Responsible Innovation in Data-Driven Biotechnology

Bruynseels, K.R.C.

DOI
10.4233/uuid:a1b062bb-876e-4271-9c00-db3c8aa866dd

Publication date
2021

Document Version
Final published version

Citation (APA)
Bruynseels, K. R. C. (2021). *Responsible Innovation in Data-Driven Biotechnology*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:a1b062bb-876e-4271-9c00-db3c8aa866dd

Responsible Innovation in Data-Driven Biotechnology

Responsible Innovation in Data-Driven Biotechnology

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op donderdag 29 april 2021 om 15:00 uur

door Koen BRUYNSEELS
Doctor in de Medische Wetenschappen,
Ingenieur voor de Scheikunde en de Landbouwindustrieën,
Bachelor in de Filosofie,
Katholieke Universiteit Leuven, België

geboren te Lier, België

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. M.J. van den Hoven

Samenstelling promotiecommissie

Rector Magnificus, Technische Universiteit Delft, voorzitter
Prof. dr. M.J. van den Hoven, Technische Universiteit Delft, promotor

Onafhankelijke leden:
Prof. dr. P. Osseweijer, Technische Universiteit Delft
Prof. dr. M. Boenink, Radboud Universiteit
Prof. dr. A. Zwitter, Rijksuniversiteit Groningen
Prof. dr. P.M. Macnaghten, Wageningen University & Research
Prof. JD B.M. Frischmann, Villanova University Charles Widger, U.S.A.
Prof. dr. B. Taebi, Technische Universiteit Delft, reservelid

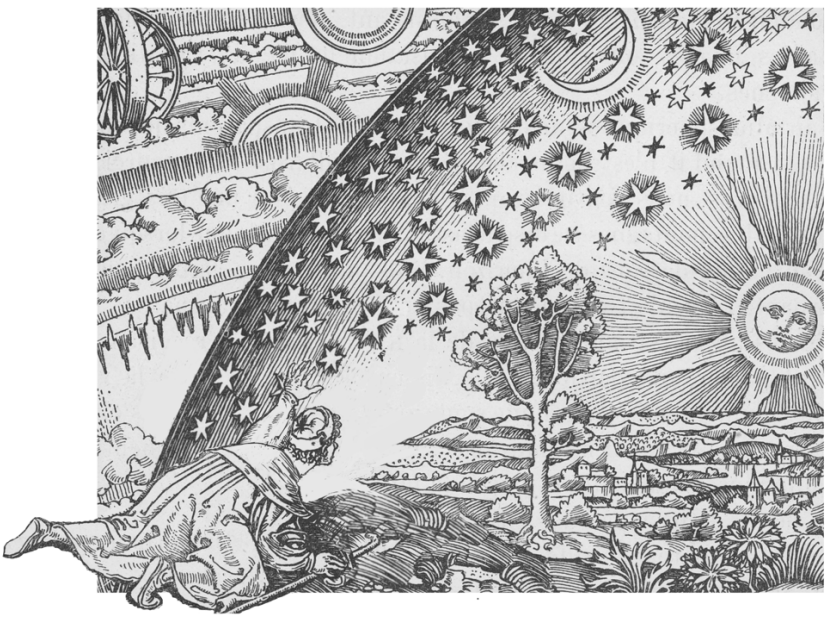
Overig lid:
Dr. F. Santoni de Sio, Technische Universiteit Delft

Research for this thesis was made possible by the 4TU.Ethics.

editors: Philip Brey, Anthonie Meijers, Sabine Roeser and Marcel Verweij

ISBN: 978-90-386-5265-8
ISSN: 1574-941X

Copies of this publication may be ordered from the 4TU.Centre for Ethics and Technology, info@ethicsandtechnology.eu
For more information, see http://www.ethicsandtechnology.eu



The Flammarion engraving, revisited

Contents

List of papers

Chapter 2

Bruynseels, KRC., & van den Hoven, MJ. (2015). How to do things with personal big biodata. In B. Roessler, & D. Mokrosinska (Eds.), Social dimensions of privacy: interdisciplinary perspectives (pp. 122-140). Cambridge University Press.
DOI: 10.1017/CBO9781107280557.008

Chapter 3

Bruynseels, K., Santoni de Sio, F., and van den Hoven, J. (2018) 'Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm', in: Frontiers in Genetics' 9.
DOI: 10.3389/fgene.2018.00031

Chapter 4

Bruynseels, K. (2020) 'When nature goes digital: routes for responsible innovation', in: Journal of Responsible Innovation, 7:3, 342-360.
DOI: 10.1080/23299460.2020.1771144

Chapter 5

Bruynseels, K. (2020) 'Responsible Innovation in Synthetic Biology in response to COVID-19: The role of data positionality' in: Ethics Inf. Technol.
DOI: 10.1007/s10676-020-09565-9

Jeroen van den Hoven and Filippo Santoni de Sio are acknowledged for granting permission to publish the co-authored papers in this dissertation.

Acknowledgements

> "It is owing to wonder that men both now begin, and at first began, to philosophize. They wondered … about the phenomena of the moon and those of the sun and the stars, and about the origin of the universe. A man who is puzzled and wonders thinks himself ignorant."
>
> Aristotle, Metaphysics 982 b 12 ff

Wonder best describes my state of mind when I entered the interview room in Den Haag. One moment I was sitting in a train, the next moment I found myself in a small conference room, at the top floor of Den Haag's central station tower building, facing a U-shaped table full of philosophy professors. This was my first real encounter with the – by then – 3TU.Ethics. I was incredibly happy when I received an email later on, stating that the committee retained me as a Ph.D. candidate. Philosophy was love at first sight when I first heard about it in secondary school. And philosophy of technology even more so, after discovering a copy of 'Die Technik und die Kehre' in the local libary of my birth-town. Now I got the opportunity to get in touch with real philosophers of technology, and to dive deep into the topic. In The Netherlands, a global hotspot of technology philosophers and -ethicists.

I am therefore absolutely grateful to my promotor Jeroen van den Hoven to have provided me with this opportunity. I want to thank him for his steady support over these years, the academic freedom he provided to shape the topic of the thesis, the many bright ideas and stimulating discussions, and for showing the positive impact philosophy can have on society. It probably took quite some patience to get me into 'the Delft way' of thinking – only carry with you what is necessary – and I'm grateful he persisted. I also express great gratitude towards my daily supervisor, Filippo Santoni De Sio, for his enthusiastic and diligent support, and his many thoughtful contributions. He was incredibly helpful in shaping efficient academic habits, as in dealing with difficult reviewer's comments, or in finding ways to approach the writing. I also want to thank David Koepsell, who was my daily supervisor in the first years. He introduced me in the ethical aspects of genetics and testified to the fact that philosophy is not incompatible with entrepreneurship. A big thanks also to the academics from the 'Sectie Filosofie' at T.U.Delft, to my fellow Ph.D. students and to the 4TU.Ethics, for providing such a vibrant intellectual environment. Diana Droog I'd like to thank for her kind support in all things practical over the years.

It was a privilege to get the opportunity to combine my work as a scientist in industry, with this dissertation work in the philosophy and ethics of technology. This allowed me to experience the amazing evolution in biotechnology from the inside, while acquiring the conceptual means to reflect on it. I'd like to thank my BASF colleagues Gerhard Ritte and Christophe Reuzeau, and the promotor of my Ph.D. in Biomedical Sciences, Florent Vanstapel, for their helpful feedback on parts of the manuscript; Paul Van Tongeren and the Hoger Instituut Wijsbegeerte, for introducing me into philosophy; and Karin, Richard, and my peers and co-workers at BASF for providing such a splendid scientific environment.

To be able to wonder requires a wonderful environment. This work would not have been possible without the support of my family and friends. My friends I'd like to thank for their support and for pointing out an incredible richness of ways to know and to approach this magnificent world. My father, mother and brother I'd like to thank for their love, and for their encouragement and support in all my endeavors in life. My three daughters and muses, Helena, Esther and Hannah, I thank for their love and inspiration, for showing me the splendor and depth of things that matter, and the importance of doing things right. To my wife Kaat, without your love, patience and support this would not have been possible. It's a blessing to have you as my companion on this wonderful path.

1 Introduction

1.1. Data-driven innovation as ethical challenge

For most of history, human endeavors by no means measured up against the forces of Nature. Man could construct ships to defy the billowy elements … but the sea itself remained indifferent[1]. This situation fundamentally changed. The avalanche of technical innovations now is a force that shapes the natural environment, society and our very self. Moreover, innovation has become an imperative in its own right. In the Schumpeterian view, innovation is "the economic activity of producing and using new knowledge and ideas about sources of value that results in the disruption and restructuring of the economic order" (Potts, 2018). This intertwinement of innovation and economic gains lead to a pursuit of innovation as a good in itself. A recent strategic note of the European Political Strategy Center indicatively advocates an Innovation Principle, which is "the positive obligation to facilitate innovation" (European Political Strategy Center, 2016).

This innovation dynamics is further intensified by the availability of vast amounts of data about the natural world. Data about human beings, ecosystems, social behaviors, nanomolecular structures, etc. provide a fertile ground for human ingenuity. This effect is pronounced in the case of biological data. High-throughput data capturing technologies led to an explosive growth of the amount of biological data over the last few decades. DNA sequencing costs plummeted, allowing to chart whole personal genomes of large human populations, and the genetic diversity in plants, animals, and microbial populations in a wide range of biotopes. Large scale read-out of multiple aspects of biology gives rise to data on genomes (Stephens, et al., 2015), gene expression behavior, proteins, metabolites, macroscopic traits, and other aspects of biology. Sensor and imaging technologies allow for dynamics monitoring of biological systems, as in wearables that monitor blood pressure or sugar levels, or in probes that measure soil moisture levels in crop fields. Satellite images provide data about entire ecosystems. Together, these bio-digitalization technologies result in large quantities of biodata that can reveal a lot about ourselves as human beings, and about the natural world that surrounds us.

