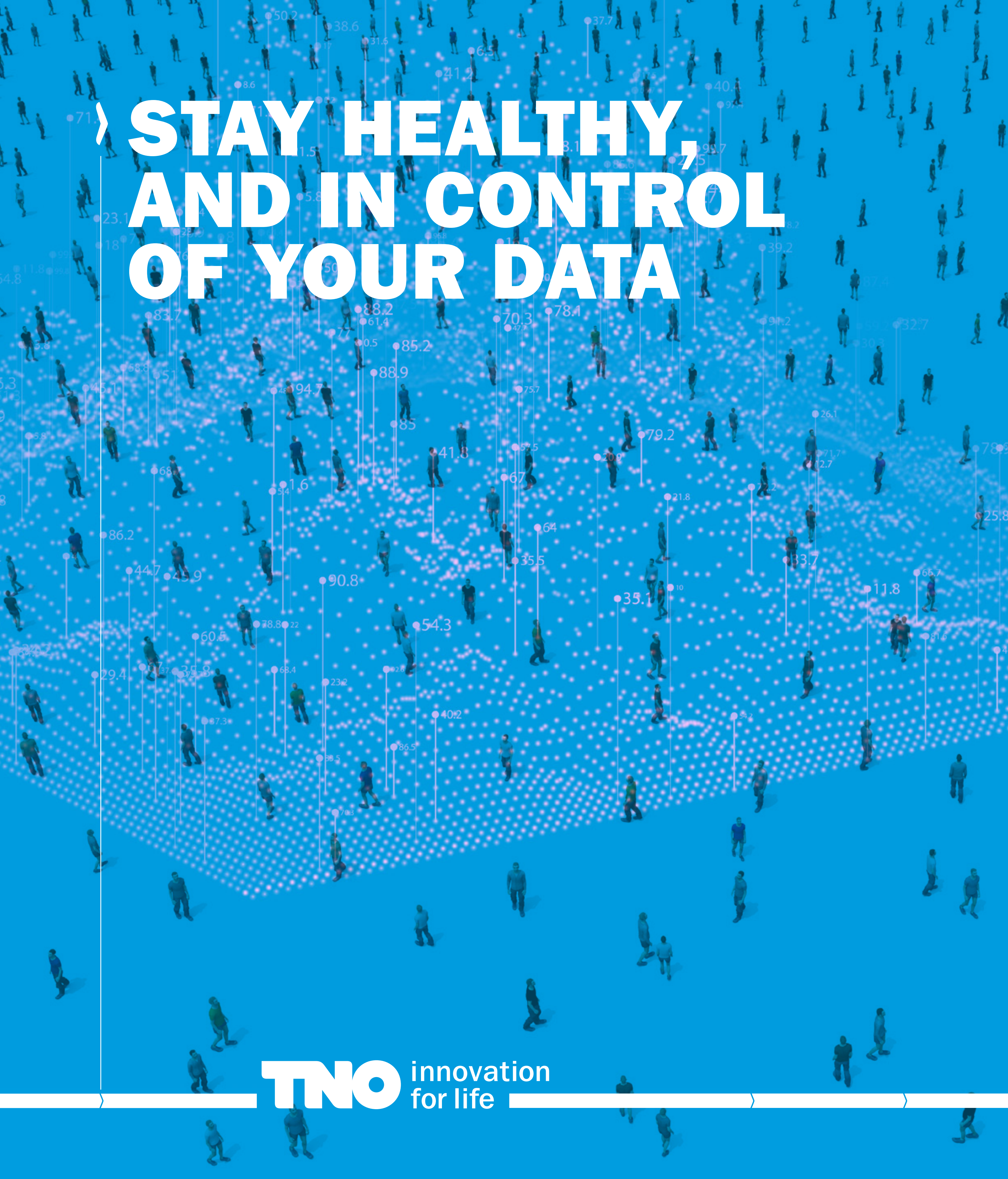
STAY HEALTHY, AND IN CONTROL OF YOUR DATA



TNO innovation for life

ABSTRACT

Nowadays the benefits of the Internet and digital innovation are permeating into practically every area of our lives and we are amassing vast volumes of data. Is it then not time that we applied more of these benefits, along with the data that helps fuel them, to improving healthcare and making it more cost effective? This paper explores how data science and Artificial Intelligence (AI) can collectively be a game changer in transforming the healthcare sector. It also highlights a few challenges to be faced along the way, such as logically connecting huge amounts of sensitive data, complying with recently introduced data protection legislation and ensuring that the individual is always in control of his or her data.