---

[1] Jonas, Hans, Das Prinzip Verantwortung, Versuch einer Etik für die technologische Zivilisation, Frankfurt am Main, Insel Verlag, 1979, p 18-19

This digitalization of biological systems unlocks a significant natural capital. It gives access to the billions of years of evolutionary testing and tinkering that is embedded in living creatures. And observing Nature always has sparked innovation. Wilbur Wright's observation of bird flight for instance eventually led to the aviation industry. He noticed how pigeons control their flight by modifying the airflow over their wings, and mimicked it in the double-winged design of the first airplane. Another prominent example is scientific plant breeding, which started with the observations of Gregor Mendel on how pea plant characteristics were inherited by hybrid offspring. Hybrid crop plants later on enabled the green revolution, changing the face of agriculture and the planet. Biology-based innovation in general is now vastly empowered by the ability to read out biological systems. For example, biodiversity sequence data provides a rich source of inspiration for drug development. About half of the FDA-approved drugs that are on the market are derived directly or indirectly from natural products (Katz, 2011). Another example is human personal genomics data, which provides the foundation for personalized medicine. The Genomics England project for instance is sequencing 100.000 inhabitants that are related to cancer or rare diseases (Marx, 2015), and aim at further broadening this approach to 5 million U.K. citizens. By relating genomic data to electronic health records, this initiative is meant to provide the foundation for a personalized medicine and a flourishing innovation landscape. Synthetic Biology is an example par excellence of bio-based innovation, being a bio-engineering field that pursues data-driven design of biological systems (Endy, 2005). It heavily builds on molecular biology and laboratory process automation, but also on algorithms and modeling techniques that can utilize large amounts of biological data in the design process (Freemont, 2019). Synthetic biology can lead new routes for drug development, increase the ability to rapidly respond to pandemics, lead to new materials that are stronger, lighter or more biodegradable, or new enzymes that can help in the generation of biofuels out of renewable feedstocks (Synthetic Biology Leadership Council, 2016). Innovations derived from synthetic biology are expected to contribute to the development of a new bioeconomy – "an economic model where knowledge-based utilization of biological resources and processes can be applied to the sustainable production and manufacturing of goods, and the provision of services across all economic sectors" (Freemont, 2019). The fostering of an innovation landscape around synthetic biology is explicitly targeted in for instance the U.S. (Si & Zhao, 2016), the U.K. (Synthetic Biology Leadership Council, 2016), France (Meyer, 2013), and Finland (Living Factories, 2017). Innovations in synthetic biology can potentially address pressing

questions related to global warming, preservation of ecosystems, sustainable agriculture, and affordable health care.

When technical innovation is left up to the market dynamics, the creative destruction might lead to better products (Schumpeter, 1943), but can also lead to devastating 'externalities'. Grand challenges such as global warming (Keeling, et al., 1976), the deterioration of the natural environment and the massive loss of biodiversity (Nobre, et al., 2016) (Johnson, et al., 2017) are anthropogenic. The sheer size of the accumulated impact of technological innovations and their application even warrants the definition of a new geological era: the Anthropocene (Lewis & Maslin, 2015), testifying to their impact. Negative effects of technical innovations most often are distributed unequally, disproportionally affecting the natural environment and parties that had no say in the development of the technologies (Jasanoff, 2016).

This strongly calls for a moral accompaniment of technical innovation. Such accompaniment is needed, since technoscience itself has an inherent openness to all that is technically feasible (Hottois, 1988). The 'technological imperative' entails a value-free exploration of the vast space of possibilities. It states that "One should try everything that is possible, execute all possible experiments, all manipulations, one should actualise all possibilities, develop all powers, all potentials of existence: of the matter, the living, the thinking" (Hottois, 1988). This 'impératif technicien' is a 'impératif de la liberté' according to Hottois. It is the radical freedom of technical tinkering that accepts no boundaries and can only be constrained by the physical resistance of reality itself. It is the attitude of genome-editing researcher Rebrikov when answering the question whether one should wait with clinical research on edited human embryos until the international frameworks are in place: "Are you serious? Where did you see the researcher willing to slow down?" (Cyranoski, 2019).

Since the technological imperative is a-ethical and a-human, it needs to be counterbalanced with an ethical imperative. One can define a 'moral innovation imperative' that demands to "bring[ing] about changes in the world so that we can fulfill more of our obligations towards the fellow human beings, the environment, life on the planet, and future generations." (van den Hoven J. , 2014). According to this imperative, innovation provides us a means to better fulfill our moral obligations. A moral perspective on innovation was already implicit at the very start of the technoscientific endeavor. Bacon's manifesto New Atlantis planted the potential of a morally guided innovation at the heart of technoscience (Bacon, 1626). In Salomon's House - Bacon's vision of research institute - technology development entailed an active pursuit for novel applications based on the insights derived from experiment

data, a pursuit which is oriented towards generating benefits for humankind. It was recently argued that responsibility and innovation are intrinsically related at multiple levels. Responsibility drives innovation by calling not to do whatever is technically possible or economically interesting, but to innovate for the better. And the products of innovations constitute means that help us to fulfill our moral obligations (Bergen, 2017). Responsible Innovation (RI) was proposed as a way to align technological innovation with values held by society. By including social and ethical aspects in the innovation process, RI provides a concrete approach for moral accompaniment of technoscientific developments. Responsible Innovation has been defined as "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)" (Von Schomberg, 2011). A broader definition brings in the time perspective in the technological evolution: "taking care of the future through collective stewardship of science and innovation in the present" (Stilgoe, Owen, & Macnaghten, 2013).

In order to be able to innovate responsibly, one needs to be 'response-able': able to respond to technical possibilities that are opening up, and able to shape technologies in response to societal needs and values. This ability is often impaired by multiple factors. The technological innovation process is often the province of experts and is disconnected from the needs and fears that live in society. Economic dynamics often drive technical innovation, thriving on asymmetries in information and power (Jasanoff, 2016). Technological innovation trajectories are mainly steered by capital and industry. And the masses of people that are involved in the industrial production, the consumption and that undergo the effects of new technologies have little to no say in the innovation process (Jasanoff 2016). Risks assessments and alignment with the bigger societal needs are happening when the technological innovation already materialized. These activities are merely a post-factum check whether already developed technologies are safe and whether they adhere to societal norms, rather than raising the fundamental question whether a technology or application should be brought into existence in the first place. Moreover, technologies can become entrenched, thereby making it hard to establish alternative routes (Collingridge, 1980). Especially in fast-paced fields that are empowered by massive amounts of data, driven by a high-tech market logic, and interweaved with multiple sectors in society, this dynamics can asymptotically approach a technological determinism – i.e. a situation in which there is little control over which future the

proliferating innovations will bring about. These questions especially hold true for bio-based innovations. Fields like biotechnology, health care and agriculture do experience an amazing ramp-up of innovations, often with a disruptive character. This speed, together with the societal importance of these sectors, calls for a responsible guidance of these innovations.

1.2. Biodata provides an entry point for responsible innovation

The core research hypothesis in this dissertation is that digitalization of biological systems provides a pivotal point for a responsible guidance of innovation.

When following the "moral innovation imperative" (van den Hoven J. , 2014), digitalizing biological systems offers novel ways to better fulfill our moral obligations. It makes the natural world deeply accessible to our human conceptual toolset, and thus to our creative attempts to improve the state of affairs for the natural environment and for our fellow human beings. Broad and open availability of scientific data in general accelerates innovation (Burgelman, et al., 2019). Access to biodata and data analysis capabilities strengthens the ability to adhere to the technological imperative, as well as to the moral innovation imperative. Data access can enhance the ability to more efficiently explore the vast space of biotechnological possibilities, and eventually let this exploration be guided by societal values. It does so by allowing for *in silico* mining and *in silico* conceptual manipulations. These computational activities provide guidance on which of the many engineering options are to be tested in the biophysical reality. For example, computer-aided design is a key element in synthetic biology design cycles. In such cycles, microorganisms are engineered to produce certain compounds or proteins (Paddon & Keasling, 2014) (Synthetic Biology Leadership Council, 2016) (Freemont, 2019). This way of designing bio-based products builds on the availability of biodiversity sequence data (amongst other data types) which provides a rich source for the identification of natural compounds (Lewin, et al., 2018) as well as for the identification of the enzymes and biochemical pathways that can produce them.

This increase in innovation capabilities does not automatically result though in a better fulfillment of our moral obligations. A responsible innovation requires the active alignment with values in the course of the innovation process. Moral accompaniment of biodata-driven innovation in this respect is special, since biodata-based innovation is not just about data. The data is tightly linked to the organisms from which they were extracted – it concerns data that are embedded in our bodies and in

our natural environment, data that become available for our human conceptual toolset, our aspirations and values, and our creativity. An ethical analysis therefore should not focus on the data aspects alone, but include the biological realities that underpin the biodata, to ensure that the analysis is not reduced to a mere data problem. Next to this, biodata is also tightly linked to our world of language, concepts, values, and innovation activities. It reveals patterns that are embedded in biological systems and makes them accessible for our moral categorizations. Biodata thus reside at the interface between the biophysical reality and our world of language and meaning. In the ontology put forward by Hottois, there is a fundamental difference between this 'naturalistic' and 'symbolic' realm (Hottois, 1995). The naturalistic realm of biophysical processes is characterized as 'operational'. This is the world of physical, chemical and biological processes that bring about or produce reality. Evolutionary processes provide the example par excellence, but according to Hottois, also technoscientific evolutions belong to this category. The symbolic realm on the other hand is the area of language and meaning, of culture and values. Here, thinking and deliberation attempt to get a conceptual grip on reality, as in scientific theories or in common language. In this view, a moral accompaniment of technoscientific innovations thus implies the interaction between two ontologically heterologous realms. This ontological framing helps in clarifying the unique position of biodata. Biodata are at the interface between the physicalist and the symbolic realm. They are extracted from biomolecules, and available for conceptual interpretation. And they are instrumental in translating this conceptual work into concrete interventions in biophysical systems, as for instance in bio-engineering. This special ontological and epistemological position has important consequences for a biodata-based responsible innovation.

Firstly, some of our very values that are at play during innovation are shaped at this interface. Patterns found in biodata can result in distinctions with significant moral loads. For instance, existing notions of health and disease become problematic when having fine grained data about a person's genetic makeup and physiological status. Likewise, the notions of what is natural and what is artificial become problematic when assessing this at the level of biomolecular data. These moral effects of biodata need to be taken into account when attempting responsible innovation. It calls for an analysis of how values and moral categories are impacted by the availability of biodata.

Secondly, the availability of biodata fundamentally biological systems into resources for innovation activities. It thereby fundamentally changes our relation to

the biological systems from which these data were extracted. For instance, biodiversity sequence data extraction can transform a biotope into a resource for the development of novel pharmaceutical compounds. Or personal genomics data can transform you into a resource for health care innovations. When pursuing responsible innovation, this calls for an analysis of how innovation can be organized in just and sustainable ways. Such ways of organizing innovation should foster a fair distribution of innovation opportunities and of the resulting benefits. They should also contribute to the flourishing and sustainability of the underpinning biological reality (the ecosystems or the populations). Question here is whether and how socio-technical systems can be designed to support distributive justice with respect to innovation opportunities and benefits, and how the design of such environments can support innovations that are aligned with societally preferred values.

In summary, the digitalization of biological systems impacts how a responsible guidance of biodata-based innovations can be pursued. Firstly, biological data are a locus where values are shaped. And secondly, biological data are a locus for organizing responsible innovation, including a fair distribution of the opportunities and benefits. Both aspects are the topic of the analysis in this dissertation.

1.3. Biological data as a locus where values are shaped

Alignment of biodata-driven innovations with societally preferred values is not one-way traffic. Biological data themselves often are a substrate for salient moral distinctions. This is particularly pronounced when it concerns human beings as the data object, as explored in Chapter 2 in the case of personal genomics data, and in Chapter 3 by using the concept of biological digital twins.

As explored in Chapter 1, raw biodata themselves are not intelligible. The data are highly complex, have a stochastic character and contain errors and noise. Raw data needs processing and interpretation in order to result in distinctions that can be used in ordinary human language. These linguistic representations are influenced by various decisions: which thresholds to take, which control datasets to compare against, which algorithms to use. Interpretations moreover can suffer from apophenia, the human tendency to see patterns where there actually are none. But the distinctions one makes do matter. Personal biological data are prone to be value laden, since they can reveal aspects of a person's physical, emotional and mental predispositions. For example, if my genomic data indicates an increased risk to develop a mental disease, this knowledge might affect how others perceive me, how I

perceive myself, the opportunities I can pursue in society, the fee I need to pay for my health insurance, etc. In the process of analysis and interpretation, patterns in the biodata are converted into statements that are part of our world of language and meaning, and are therefore value laden. The fact that biodata is tightly related to the bodies of the individuals adds to the value-laden character of these statements. Personal biodata are related to the body, to family ties, to otherwise hidden traits. For instance, if careful comparison of my genomic data reveals that I have close family ties with a certain person, this probably will change my mental attitude towards that person. Biological data can also be value-laden outside of the realm of human biology. In plant breeding for example, Crispr-CAS genome editing and conventional random mutagenesis both result in nucleotide changes in the plant's genome, raising the question what should be considered 'natural' and what 'unnatural'. These discussions often revolve around the question whether such distinction can be meaningfully made at the information level.

Responsible innovation for these reasons also implies modesty in the epistemic claims that are made based on biodata. Epistemic modesty is proposed as a guiding ethical principle in dealing with the moral loads that can be assigned to biodata (Chapter 2). The principle implies the avoidance of making strong moral claims based on biodata about a subject, taking the position that there is always more to say about a being than can be deduced from derived data. The principle can be illustrated most saliently in the case of human data. A person will not regard herself as a mere collection of data. A person pursues moral autonomy (van den Hoven J. , 2008), along the lines of Bernard William's proposition that respect for persons implies attempts at moral identification. This is an important aspect of what it means to respect someone: we owe the other an attempt at identification (not clinical or forensic) as the person who she wants to be identified with. In other words, as someone who is engaging in self-identification, who is using her freedom to determine for herself how she sees herself. This implies that others appreciate her as someone who is engaging in self-identification – rather than as a collection of data. Respect for a person thus implies modesty about what one knows when one has access to a person's biodata.

Responsible innovation for these reasons also implies the acknowledgement that availability of biodata can impact existing moral distinctions, and taking an explicit stance with respect to these moral distinctions that are impacted by the biodata. Digital Twins are proposed as a conceptual tool to make this aspect tangible. Digital Twins are an emerging practice in engineering, in which computer models are built

that closely reflect individual artifacts. This tight interaction between the digital representation and the artifact makes it possible to increase predictiveness, up to the level that predictive maintenance becomes possible. It is adapted by engineering companies to maintain their machine park via predictive maintenance, a practice in which the digital twin of a machine allows to predict which parts are about to break down and should be replaced. This engineering practice is emerging now also in human medicine. Digital twins of the hearts of individual patients were for instance constructed to study effects of surgical interventions *in silico*. The computer models are based on biodata from many patients, but parameterized to the heart of each particular patient, by factoring in Magnetic Resonance Imaging, Computed Tomography scans and other measurements on that patient's heart. The heart model is also linked with the patient's heart via sensors, so that it continuously reflects its actual status and dynamics. In the context of this dissertation, biological digital twins are taken as a conceptual tool rather than as a claim about comprehensive representations of biological systems (the complexity of biological systems simply defies such claims). The digital twins engineering concept makes explicit what the consequences are of the tight relation between the biophysical reality (people, biotopes), extracted biodata, and our world of language and meaning. For example, how health, therapy, preventative care, and human enhancement are distinguished will be increasingly driven by the usage of human biological data, as well as by our ways to assign meaning to them (Chapter 3). High resolution data of individuals functions as a magnifying glass for the existing moral distinctions, and potentially triggers shifts in these moral distinctions. For instance, it can challenge where intensive medical follow-up (as is done with athletes) ends, and where human enhancement begins, or it can challenge when exactly interventions transgress existing moral boundaries.

Biodata-driven innovations thus happen at the locus where values need to be deliberated and determined, in confrontation with the patterns identified in Nature. The perspective taken in this dissertation builds on approaches in literature that focus on biodata as a mere data issue, but explicitly takes the tight relation with the biophysical reality that underpins it into account. Responsible innovation – the steering of innovations towards societally preferred directions – when involving biodata, takes place on this interface between the natural world and culture. Biodata provides the resource for serendipity and shows what is possible. Whether possibilities are pursued or not is driven by the scientific interest to explore, the economic interest to open valuable markets, but also by what we value and deem worthwhile.

Any of these decision rationales requires taking a moral stance towards the patterns found in the biodata.

1.4. Biological data as a locus for organizing responsible innovation

Biological systems undergo a fundamental transformation when being digitalized: they are transformed into resources or goods. The biological system becomes a source of information goods that provide the basis for economically and/or societally valuable innovations. And the biological system itself also transforms into a good, a good that can be accessed and put to use for human wants and needs in very different ways. One can for instance derive novel pharmaceuticals based on the biodiversity data from a biotope, instead of using it for agricultural purposes.

This transformation of biological systems into goods has consequences for responsible innovation approaches. It raises the question about fairness and sustainability. How to obtain a fair distribution of the ability to participate in innovation, as well as a fair distribution of the benefits that result from innovations? And how to broaden the range of beneficiaries to also include the human populations or the ecosystems themselves, so that there is an inherent focus on the benefits for society or on the sustainability of the ecosystems? RI approaches often focus on governance to shape academic research and development activities, or on the question how individual innovation processes should be shaped in order to occur responsibly. This perspective is broadened here, to allow for an assessment of how innovation environments around biodata can be shaped (be it academic, corporate, or a blend of both) so that responsibility is an inherent property of them.

Commons are used as a model, since digitalized biological resources often have characteristics of a 'common pool resource': they are often managed and used by multiple parties, and are therefore prone to social dilemmas. Hess and Ostrom (Hess & Ostrom, 2006) argued in general that:

> New technologies can enable the capture of what are once free and open public goods. This has been the case with the development of most "global commons", such as deep seas, the atmosphere, the electromagnetic spectrum, and space, for example. The ability to capture the previously uncapturable creates a fundamental change in the nature of the resource, with the resource being converted from a nonrivalrous nonexclusionary public good into a common-pool resource that needs to be managed, monitored, and protected, to ensure sustainability and preservation.

Common pool resources are susceptible to over-exploitation and under-support. One way of dealing with these social dilemmas is collective self-regulation. This happens in a "common": communities and their institutions that govern the use of the common pool resource. In Chapter 4, twin commons are proposed as a framework to organize responsible innovation in biodata-centered communities. Twin commons build on the natural resource commons framework developed by Ostrom et al (Ostrom, 1990), the knowledge commons (Strandburg, Frischmann, & Madison, 2017), and the innovation commons (Potts, 2018). Core to the twin common is the twin resource: the biological system and its digital representation. The innovation dynamic in a twin common is dependent on the *in silico* data, but also on the community of stakeholders and how they interact, on the biological system, on the technologies for interacting with the biological system, and on the innovation resources (including the tacit knowledge) that are available to the community.

Responsibility is foundational and motivational in the governance of common pool resources. It provides an antidote for the rational self-interest of individual stakeholders and thereby avoids a tragedy of the commons. Self-organization of collective ownership can counteract negative outcomes of social dilemmas, thereby allowing for a sustainable management of both the biological resource and the derived biodata. One of the rules for successful commons (Ostrom, 1990) consists of the monitoring and sanctioning of irresponsible behavior towards the commons or towards the community of stakeholders (Poteete, 2010). This dynamic is empirically illustrated in the "Ultimatum game" (Nowak, Page and Sigmund 2000). The Ultimatum Game is an economic experiment in which a player gets a sum of money and has to decide on how much of this sum to give to a second player. If the second player rejects the offer, none of the players receive money. If the offer is accepted, then the proposed amount of money is transferred to the second player. The empirical outcomes of the game vastly deviate from the ones predicted by this classical economic theory. If the proposed amount is rated to be unfair, the deal is very often rejected by the other party, even if this means losing out on a small amount of money. These types of experiments indicate that values like fairness, and related social rules like reciprocity and equality of opportunity, play a vast (an important) role in social interactions. Twin commons provide a framework to identify responsibilities and values that are at play in digitalized biological systems. Such clarification of values is instrumental when fostering responsible innovation. Moreover, where

necessary and feasible, values then can be embedded in the socio-technical system that supports the twin common.