To illustrate the feasibility of our vision, the paper makes it more tangible by describing some of the concrete steps that TNO has already taken in this area. These include monitoring children's growth data to provide insights into their generic health and flag up potential anomalies, the early detection and diagnosis of chronic diseases and improving people's health through personal nutritional advice. Things that, with the aid of data science and AI, will deliver bespoke ways of keeping people healthy and lowering the cost of doing so.

TABLE OF CONTENTS

STAY HEALTHY, AND IN CONTROL OF YOUR DATA

Data science and artificial intelligence (AI) can do great things for healthcare. They have the potential to make the healthcare sector more efficient by helping both individuals and doctors to define what exactly is healthy, and for whom. And it has already started.

Data science and AI are currently used in many healthcare tools and solutions, such as AI-assisted robotic surgery, virtual nursing assistants, image analysis, workflow improvement tasks and several in-home monitoring devices. Data is currently seen as the 'new gold' and the growth of companies like Google, Apple and Facebook are the living proof of this. At TNO we believe data science and AI can play a crucial role transforming the healthcare sector and reducing costs. They can help us do the right things to stay as healthy as possible, and make the right choices when, or even before, we fall ill. This paper explores how data science and AI can transform the healthcare sector; but not without first pointing out a few of the most important challenges to be faced along the road to transformation. By way of illustration, it also provides a number of success stories so far.

ARTIFICIAL INTELLIGENCE (AI)

In computer science, AI research is defined as the study of "intelligent agents": any device that perceives its environment and takes actions that maximise its chance of successfully achieving its goals. Colloquially, the term "artificial intelligence" is applied when a machine mimics "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving". The traditional problems (or goals) of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception and the ability to move and manipulate objects.

DATA SCIENCE

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured similar to data mining.

RESPONSIBLE AI

A term used to establish which ethical backbone is considered in the models. For example, do we optimise for the best health outcome or the lowest costs?

EXPLAINABILITY

The outcomes of the AI models should be explainable, for two main reasons. Firstly, individuals and health providers will not trust a model if they do not know what its outcomes are based on. Secondly, who is responsible if its prediction is found to be wrong? Who takes the blame?

THE CHALLENGES POSED BY USING DATA IN THE HEALTHCARE SECTOR

There are currently four challenges that prevent the full implementation of data science into healthcare:

DERIVING VALUE FROM DATA

The first, and most important, challenge is that data in itself does not bring value. It's the interpretation and explanation of that data – along with the correct predictions – that deliver the real value. For this, expertise is required in the areas of biology and physiology, as well as the system understanding to effectively apply data science and AI.

A SHIFT OF FOCUS?

A second challenge is how to shift from sick care towards healthcare, whereby diseases are prevented or cured. The ultimate ambition is to combine various sources of information – such as a patients' clinical and personal real-world data, digital knowledge from literature and protocols – and use them in algorithms to retrieve new knowledge.
This may well shift the current focus, which is on the treatment of illnesses, towards the maintenance of good health. However, the sheer complexity of all this will be such that AI will be needed to reconcile all the parameters correctly. Moreover, explainability and responsibility of these models will be key and safeguarding the privacy of the data will be crucial.

---

data in itself does not bring value

SENSITIVE, BUT SENSIBLE

A further challenge lies in how we can logically connect massive amounts of highly sensitive data. The volume of health data we have been amassing has been increasing exponentially for about 20 years now, and for several reasons, two of which we will highlight. Firstly, ever-developing technology (the omics revolution) has made it possible to measure anything from hundreds to hundreds of thousands of parameters simultaneously. Secondly, patient data is becoming increasingly digital and thereby more reusable. And much of this data is collected by people themselves, through health monitors and other do-it-yourself devices. Traditionally, the healthcare sector has been a relatively data-poor environment, with data always stored in separate silos that were difficult to connect to one another. But during the past few years more connections have been made, particularly in enriching medical data with data from those newly available do-it-yourself devices, including activity monitors and cholesterol-measuring equipment. More recently, the General Data Protection Regulation (GDPR) has rightly erected a set of strict boundaries as to how data, including healthcare data, should be harvested and stored. Healthcare data is, by definition, sensitive information because it is about a specific individual. If we are to use this data to its full potential, while respecting the relevant privacy issues and requirements of the GDPR, models based on AI technology must be developed. These models will then help by offering benefits in several areas of healthcare, such as the diagnosis, prevention and better treatment of diseases.