Biodata-based responsible innovation requires access to biodata. In this dissertation it is argued that biodata are inherently positional (Chapter 5), which means that there is a tendency of information not to be free, but to be subject to 'artifical scarcity' (Zinnbauer 2018). In research settings, biodata have often been framed as global public goods (Chadwick & Wilson, 2004). In practice though, multiple hurdles can be identified that compartmentalize biodata. This compartmentalization is driven by, or leads to, positionality: an economic term that indicates that the value of a good depends on its exclusivity. The less parties have access to the data, the higher its value. Positionality of data comes in various flavors, depending on the 'data frictions' the data encounters (Bates, 2018). The constraints in data access can be temporal or spatial, because of ownership regimes, monopoly over the data, or because of the fact that it concerns a new scientific frontier. These vertical positionality effects are complemented with horizontal positionality. The ability to derive meaning from a biological dataset often requires access to other data sources. The synthetic biology responses to the COVID-19 pandemic provide an example of how data positionality impacts the ability to innovate. Synthetic biologist firms and academics aim at a disruptive innovation of vaccine development, therapeutics development, and testing. A vast efficiency increase of the current innovation approaches is beneficial given the likelihood that pandemic situations can occur more regularly and have vast negative economic and societal impacts. It is crucial though to guarantee biosecurity and biosafety, and to meet societal values like privacy, a fair access to therapies and vaccines, and a fair distribution of the benefits. Data availability and capabilities to make use of the data though are heterogeneous, and 'data frictions' can be experienced. The level of data sharing significantly impacts the ability to respond to a pandemic situation, as was clear during previous viral outbreaks with pandemic potential. Organizing a responsible innovation therefore requires close attention to the aspect of data accessibility and data mining capabilities. Positionality for instance makes it difficult to steer innovations: information asymmetries are at the core of positionality, and those hamper inclusiveness in the process, which is a common dimension of an RI approach (Burget, Bardone, & Pedaste, 2017) (Bogner & Torgersen, 2018).

1.5. Dissertation Outline

The extraction of data from biological systems gives rise to an avalanche of innovations. It thereby also gives rise to thorny ethical questions. Given this pivotal role, extraction of biodata from biological systems also provides venues for responsible innovation. A responsible biodata-driven innovation is likely to play an important role in meeting the 'grand challenges' that humanity and the planet are faced with, like a sustainable agriculture, citizen health, clean oceans, and global warming. Biological systems that have an *in silico* representation become subject to new ways of usage, control and steering, which opens new venues to better meet our moral obligations, as well as new risks that need to be mitigated. The massive extraction of biological data from human populations, ecosystems and biological populations therefore demands a rethinking of how to responsibly guide biodata-based technological innovation.

Central to this dissertation is the positioning of biodata at the interface between the natural realm, and the realm of human language and meaning. This positioning requires considering ethical questions around biological data not just as data questions – but in very close relation to the biological systems the biodata originated from. Digital twins and twin commons are used to analyze this intertwinement and to illustrate the ethical consequences and resulting opportunities for a biodata-based responsible innovation.

Chapter 2 zooms in on the relation between values and human personal genomics data. Large scale population sequencing initiatives chart the full genomes of entire populations of citizens, with the aim to develop novel treatments and an improved (personalized) healthcare. This sharply raises the question about the status of personal genomics data. Do personal genomics data have a special status, and do they thereby deserve a special treatment? The analysis builds on how the data relates to the biophysical reality, and how the data relates to our common language interpretations and moral assessments. The fact that the underlying biophysical reality escapes a full conceptual explanation, the symbolic effects that patterns in the data trigger, and our obligation to owe people the recognition of their capacity for moral identification, leads us to the concept of epistemic modesty in the field of genomics.

The focus of Chapter 3 is on Digital Twins - emerging data-driven engineering practices – as a conceptual tool to analyze how value-laden categories can become impacted by innovations in data-driven healthcare. Human personal biodata (genomics data, but also molecular phenotypic data, patient records, behavioral data, etc.) increasingly capture significant aspects of a person's biophysical, emo-

tional and mental constitution. As such, they are used in an engineering approach to medicine, which is asymptotically similar to the Digital Twins engineering practices that are applied in industrial predictive maintenance. This perspective sheds light on the operational character of human biodata: biodata-based innovations and practices have the potential to challenge deeply rooted moral distinctions. The distinction between healthy and diseased becomes challenged. Likewise, the distinction between therapy, preventative care and enhancement gets blurred when taking this biodata-based engineering perspective. Using digital twins as a conceptual tool sheds light on these moral questions that are triggered by biodata-based innovations.

In Chapter 4, complementary strategies are investigated for a responsible biodata-driven innovation. Biological data as an information resource can have characteristics of a Common Pool Resource, when used and managed by multiple stakeholders. The concept of 'Twin Commons' is proposed: the institutional arrangement of natural resources that have a tightly linked digital component which is shared and governed by a community, and that have research and innovation as important outlets. This concept helps in identifying values in the socio-economical and techno-scientific system that underpins the data-driven innovations. This can complement stakeholder involvement as a strategy in a responsible innovation approach. Responsible innovation based on natural resources is explored using the common pool resource framework and using the emerging field of biodiversity sequencing as an example.

Finally, Chapter 5 investigates the economic properties of biological data and the related biological systems as economic goods. Arguments are put forward why some biodata are positional goods. Biodata can be subject to manufactured scarcity or can be subject to other types of 'data frictions'. The resulting positional effects contribute to the ability to organize responsible innovation. This point is illustrated with synthetic biology innovations in response to the COVID-19 pandemic. Differences or delays in data access impact the ability to innovate, as well as differences in the ability to put data to use. The roles of governance and of collective self-regulation in commons is highlighted as venues towards responsible innovation, and opportunities to mitigate negative positional effects are identified.

2 How to Do Things with Personal Big Biodata

2.1. Chapter abstract

Genetic privacy is increasingly hard to guarantee due to the growing volume of personal health care data stored in databases. Although attempts are made to make the data anonymous or un-linkable, it was shown that individuals are at risk of being identified and re-identified. Anonymous DNA data was demonstrated to be linkable to individuals on the basis of publicly available information on the Internet. Utilization of linkable data can result in harm, inequities and discrimination since these data potentially reveal intimate personal, medical and family details. The increasing availability of genomic data – and more generally 'personal big biodata'[2] (which comprises a wide variety of medical and health care data covering both medical images and a panoply of biomarkers) – combined with the computational power and analytical tools of bioinformatics calls for a rethinking of privacy. In this paper we argue that in the age of personal big biodata, privacy implies first and foremost the responsible appraisal of the limits of what data allow us to know about individuals and we suggest furthermore that respect for human persons and their dignity implies an acknowledgement of the fact that there is always more to know about them than even the most comprehensive set of data may offer in terms of knowledge. We refer to the ideal of acknowledging the limits of our knowledge of persons as 'epistemic modesty'. We offer the epistemic modesty account of what privacy entails in the age of advanced genomics as a partial explication of the fundamental principle of the International Declaration on Human Genetic Data adopted by UNESCO in 2003: "(…) to ensure the respect of human dignity". (UNESCO, 2003) Personal big biodata carry the risk of epistemic immodesty. We argue that privacy is instrumental in ensuring a person's ability for self-determination in view of personal big biodata and the acts of epistemic immodesty of others. In addition to ex ante approaches to data protection such as privacy enhancing technologies, we also draw attention to

[2] The term 'Genomic data' in this paper refers to whole genome sequence data or to genotyping data. 'Genetics data' refers to data on a set of markers in the genome. 'Big biodata' includes also other types of molecular data (e.g. transcriptomics, metabolomics, proteomics) and phenotype data (e.g. fMRI images)

the potential of ex post approaches, e.g. tools and mechanisms supporting proper and modest interpretations of genomic data, as well as the detection of prohibited use of genomic data for certain ends.

2.2. A Call for Rethinking Genomic Privacy

Biomedical sciences currently witness a flood of data on research subjects. Techniques like personal genome sequencing, RNAseq expression profiling, metabolic profiling and medical imaging provide large volumes of personal biological data. These data potentially contain sensitive information, especially when they are combined with other types of health care data and lifestyle data, often voluntarily collected now by individuals with the help of smart wearable devices. The 'Baseline Study' initiated by Google is indicative of the increased interest and ability to execute combined analyses on a multitude of biological and other datasets. Google is collecting genetic and molecular data from 175 participants and aims to scale this up to thousands. Smart algorithms will be applied to pinpoint how a healthy human being should look from a data point of view. This baseline will be used to identify biomarkers for disease states.

Currently, de-identification of electronic records is commonly used as a measure to protect the privacy of research participants and patients. This however does not provide an absolute guarantee of privacy. Privacy regulations like the HIPAA Privacy Rule therefore request measures that minimize the probability of information disclosure, rather than demanding absolute guarantees of privacy since the latter cannot be guaranteed. According to the HIPAA privacy rule, a record is considered de-identified if either an expert assesses the remaining risk of de-identification to be 'very small', or if a fixed set of identifiers is removed (HIPAA Privacy Rule, 2007). The probability that de-identified records can be re-identified though is bound to sharply increase, as personal biological data are becoming ubiquitous and more easily accessible (Malin, Loukides, Benitez, & Clayton, 2011). Especially genomic data prove to be prone to re-identification (El Emam, 2011). Small sets of genomic features can already function as a unique identifier for a person. Moreover, a growing number of people have personal genotype data and even whole genome data stored in databases. In a research setting, the Thousand Genomes Project Consortium (The 100.000 Genomes Project, 2012) reported on the sequencing of the whole genomes of 1092 individuals. The Personal Genomes Project aims at sequencing 100,000 individuals (Lunshof, et al., 2010). Public healthcare is also

shifting its interest from genetic profiling towards full genome sequencing. The Faroe Islands started the FarGen initiative, an endeavor to sequence all 50,000 citizens in about five years time (Kupferschmidt, 2011). The results will be stored in a database that is linked to the Genetic Biobank, which contains samples from island inhabitants and medical and genealogical records. The UK Biobank is collecting samples and medical data from half a million Britons (Gottweis, et al., 2012), and the UK 100,000 genomes initiative aims at sequencing that many patients and their close relatives. Next to the traditional realm of research and healthcare, commercial companies are building proprietary databases with consumer genetics data. Companies like 23andMe, Navigenics and DeCode genetics experimented with genotyping services directed towards consumers. As a result of these and related activities, genetic and genomic data are accumulating in a multitude of databases.

This new data situation can lead to an increased risk of re-identification. Gymrek et al. (Gymrek, McGuire, Golan, Halperin, & Erlich, 2013) traced the identity of participants in genomic studies, based on public data. Sweeney et al. (Sweeney, Abu, & Winn, 2013) used previously published methods to re-identify volunteers from the Personal Genome Project. The authors were able to identify about 40 percent of the individuals out of 579 anonymous records that contained postal code, birth date and gender. Algorithms that link de-identified family relations to named people have already been developed (Malin, 2006). Homer et al. (Homer, et al., 2008) demonstrated that genotyping data provides a very strong identifier, by using it to determine whether the DNA of a person is present in a DNA mixture of up to 200 individuals.

This evolution towards a ubiquitous production of personal biological data, and the related risk of re-identification demands a rethinking of genomic privacy and big biodata privacy. There is a longstanding intuition that storing and sharing genetic information deserves special caution. "Genetic privacy" is the term that is most often used to refer to a cluster of rights associated with this point of view (Roche & Annas, 2001). Central is the right to protection from non-voluntary disclosure of genetic information (Lunshof, Chadwick, Vorhaus, & Church, 2008). This right has been embedded in legislation in many countries around the world. Measures to guarantee genetic privacy range from protective (Soini, 2012) to more liberal. The frameworks were developed for a setting where genetic information is mainly used for research purposes or clinical testing. This context changes now genetic data are frequently used beyond research. Moreover, the sheer amounts of data that are produced shift practices to whole genome analysis and other types of big biological data processing.

This radically new situation calls for a careful revision of the existing practices and technologies that are used to store, mine and communicate personal genomic data (Erlich & Narayanan, 2014) and big biological datasets in general (Sarwate, Plis, Turner, Arbabshirani, & Calhoun, 2014) (Choudhury, Fishman, McGowan, & Juengst, 2014).

2.3. Rationales for the Special Status of Genomic Data

The privacy concerns regarding genomic data have always been prominent in public debates. This prominence can be explained by the fact that they are accorded a very special status, an idea sometimes referred to as genetic exceptionalism. Genetic exceptionalism is the thesis that genomic data are different from other types of biological and medical data, and therefore require a special status and special protection. In this view, genomic data are seen as being extraordinarily informative. Green and Botkin observe for example that "Right or wrong, genetic information is believed to reveal who we 'really' are, so information from genetic testing is often seen as more consequential than that from other sources." (Green & Botkin, 2003). Article 1 of the Universal Declaration on the Human Genome and Human Rights (United Nations, 1998) proclaims: "The human genome underlies the fundamental unity of all members of the human family, as well as the recognition of their inherent dignity and diversity. In a symbolic sense, it is the heritage of humanity." A large body of literature focused on the question whether genomic data are indeed exceptional. The special biological roles of the DNA molecule often were put forward as a main argument. Such special roles are the immutable nature of the genome base sequence, the fact that it uniquely identifies an individual, the informational nature of DNA, and the fact that parts are shared among family members and ethnic communities (Green & Thomas, 1998) (Sarata, 2008) (Ilkilic, 2009). DNA is the central information carrier in biology and therefore potentially can reveal information about future illness, genetic defects, or unknown facts about familial relationships. Genomic sequences not only reveal information about a particular individual, but also about her or his relatives and the ethnic group the person belongs to. Taken on their own these features do not uniquely distinguish genetic information from other types of medical data (Green & Botkin, 2003). One can argue that not one unique property in itself, but rather the combination of properties distinguishes genetic data from other types of medical data (Sarata 2008). It is implied in the general definition of a category that it is associated with a unique

combination of properties that sets it apart from other categories. The fact that DNA based data is characterized by a specific set of properties therefore does not provide a rationale for a special status of the data, it just implies that it is different form other categories of data. Moreover, deducing a special moral status from scientific facts about a molecule can be regarded as committing the naturalistic fallacy (Ilkilic, 2009). Nevertheless, there is a persistent intuition that genetic data is special. Surveys show there is a belief that genetic information needs special consideration (Sarata, 2008). This belief has been explained in different ways. One explanation is that the scientific discoveries in molecular biology gave rise to an aura of power when it comes to DNA. Such sociological explanations leave open the question whether there are more fundamental reasons for a special status of DNA based data that remain overlooked, and that warrant genetic and genomic privacy.

A different set of rationales can be provided to account for the status and salience of genomic data as revealing "who we really are". These rationales are based on (i) the information content of genetic information (ii) the naturalistic connotation or 'naturalistic load' genetic information carries, i.e. the fact that it is construed as a veridical and direct representation of what the world is really like and (iii) the unique identification of individuals that genetic information makes possible, which depends in turn on the immutability and the uniqueness of the information. The first two arguments – in varying degree – not only apply to genomic data but also to other types of big biodata.

(1) Genomic data and big biodata can contain sensitive information. Genomic sequences for instance are related to how a person physically develops, to what extent she or he is prone to develop certain diseases, reacts to certain medication, is likely to display certain physical characteristics like muscle strength or longevity, etc. Behavioral traits can also have a hereditary component, for example verbal and numerical intelligence, susceptibility to addiction and to certain mental diseases, and certain character traits. Many of these characteristics are morally salient, and the derived claims are not restricted to an individual but also to relatives and ethnic groups. It also needs to be noted that genomic data are bound to contain more information than current methods are able to extract from them. For example, about 80% of the variation in height in a human population can be attributed to heritability, but the loci currently known to relate to height are only able to explain about 10 percent (GIANT Consortium, 2010). Also intelligence has a hereditary component, but even with very large samples of individuals no common genetic variants related to intelligence can be identified (Le Hellard & Steen, 2014). This

problem of missing heritability makes it plausible that more information will be derived as insights are gained into how to interpret genomic data. For other types of big biodata, the sensitive nature of the information can even be more prominent. Brain scans provide more direct information about a person's mental health and gene expression studies can provide information about disease states. Nevertheless, the sensitivity of the information that currently can be derived only partially explains the exceptional weight that is attributed to big biodata, since usage is currently limited. Translating personal genomic data into concrete medical advice for instance proves to be challenging (Ashley, et al., 2010). Health claims based on genetic data can be conflicting as they depend on the specific subset of genomic features that is probed, the specific genome wide association studies that are used in the interpretation, etc.

(2) Big biodata are derived from biological molecules or structures. These data are therefore perceived to reflect structures inherent in Nature itself. Big biodata are derived from molecules (DNA, mRNA, proteins, metabolites, etc.), i.e. from entities in a physicalist realm. These molecules take part in complex networks of biochemical interactions, which are the result of billions of years of evolutionary processes. Statements that are derived from big biodata therefore are perceived to carry a naturalistic load, which provides an additional rationale for attributing a special status to big biodata. On the other hand, big biodata are also part of the world of language and meaning. One can try to interpret the information that is embedded in these molecules and networks, and use this information to build theories, use it in discussions and assessments. Gilbert Hottois (Hottois, 1995) highlighted the fundamental difference between this 'naturalistic' and 'symbolic' realm in his analysis of technosciences. Hottois characterizes the naturalistic realm as 'operational': it concerns physical and chemical processes that bring about or produce reality. The symbolic realm is the area of language and meaning, of culture and values. One can try to represent the operational reality by making use of symbolic systems, as in scientific theories or in common language. But there is a limit to this. Both realms are fundamentally different and therefore have different dynamics. Following this distinction, we can state that big biodata have a unique position at the interface between the physicalist and the symbolic realm. This can be illustrated with genomic data. DNA is a molecule that is the result of evolutionary processes: it 'works' by bringing about organisms via its interactions with other molecules, but it is clearly not designed to be 'read'. On the other hand, DNA has the character of a text, with modularity, structure, compositionality that can be read and interpreted. The

sequence of the bases in the molecule constitutes a quaternary code that can be represented in a computer. Notwithstanding the limited ability to interpret genomic data, it is clear that the DNA carries crucial information for the development and functioning of the person. It is an information carrier and thus also part of the symbolic realm. For these reasons, genomic data have been referred to as the 'genetic blueprint', the 'genetic code', or the 'genetic program' of a person. Other big biodata types similarly reside at the interface between Nature and culture. fMRI measurements of electromagnetic activation patterns for instance reflect biological processes in the brain via imaging software that implements models, but can eventually be translated into claims about mental states. This particular position of big biodata – at the interface between the physical and the symbolic levels – affects the weight that is attributed to the derived symbolic statements.

The fact that this impact is assigned can be accounted for by looking at the way the relation between big biodata and natural kinds is conceived. In a realist interpretation of the world, certain groupings of entities are not merely man-made distinctions but reflect the way reality itself is structured. When organizing a collection of stones, one can for instance group them by shape: round stones, square stones, heart shaped stones, or by the role they play in religious ceremonies, or by their color and patina (and other so-called secondary properties, etc.). The stones can also be grouped based on their inner atomic structure, crystal lattice structure, chemical composition, which is tightly linked to their geological formation: igneous stones, sedimentary stones, metamorphic stones. A realist will state that the latter classification is a better reflection of the way reality itself is structured (Wilkerson, 1995). The hidden inner structure of the stones determines their membership of a natural kind. In chemistry, natural kinds seem to be unproblematic. Water for instance is defined by the chemical structure H2O. Realists hold that this is the case in every possible universe. Even if a person lacks the epistemic means to uncover the hidden structure, membership of the class 'water' will depend on whether a sample X and the reference 'water' sample have the same hidden structure. The underpinning idea is that natural classes exist independent of human interests. This notion of hidden structures and natural classes is central to physicalist accounts of the universe. Natural kinds supposedly 'carve Nature at its joints'. They underpin a categorization that is not arbitrary but reflects the way reality is structured (Wilkerson, 1995). For many centuries, biology delivered the archetypical examples and paradigm cases of natural kinds. Species were believed to be natural kinds, and the essence of an organism determined its membership of a species. Extrapolating this essentialist

worldview to the genomic era puts a heavy metaphysical weight on genomic data, since claims based on genomic data will acquire the authority of reflecting the 'hidden structure' or the essence of a person, analogous to the chemical structure of a water sample that determines its membership of the natural kind 'water'. Distinctions that are based on genomic data are strong, since they have the appearance of not being the result of a cultural and social convention, nor being an accidental attribute or secondary property of a person. They are easily equated to epistemic claims about the 'very essence' of a person and easily give rise to essentialism about persons and their properties.