CHAIN OF CONSENT

But yet another challenge posed by the use of healthcare data is that the chain of consent regarding that data currently leaves much to be desired. Given that health data is always personal and confidential, as we discussed in the previous paragraph we believe that the individual should always have the power to control the use of that data.

The healthcare sector has been a relatively data-poor environment, with data always stored in separate silos

POWER (OVER THEIR DATA) TO THE PEOPLE

CRUCIAL PREREQUISITES

The better news about improving the chain of consent is that access to health data can be organised by so-called Health Data Cooperatives or Communities (HDC), in which people themselves are the true controllers of their health data. But two very important prerequisites are needed for such an organisation. The first of these is strong governance, to support people in deciding about whether and how they should give other parties access to their health data. Patient communities, for example, could provide advice on the delegation of data sharing. Furthermore, the governance should be clear about commercial interests. The second prerequisite is that we must strongly encourage the development of new techniques to allow algorithms to self-learn from data, without the need to create additional copies of that data, and to allow algorithms to throughput encrypted data, so that privacy is safeguarded.

SOLUTIONS NEEDED

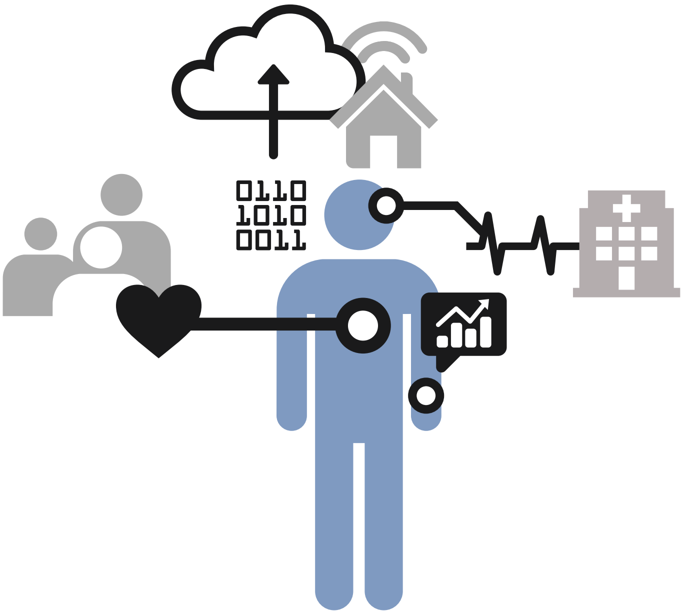
If the HDC is to arrange the relevant chain of consent, data must be stored securely and reliably. Health data is typically stored by the family doctor, hospital and/or the individual's smart phone or other device. This means that solutions must be found for the resulting complex models if they are to self-learn from all these data sources. Given that central storage will not be a feasible option (because the data has already been distributed and is thus not secure and not scalable), the HDC must facilitate decentralised personal data vaults that are controlled by the individuals. At TNO we are developing AI technology that has exactly these capabilities, i.e., secure machine-learning technology. One of the technologies comprises a toolbox of machine-learning algorithms that are integrated with cryptography, in other words secure multiparty computation (MPC). Another is distributed learning.

LARGER TNO INITIATIVES

The solutions alluded to in the previous paragraph are part of larger initiatives in which TNO actively participates, such as the Personal Health Train. A health train is a concept in which data stored in different locations (stations) is standardised so that 'trains' travelling between these stations are loaded with models to self-learn from the data. The main objective of a health train is to learn from health data and thus make it more usable. The trains connect different health data stations, in which the data is made FAIR (Findable, Accessible, Interoperable and Reusable) so that it can be linked. The 'tracks' between the data stations are the standards with which the models that process the data must comply. The models are loaded on the trains and then these models use the tracks to find the required data in the applicable stations. A train's access to a particular station is governed by the appropriate validation, which can include permissions given by an individual to share certain data, for example. The structure and details of this governance is defined in the HHDC (Holland Health Data Cooperatives).