(3) In addition to the information content and the naturalist load, there is another factor that supports the claim that genomic data are special, to the extent that they may be called 'exceptional and unlike any other information about a person'. The reason can be found in the unique identifying power of genomic data. Genomic data uniquely identify a person. They are given at birth, and epigenetic modifications and mutations aside they are probably one of the most constant characteristics of a person. This unbreakable link between a person and her or his genomic data adds to the reification of claims based on a person's genome. An essentialist interpretation of personal genomic data reifies the symbolic claims that are derived from it (Barnes & Dupré, 2008). The strength of the identification between a person and his or her genomic data is carried over to value-laden claims that are made based on these data.

2.4. Big Biological Data as a Substrate for Social Classification

Because of this widely shared conception of the special status and salience of genomic data, genomics has also inserted itself into our thinking about the social classification and categorization of persons.

At this point a distinction should be made between the use of genomic data (i) in forensic practices as evidence for (re-) identifications of individuals, (ii) as a basis for classifications in clinical and clinical research practices and (iii) as basis for categorizations in social and institutional practices. The use of DNA in forensics has been widely accepted as a highly reliable technique of establishing or confirming the identity of both criminals and victims. Genetic material is in both cases a great help in finding out 'who is who'. Statistics and Bayesian probability theory have been applied to increase the reliability of our reasoning with genetic evidence in criminal

justice procedures so as to prevent errors in inferences in these contexts. Genetic evidence is one of the strongest, unique identifiers of individuals in a forensic sense.

Secondly, the use of genomic data in clinical diagnosis, therapy and research is ubiquitous. Our knowledge of diseases and health problems has benefited tremendously from our study and understanding of the human genome. But it also became obvious that the picture is far more complex than initially anticipated. The complexity of the information embedded in DNA proved to go far beyond the approximately twenty thousand protein coding genes. Studies over big populations show that phenotypic traits are often associated with large sets of genomic features. The non-protein coding part of the genome plays a significant role, as highlighted by the ENCODE project (ENCODE Project Consortium, 2012). In most cases a complex network of multiple genes, gene control elements, etc. contribute to the phenotype. The secrets of genetic imprinting, epigenetics, and proteomics are being unraveled and add to the complexities. For many diseases with a genetic component it is anything but clear that genetic material allows for easy and quick clinical diagnosis with high reliability. Reasoning from genomic data implies data reduction: out of the 3.2 billion base pairs of the human genome, characterizing features and discrete categories need to be extracted. For instance, in the case of personal genomic information, the process starts with a biological sample taken from an individual. The DNA in the biological sample is extracted, prepared and run through a sequencing machine, which will generate a series of signals that correspond to the sequence of bases in the DNA fragments. The signals are recorded, stored in a computer system, and assembled via bioinformatics tools into a genomic sequence. The end result is a series of character strings on a computer hard drive. The features in this personal genomic sequence can then be analyzed in order to derive personal genomic information: statements about e.g. disease susceptibilities or character traits. This process of interpretation often relies on the output of statistical inferences like Genome Wide Association Studies (GWAS). At the highest level of abstraction, interpretation of a personal genome sequence results in symbolic statements. For example: "Person A belongs to the group of people that has a higher risk of developing disease D". This statement divides the population in two groups: the group non-D that has a low risk to develop the disease, and the group D that has a high risk. Given the complexity of biological systems and the underlying probabilistic nature of the inferred relations, this statement is a simplification of a complex molecular and biochemical reality. Interpretation of genomic data often renders probabilities, rather than deterministic certainties or memberships of clearly distinguishable

classes. A classic example is the presence of the ApoE4 allele that relates to an increased risk of developing Alzheimer's disease. The information needs to be interpreted as probabilistic: one can be carrier of the allele and nevertheless grow old without developing the disease. The same mechanism holds true for other types of personal big biodata. Also, for e.g. metabolite data or brain images, features need to be extracted to put these data to use. Reduction of big biological datasets to symbolic statements allows for talking in a common language – closer to ordinary language – about what these complex datasets can mean for us. It is a necessary process if we want to use these data in everyday assessments, comparisons, or categorizations. We need to keep in mind though that such statements are always simplifications of an underlying complex biological reality.

Thirdly, as far as the evidentiary role of genetic information in social classifications (e.g. entrepreneurial, leadership, impulse control, alcoholism) is concerned, the story is even less straightforward, and the following considerations need to be taken into account. As indicated, interpretation of big biological datasets implies a data reduction, in which complex characteristics are used to classify individuals in terms of membership of discrete social or clinical categories. Some of these categories represent naturalistic properties in a natural population, others however are social constructs which may be socially controversial and contested, categories such as "Attention Deficit Disorder", "homosexuality", "weak impulse control" and "alcoholism". In social, institutional and in non-scientific discourse the certainty and special evidentiary role that genomic data can play in forensic identification and in a clinical context, as discussed above, cannot be assumed to carry over to social classification practices. Simply because genomic data can help to identify a person uniquely with high degrees of certainty, or may help to underpin statistical inferences in clinical research, we cannot infer that symbolic interpretations and the corresponding social labels attach with the same degree of certainty to a given individual. This process implies jumping beyond the evidence and endorsing claims that often cannot be justified. Nevertheless social categorization and classifications based on a person's genomic data may give rise to entrenched perceptions or 'frames' of a person's social identity and may eventually constrain an individual's ability to choose how he or she presents and defines himself or herself in social contexts. The genetic basis of novelty-seeking behavior can be taken as an example. Some allelic variants of the dopamine receptor genes were associated with differences in dopamine binding and with phenotypes related to novelty-seeking behavior (Padmanabhan & Luna, 2014). A genetic test thus can be designed to label individuals as either

carrying this allele or not carrying this allele. The genetic evidence, though, will not warrant tagging the individuals with the labels 'novelty-seeking' or 'not novelty-seeking'. The presence of the allele does not say much about the novelty seeking tendencies of a particular individual. At best, it may indicate a statistically higher likelihood of displaying such behavior. Without caution the information on whether an individual carries the allele may easily get translated into a claim about how a person is really, namely being eager or not eager to engage in novel experiences.

We refer to this cognitive attitude towards genomic data as epistemic immodesty. Epistemic immodesty (van den Hoven J. , 2008) is defined as the making of knowledge claims about persons that are not fully supported by the evidence available with potentially significant cost or negative moral consequences to the target person, while ignoring how the person involved would like to be identified. In epistemic immodest judgments one draws conclusions about who a person is, or one claims to be acquainted with one or more of his or her properties on the basis of limited or irrelevant evidence. Because of the reasons mentioned before, big biodata inherently carry the risk of giving rise to immodest claims on what one knows or can know about a given person. Classification of people based on biological characteristics transforms statistical claims into symbolic claims about who a person is. These claims are strengthened by the naturalistic load of the classifications and by the unique identifying power of genomic data. Moreover, genomic studies often concern morally salient characteristics, increasing the proneness to epistemic immodesty. As an example, a study on patients with bipolar disease revealed a genomic region that was indicative of an increased proneness to suicide attempts (Willour, et al., 2012). Epistemic immodest claims can arise when such probabilistic findings are put to use beyond the field of science. When for instance applied by a future employer or to customers in a consumer genetics setting, the focus is not on identifying generalizations out of the data of many research subjects. In this different pragmatic context, the focus shifts towards the labeling of particular individuals by putting the generalized data to use. It can be foreseen that the claim to know that person will give rise to moral judgments that may negatively affect the person. The immodesty is constituted by the fact that the judgment fails to take into account the subjective experience of individuals and the way they want to be identified.

2.5. Genomic Privacy and the Right to Self-Determination

The question then is how to deal with personal big biodata and how to respect and protect them so as to prevent making data subjects vulnerable in the light of the availability of the data. Personal big biodata need to be protected to prevent harms, inequities and discriminatory practices of which patients and data subjects may become victims (for a taxonomy of moral reasons for data protection see van den Hoven (van den Hoven J. , 2008). It is obvious that individuals can be harmed, wronged, disadvantaged, targeted and exploited in numerous ways on the basis of the access that others have to their genetic information. A right to control access to your own big biodata, the requirement of informed consent is still the central point of all existing data protection legislation. There are four types of moral reason for this: 1. Prevention of (information-based) harm, 2. Equality of opportunity and fair treatment in markets for commodified personal data, 3. Informational justice and discrimination and 4. Moral autonomy (van den Hoven J. , 2008).

First of all, it is clear from the discussions above that personal big biodata potentially can be the cause of information-based harm. Information based harm is here defined as harm that could not have been (easily) inflicted if particular information would not have been available. Secondly, another type of moral wronging may occur when big biodata are commercially exploited without proper benefits to the data subject, or without him or her even knowing about this. Thirdly, a form of moral wrongdoing occurs when the data are produced in a research or clinical context, but become available in a very different social sphere, such as the world of insurance or potential employers (market) or police or criminal courts (criminal justice). The use of information about a person at a time and place where it is deemed inappropriate or is irrelevant is a form of discrimination. The prevention of discrimination calls for controlling the boundaries of social spheres in which the big biodatasets are produced and used. Also the different practices (forensic, clinical research, social / institutional), their different associated standards of care in reasoning and methodology, the governance, and their norms for the fair allocation of access ought to be observed and separated.

Autonomy is the fourth moral reason to be concerned with the protection of personal big biodata. Some types of personal big biodata like genomic data and brain structures are largely immutable during a person's lifetime. Such immutable data that is intimately linked to a person vastly increases the impact when releasing this data, since it can impair a person's ability to determine how he or she presents himself or herself to others. The focus of privacy concerns in this case is to prevent

one specific form of wronging moral persons, namely the fact that they are made subject to social sorting, categorization and classification, on the basis of personal genomic data or personal big biodata in general. When people are classified in this way and given the perceived reality of genetic exceptionalism, there is a lack of plausible deniability of the evidence for these classifications on the part of an individual (assuming he or she is not an expert in genetics). This interferes with the moral autonomy of persons, the moral right of persons to define themselves and shape their own identity, present themselves and have their chosen public persona recognized and respected by others. Controlling the way one presents oneself provides the necessary space for self-determination. To be recognized as a being who is able to conceal information is fundamental to being perceived as a 'self-representing being', as argued by (Velleman, 2001). Norms of privacy dictate that certain things should be allowed to be concealed. If the ability for self-presentation is impaired, one is naked in the sense that one is exposed in a way that fundamentally impacts one's standing as a social agent.

In the case of personal big biodata that are immutable, change in a person's profile can only occur at the level of the interpretation of these data. It is thus important to instill mechanisms to control the process of the inscription of meaning to the data. This process though is largely outside the control of the person. A first reason for this lack of control lies in the fact that new interpretations emerge as technologies and insights evolve. This fact is for instance acknowledged in the informed consent form of the Personal Genome Project: "because the science in this area is evolving, and data will be collected on an on-going basis by the PGP, the risks involved due to your participation in this study, as well as the likelihood and severity of such risks, will change over time" (PGP Consent Form, 2014). Secondly, information systems that are used to store personal data can constrain the freedom of a person to manage his or her own identity (Manders-Huits & van den Hoven, 2008) (Manders-Huits, 2010) irrespective of the type of personal data that are stored and processed. In the case of personal big biodata, these systems represent a person via a digital record, which can contain the person's genomic sequence, medical record, etc. Digital representations reduce a person to his or her representation in the system, and allow for the creation of 'types' of people by clustering such representations. Together with the fact that data in these systems tend to persist, digital identities result in a reduced ability to reshape one's identity. Central to this moral autonomy reason for data protection (van den Hoven J. , 2008) is Bernard William's proposition that respect for persons implies attempts at moral identification, namely

the identification of a person as someone who is engaging in self-identification and who ought be identified as such. This is an important aspect of what it means to respect someone: we owe the other an attempt at identification (not clinical or forensic) as the person who he wants to be identified with. As Malcolm Forbes has already observed: "We'd all like to be taken for what we'd like to be." The object of knowledge claims therefore becomes a more complex one, namely an objective physical human being who is characterized in terms of a panoply of biomarkers and a genome, but who has a subjective conception of him- or herself. Self-identification encompasses the ability and freedom to determine and reorient oneself. In databases this ability can be constrained when a person's identity is reduced to a digital record. Defining one's course of life can be hampered when the person has no ability to control this digital information.

Epistemic modesty is thus a moral source of restraint in the process of the inscription of meaning when this process of inscription of meaning is out of the control of the subject. In addition to the avoidance of information based harm and the prevention of inequity and discrimination, genomic privacy provides a person with the freedom to shape his or her own social and moral identity, and to relativize or completely undo previous and external determinations made by others and actively engage in self-presentation in a range of social contexts. The epistemic hubris or immodesty that is so easily produced by practices associated with collection and utilization of personal big biodata should be counteracted by instilling and institutionalizing forms of epistemic modesty about claims that we know 'who someone really is' if we know this person's genomic blueprint.

Ensuring privacy as a property right in one's own biological data, and creating markets for personal biological data may not work to ensure a fair processing of these data. Markets where people sell and buy personal biological data are highly problematic, as is the case in e.g. markets for organs. The items changing hands are unique and in a relevant sense belong to the individual, or can be seen to be co-constitutive of the individual. And the transfer is irreversible, since once the data is out, it is practically impossible to get control over it again. The idea of a market also implies freedom, informed consent, and full information, which are all precisely what is at stake. Genetic data protection by means of data markets and transferable property rights therefore seems a bad idea. An alternative to a market is the "information altruism" as highlighted by Lunshof et al. (Lunshof, Chadwick, Vorhaus, & Church, 2008). In this setting, the individual shares his or her genomic data in order to support scientific progress, aiming at generating benefits for many. Information

altruism is a peculiar form of altruism since the audience one provides the information to is unknown and global, and so are the potential beneficiaries. By subscribing to the open consent form, the research subject allows the storage of his or her personal genomic data in publicly accessible databases without any guarantee of anonymity, privacy or confidentiality. Withdrawal from a study is possible at any time but the research subjects need to acknowledge that their data might not be completely removable. They also need to acknowledge that the information release might not be to their benefit and can even harm them. An information altruistic policy therefore can lead to the vulnerability of research subjects.

The tension between the personal and the societal interests is bound to be a central topic in big biodata privacy debates (Knoppens, 2009). The individual benefits from a strong protection of his or her big biodata. The community benefits from an improved healthcare that is the result of a better understanding of biological data. Inaccessibility of personal genomic data for the research community can hamper scientific progress and the related societal benefits. There is a need for solutions that ensure protection of the individual, while allowing researchers to mine the data. Whether technical solutions can dissolve this moral dilemma is being investigated (Ayday, De Cristofaro, Hubaux, & Tsudik, 2013). For instance, Baldi et al. (Baldi, Baronio, De Cristofaro, Gasti, & Tsudik, 2011) applied encryption technologies to full personal genome sequences. Such technologies allow for the *in silico* execution of genome tests, without disclosing the information outside of the intended audience. In this setting, a person can have personal genomic data available on an electronic carrier in an encrypted way. The person then can consent to let a service provider run an algorithm on his genomic data, without having to disclose the data to this service provider. Such encryption strategies can provide the basis for innovations in healthcare and consumer genetics that ensure genetic privacy, while providing researchers the proper access to analyze the data.

In any setting, it will be crucial to create room for self-determination and for mitigating the effects of epistemic immodesty, by giving persons more means to control not only the access to their big biodata, but also by giving society the means to ensure proper interpretation and usage of these data. Such means imply pluriformity in available models, frameworks and rival theories, access to countervailing interpretations, checks and balances in actions based on interpretations of genomic data. Good governance of personal genomic data is a way to instill epistemic modesty in users. Practices should also extend to training and support for genetic counselors, in systematic reviews of the interpretations offered by genetic services,

the required indication of probabilities in claims, creation of modes of contesting available categorizations, and the dissemination of scientific information in society.

3 Digital Twins in Therapy and Enhancement

3.1. Chapter abstract

Personalized medicine approaches use fine grained information on individual persons, to pinpoint deviations from the normal. 'Digital Twins' in engineering practices provide a conceptual framework to analyze these emerging data-driven health care practices, as well as their implications for therapy, preventative care and human enhancement.

Digital Twins stands for a specific engineering paradigm, where individual physical artefacts are paired with a digital model that dynamically reflects the status of that artefact. When applied to persons, Digital Twins are an emerging technology that builds *in silico* representations of an individual that dynamically reflect molecular status, physiological status and lifestyle over time.

We use Digital Twins as the hypothesis that one would be in the possession of very detailed bio-physical and lifestyle information of a person over time. This perspective redefines the concept of 'normality' or 'health', as a set of patterns that are regular for a particular individual, against the backdrop of patterns observed in the population. This perspective also induces a normative shift in how therapy and enhancement can be distinguished, as can be illustrated with the cases of the 'asymptomatic ill' and life extension via anti-ageing medicine. This normative shift relates to how meaning is derived from measurement data. We use a promiscuous realist account to clarify how moral distinctions based on digital twins are the result of both cuts made at Nature's joints, and the world of language and meaning that is grafted on these structures.

Ethical and societal implications of Digital Twins are explored. Digital Twins imply a data-driven approach to health care. This approach has the potential to deliver significant societal benefits, and can function as a social equalizer, by allowing for effective equalizing enhancement interventions. It can as well though be a driver for inequality, given the fact that a digital twin might not be an accessible technology for everyone, and given the fact that patterns identified across a population of digital twins can lead to segmentation and discrimination. This duality calls

for governance as this emerging technology matures, including measures that ensure transparency of data usage and derived benefits, and data privacy.

3.2. Personalized medicine – therapy as digitally supported engineering

Engineering approaches are ubiquitous in modern medicine. In current health care practices, one engineers a vascular bypass to restore the blood flow in case of atherosclerosis, repairs a heart valve, or replaces and old lens in the eye of a patient suffering from cataract. These engineering practices are rooted in the explanatory power and practical successes of the mechanical philosophy that has gradually emerged since the Renaissance. For instance, the drainage of the Low Countries provided significant improvements in the understanding of pumps, valves and hydraulic systems. These evolutions resonated in the work of contemporaries that studied vascular anatomy and the working of the heart (Novell, 1990). The description of the heart as a pump with one-way valves eventually opened the route to engineering actions like heart valve replacement. The engineering perspective developed into an important paradigm in current health care and therapy. Many Technical Universities in the world now train and educate engineers in clinical technology curricula, and doctors routinely work with engineers with a range of different backgrounds.

This engineer's point of view also forms the hidden premise in many debates about human enhancement. When it is possible to replace broken parts in the body, and to tweak, fine tune, and optimize them, it is in principle also possible to extend this body with new functionalities. Neural implants can for instance be used for visual prosthetics for blind people, but they also open the route towards capabilities going beyond normal human sight and give access to a range of normally inaccessible parts of the electromagnetic spectrum. Drugs like Ritalin can be used to help ADHD patients to focus, but can also be applied to boost mental performance in people that don't suffer from ADHD. The engineer's perspective becomes especially striking in the case of human germline editing with the aid of Crispr-CAS (Liang, et al., 2015). In therapeutic applications, one could consider the editing of the nucleotides that give rise to severe Mendelian diseases, thereby preventing a lot of human suffering. With the same engineering approach, one can potentially bring in traits that go beyond current human capabilities. For example, one could consider engineering human hemoglobin to be more like shark-hemoglobin, thereby allowing humans to store more oxygen in the blood. It will though be very difficult if not

unfeasible. The engineering approach to health in contemporary medicine is confronted with the sheer complexity of the human body and its operations. Here a purely mechanistic approach has revealed insufficient. It is for instance very difficult or impossible to precisely predict the efficacy of a drug and its side effects in a concrete patient. A large quantity of the massively prescribed blockbuster drugs therefore has suboptimal effects. Complex multifactorial diseases prove to be very hard to tackle via an engineering approach. Along these lines, human enhancement will require the engineering of complex and interconnected traits. This might well be impossible to achieve with current medical engineering approaches.

To get a better grip on this complexity, large initiatives are established to generate detailed molecular data of patients and healthy research subjects. Publicly funded initiatives like Genomics England (Marx, 2015) or the US precision medicine (PMI working group, 2015), and private initiatives like Longevity and the Mayo Clinic Centre for Individualized Medicine gather genomic information on large numbers of individuals. These initiatives ultimately aim at the development of digital models of certain aspects of patients, allowing for more targeted health care interventions. Instead of using an overall scheme of the average human body and its responses, personalized medicine starts from the premise that health care can vastly benefit from detailed molecular and lifestyle data of each individual patient. In the case of picking the right drug to treat a cancer, the efficacy of this approach already has been proven. Genotyping an individual's tumor tissue provides clues on which drug will result in the biggest impact and the smallest side effects (Kummar, et al., 2015). Personalized medicine also carries the promise to lead to predictive medicine, where diseases can be predicted and thereby also preventatively treated.