It stands to reason then that HDCs, in combination with the Personal Health Train, could offer a significant degree of control over personal health data. It could also establish and fuel a personal FAIRtrade health data marketplace, allowing people to truly benefit from their data and thus stimulating a new health data economy. All this is far from being just pie in the sky. The conceptual premise for it is already in place; all that's needed is for the infrastructure to be set up and implemented.

MAKING IT REAL: EXAMPLES OF HOW DATA SCIENCE CAN BE USED IN DIAGNOSIS AND TREATMENT



IN THE YOUTH HEALTH DOMAIN

Monitoring child growth is an important indicator of a healthy development. In the youth health domain it is used to detect abnormal growth patterns, which, in turn, may reflect nutritional problems or, even worse, an underlying illness. In the context of the larger population, the data that is generated could provide insights into the generic health of this population segment, such as the risk of being overweight or obese, or average height, for example. TNO has carried out large-scale national growth studies in children every 10 to 15 years since 1955 and the results have been translated into growth charts that are used to the benefit of juvenile healthcare. These growth studies also allow comparisons to be made of children's height and weight over time. The integration of electronic health records using AI models will provide personalised and accurate predictions and referral criteria, and will thus be of immense value for healthcare professionals. TNO has also applied curve-matching methodology on growth. This data-driven technique can be used to improve the individual prediction of childhood growth. The idea is to match information of the current child with data from existing children in existing databases. The growth patterns of the matched children predict how the current child is likely to develop in the future. These new ways of integrating platforms, re-using data, and personalised and accurate advice will support professionals in making evidence-based decisions.

IN THE EARLY DETECTION OF CHRONIC DISEASES

Another example of how data science can be used in healthcare is the diagnosis and subsequent treatment of liver disease. Non Alcoholic Fatty Liver Disease (NAFLD) has become the most common chronic liver disease in developed countries and its increase is closely associated with the incidence of obesity. NAFLD can be associated with hepatocellular damage and inflammation and is then called Non-Alcoholic Steatohepatitis (NASH). Currently, in the US, for example, there are no FDA-approved drugs for NASH on the market. This is mainly due to the lack of knowledge regarding the key molecular processes, the lack of translational pre-clinical models and the non-availability of blood-based biomarkers.

At TNO, by using Systems Biology approaches and machine-learning (ML) techniques, we are developing models to help with the early detection and more efficient treatment of NASH. Systems Biology integrates datasets from multi-omics analyses platforms to obtain relevant biological pathways, which contribute to the mechanisms' underlying disease development. The applied pattern recognition tools have led to the identification of molecular patterns in liver transcriptome that correlate with functional proteomics readouts for collagen synthesis, as determined by a D2O-labelled detection method. An early molecular signature is described which predicts the development of NASH and fibrosis before pathology becomes manifest. This contributes to the more efficient development of personalised treatments for NASH and fibrosis.

WATCH THIS SPACE!

Another tool that TNO has specifically developed to handle multi-omics data is the health space. The technology on which this tool is based is regularised regression and it projects the high dimensional feature space to a single dimension, representing biological processes (such as lipid, carbohydrate and protein metabolism) or those in the health domain (such as muscle health, metabolic age, inflammation). This single dimension can then be reported as a user-interpretable value, such as your score for muscle health being 7 out of 10, for example. An individual's position in the health space can be interpreted as an indication of his or her health status, and interventions can be based on the movements of others in the same space.

BOOSTING HEALTH THROUGH PERSONAL NUTRITION ADVICE

Personalised insights into health are being made possible by integrating data, smart algorithms and expert knowledge in TNO's Personalised Health Toolbox. We are all unique individuals, and this also applies to how our bodies react to food and nutrition. To find out what kind of food and nutrition works best for us, TNO has developed an automated expert system that converts personal knowledge into nutritional advice. Habit, a US company that specialises in personalised nutrition, is the first business to apply such an automated knowledge approach commercially. Using a systems biological approach, data on general characteristics, food preferences and personal health goals are combined with the output of DNA testing, and blood analysis after drinking a liquid meal. The personal data generated is analysed by algorithms devised by TNO and this culminates in a personal health report and balanced personal nutritional advice. The outcomes of these algorithms connects to Habit client services, such as online coaching, user apps, personal recipes, and shopping lists.

Current algorithms could be further improved if data collected from all participants could be re-used, such as data from a challenge test, for example. This would be done by giving participants a metabolic trigger in the form of a liquid meal and then analysing the response beforehand and then after a while by doing finger-prick blood tests. On the strength of the ensuing data, adapted AI models could be developed that can sub-type participants on the basis of their responses. Then, depending on the applicable sub-type, different nutritional or health-related advice could be given. Bayesian networks, based on expert knowledge, would be used to generate the relevant recommendations. These models would have the ability to learn from data that is harvested during the field deployment of said systems, essentially leading to a self-learning system.

THE FUTURE OF DATA SCIENCE IN THE HEALTHCARE DOMAIN

TRUST IN PRIVACY WILL BE CRUCIAL

Data science is still a relatively new discipline, which is why it is quite difficult to predict what to expect from it. A few generic trends in data science that are specific to the healthcare sector are very much under discussion. Mainly, these trends are related to the fact that privacy and security play such important roles in the field of health. ‘Privacy by design’ is set to become increasingly important, not just because of regulations and welcome legislation such as the GDPR, but because trust in privacy will be the main reason that people will be prepared to trust AI-derived advice.

A distributed approach to storing data will also become very important. Privacy in this respect translates to an individual having control over what happens to his or her data. Distributed storage – where the individual also controls what information is stored, where, and for how long – is more privacy-friendly than storing data in one place and then having to copy it every time a different organisation needs it. This form of decentralisation also has other distinct advantages. For starters, copying data is always accompanied by the risk of loss or damage and it can be very expensive. Furthermore, decentralised data is more likely to be retained. Current privacy legislation stipulates that medical data must be automatically destroyed after a certain time. This makes it difficult to retain the medical data of young children in the long term, while that is the very data that can be so useful for future research.

TNO'S FUTURE PERSPECTIVE OF DATA SCIENCE

This paper has discussed the challenges that need to be tackled if data science is to be successfully applied in the healthcare sector and become the much-needed game changer in scientific studies and the prevention of diseases. As the examples have shown, thanks to data science we can analyse datasets and produce bespoke ways of keeping people healthy, thereby reducing health risks, extending life expectancy, and lowering the cost of healthcare. However, as with any new technology, it throws up new challenges. First and foremost are the understanding and interpretation of the analyses results generated by data science, and how these results can be applied in the most meaningful ways. Then there are the privacy obstacles, which can be surmounted by a decentralised approach to data and by giving people more control over what they share. At TNO we are confident that as soon as these challenges are overcome, data science will transform the healthcare sector as we know it.

thanks to data science we can analyse datasets and produce bespoke ways of keeping people healthy

SIDEBAR

TNO is committed to playing a relevant role in society in ensuring that health-related data is used for the benefit of the people it pertains to. To this end we are researching how best to organise it all and already partly implementing, working together at every opportunity with scientific organisations governments and industry alike to establish a national and international code of conduct to deal with it. All this involves initiatives, projects and knowledge that place health data under the control of its owners, rather than non-healthcare players, such as tech companies and health insurance providers, as is currently the case. It requires different business models to those that are currently available, models such as FAIRtrade Health Data, for example. This is an initiative of TNO to ensure that in addition to their rights guaranteed by the GDPR, citizens are given safe and reliable access to their health data and the ability to control its use.

LITERATURE:

(1) https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy
(2) Data for the People, how to make our post-privacy economy work for you. Andreas Weigend, Basic Books, New York.
(3) https://www.futureagenda.org/news/future-of-patient-data-global-report
(4) Wahl, B., Cossy-Gantner, A., Germann, S., & Schwalbe, N. R. (2018). Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Global Health, 3(4), e000798
(5) For a visual explanation of the health train see: https://vimeo.com/143245835
(6) For an explanation on Multi Party Computation (MPC) see: https://www.tno.nl/en/focus-areas/information-communication-technology/roadmaps/trusted-ict/secure-multi-party-computation-jointly-analysing-sensitive-data-without-sharing-it/

CONTACT

TNO
Schipholweg 77
2316 ZL Leiden
Postbus 3005
2301 DA Leiden
088 866 90 00

Dr. J. (Jildau) Bouwman
Senior scientist systems biology
Microbiology and Systems Biology
jildau.bouwman@tno.nl

Ir. C.J.E. (Carla) Rombouts-Gordijn
Business Development
Digital Health
carla.rombouts@tno.nl

TNO innovation for life

TNO.NL

18-10432