At a fundamental level, the mentioned initiatives share the belief that refined mathematical models of patients, fueled by big biodata, will drive more precise and effective medical interventions. The availability of molecular readout technologies and of sufficient computational power increasingly makes it possible to build such models. The models can be complemented with continuously tracked health and lifestyle parameters, eventually resulting in a digital representation of certain aspects of an individual patient. The concept of a "virtual patient" or even a "virtual self" therefore was proposed as a venue for European healthcare: "realistic computer models that are built and validated upon experimental big data collected by the most advanced technologies from molecular to macroscopic scales" (Lehrach, Ionescu, & Benhabiles, 2016). This manifesto projects vast health improvements, reduction of

health care costs, and an increased personal freedom in dealing with our own biology.

Provided such 'virtual selves' indeed become available, they will take the engineering practices in health care to a different level. In first instance, they can vastly increase the resolution at which one can define normality and disease. The personalized high-resolution models of the "virtual self" will provide a detailed map that allows to better pinpoint disease states, i.e. deviations from the normal. This 'normal' or healthy state can be defined at a high resolution, using molecular, phenotypic and behavioral level of entire populations. Instead of basing medical interventions on the responses of the average person, digital models carry the promise to tailor healthcare to the anticipated responses of individual patients. Secondly and relatedly, personalized high-resolution models also have the potential to redefine the concepts of therapy and enhancement. In the following sections, we argue that the conceptual and practical implications of this redefinition can be best captured with reference to the concept of Digital Twins, an emerging practice in engineering that has individualized digital models a central instrument.

3.3. Digital Twins as conceptual tool: relevance for the therapy and enhancement debate

Much can be learned about therapy and enhancement, by putting the quest for accurate digital models of patients in the light of the emerging Digital Twins engineering practices. The data and model driven healthcare trends namely bear a striking similarity with the usage of Digital Twins in engineering.

A Digital Twin consists of two systems: a physical system (e.g. one particular machine) and a computer model that closely reflects the architecture and the dynamics of this one particular system. Unlike traditional engineering models, Digital Twins reflect the particular and individual, the idiosyncratic. Traditional engineering models reflect the generic: they apply to multiple instances. A Computer Aided Design model for an airplane jet engine reflects the structure of all the jet engine instances that were built based on this model. Sensors that allow for continuous monitoring of technical systems increasingly make it possible to individualize such digital models, in order to represent the status of one particular physical object. This type of model has been termed 'Digital Twin', since it closely reflects the inner state of the physical twin object. Digital Twin models are used in predictive maintenance. In this case the dynamic behavior of a particular object is closely monitored, for instance a part of an engine. These data are then related to large amounts of

sensor data from similar objects. This allows to identify anomalies long before parts actually break down. Digital Twins are also applied in the engineering process, to simulate the outcome of technical interventions like fixes and upgrades.

The Digital Twin concept was applied by NASA in the development of aerospace vehicles that last longer and endure more extreme conditions. A Digital Twin in this context was defined as "an integrated multi-physics, multi-scale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its corresponding flying twin. … By combining all of this information, the Digital Twin continuously forecasts the health of the vehicle or system, the remaining useful life and the probability of mission success. The Digital Twin can also predict system response to safety-critical events and uncover previously unknown issues before they become critical by comparing predicted and actual responses." (Glaessgen & Stargel, 2012). The concept emerges also as a key element in Industry 4.0 strategies. It was termed "a living model of the physical asset or system" that allows to "continuously adapt to changes in the environment or operations and deliver the best business outcome" (Infosys Insights, 2016), a "digital copy that is created and developed simultaneously with the real machine" (Siemens, 2015), "the bridge from the physical to the digital worlds, providing understanding of each unique asset over time" (GE, 2017). Digital Twins have been applied to optimize the operations of power plants, wind turbine parks, critical jet engine components, etc.

Digital Twins driven engineering in industry bears striking resemblances to the emerging data-driven personalized health care practices described above. These novel engineering approaches to health care also build on dynamic and high-resolution digital models of genetic, biochemical, physiological and behavioural aspects of individual persons. Digital Twin based medicine is far from being an established fact yet. Various initiatives nevertheless pave the path by gathering detailed molecular data from individual patients (The 100.000 Genomes Project, 2012), (Telenti, et al., 2016). Closer to the engineering of artefacts, attempts are currently already undertaken to develop Digital Twin models of the heart (Scoles, 2016).

Next to being an emerging practice in health care, the concept of Digital Twins provides a very viable instrument for analysing concepts like health and disease, therapy and enhancement. It does so for multiple reasons. Firstly, the perspective taken in contemporary medicine is that of rational maintenance, optimization and even design of (very complex) bio-physical systems. Interventions in both engineering

and medicine can be considered as engineering actions. Probabilistic models of human individuals in personalized medicine aim at supporting the engineering of a healthy status. This includes an approach analogous to predictive maintenance in industry. Molecular biomarkers can provide an early identification of upcoming disease states, even before the disease is manifest. Interventions can then be done to restore the system to a healthy state. Further along the same lines, human enhancement scenarios implicitly assume that humans are (eventually amongst other things) biophysical system of which the components and the functioning can potentially be understood in terms of mechanistic processes, and are therefore amenable to engineering of current features, and the engineering of novel ones. Secondly, these activities in both fields are guided by big data and by mathematical models that represent one individual person or artefact. In both engineering and medicine there is a strong belief that interventions will be more precise and effective, when individualized mathematical models are used that capture the actual status of one particular artefact or person over time. Models of artefacts are evidently much more comprehensive than models of an organ or of the metabolic status of a person. Artefacts have building plans and are much less complicated than human beings. Models in medicine are still very partial and coarse grained, but nevertheless already show effectiveness, as can be seen in the field of cancer treatment. By combining various types of omics-levels one can anticipate that a much higher level of predictivity can be achieved than when using only individual data types, like genomic data.

3.4. Digital Twins and the concepts of the normal

To the extent that physicians already tailor treatments to the medical history and actual status of their patients, one can say that medicine has always been personalized (Brenner, 2012). A Digital Twin approach though would not only take the disease history of an individual patient into account, but also her healthy state, in great molecular and behavioural detail. Digital twin approaches in health care will heavily rely on a detailed picture of the healthy state of an individual, not merely on a record of disease states. 'Normal' in this context refers to the typical molecular, physiological and behavioural patterns observed in the individual, interpreted against the backdrop of the patterns observed in the entire population. Blood pressure readouts provides a simple illustration of this point. The sphygmomanometer is available for more than hundred years, nevertheless there is not yet a clear understanding of what is a 'normal' blood pressure. One of the reasons is that this

cuff-based blood pressure determination method results in sparse measurements over a person's lifetime. (Steinhubl, Muse, Barrett, & Topol, 2016). This makes it impossible to assess the impact of day or night, age, caffeine consumption, stress conditions, and so on. The result is improper management of hypertension in many cases. Wearable devices nowadays can monitor an individual's blood pressure continuously. A "virtual medical assistant" has been proposed that uses machine learning to mine these data streams and identify the blood pressure trends that are unique to that particular person. Such information can provide an individualized concept of what is a normal blood pressure, against the backdrop of trends observed in people with similar age, life style, etc. (Steinhubl, Muse, Barrett, & Topol, 2016). Similar approaches are relevant for molecular biomarkers. Identification of the risk to chronic heart failure can benefit from serial measurements of biomarkers over time, rather than from single values (Miller & Jaffe, 2016). This Digital Twin approach is in sharp contrast with current normal function accounts that define a normal or healthy state simply based on population statistics, with no reference to any individual's conditions. Digital twin models are continuously fed with all types of information during the lifetime of a person. This will allow to determine what the statistically normal patterns are for that person for a manifold of parameters. These normal patterns for the individual might well lie out of range when compared to the ones observed in the population.

Next to the fact that the normal will be defined by the individual, the normal will also have a multi-dimensional and high-resolution character. Natural variation amongst individuals, that make it otherwise difficult to pinpoint what is exactly normal, can be mapped in a high dimensional space of all different sorts of data. Such approach will allow to obtain a much sharper statistical definition of the normal or healthy state of an individual, and likewise of disease states or disease susceptibilities. Confounding factors like age, lifestyle, genetic background can be factored in in these models.

Thirdly, Digital Twin models will allow for comparing normal patterns across individuals in great detail. The multidimensional space of properties across Digital Twins can be used to cluster similar individuals. Currently comparison with the normal range is mainly based on age and gender. One can expect that a high-resolution picture will lead to a great heterogeneity of types of human beings, each of them characterized by their own normal patterns. This effect already becomes apparent at the genomic level. High resolution genomic sequence data of multiple individuals revealed that human genomic variation was larger than originally

anticipated (Telenti, et al., 2016). Variation in genomics regions that were previously perceived as junk seemed to have functional significance when having more data at hand. Similarly, it has been suggested that there might be a manifold of healthy states in human microbiomes, and therapy boils down to moving the composition of the microbiome towards one of these healthy attractors (Lloyd-Price, Abu-Ali, & Huttenhower, 2016). This transparency in the heterogeneity of what is normal raises the question on whether natural levels are optimal and are prone to engineering (Kahane & Savulescu, 2015).

High resolution models of what is normal or healthy constitutes the cornerstone of upcoming personalized medicine approaches. A detailed picture of the healthy assumedly allows for a better identification of potential or actual disease states that need to be remediated. For example, assessment of which particular chemical is optimal to treat a cancer in a specific patient requires classification of that cancer by its driver mutations. This implies a precise understanding of how a healthy genome looks like, and which deviations from this normal situation are harmful. The approaches though often base the concept of the normal on the population, not yet on the individual. Early initiatives like the Framingham Health Study used physical examinations and lifestyle interviews on a set of healthy individuals. These studies played an important role in understanding the impact of lifestyle on cardiovascular diseases (Framingham Heart Study, 2017). Population genomics studies sequence large amounts of citizens to infer genetic diseases, and by consequence build a picture of a healthy genome. Initiatives like the Metagenomics of the Human Intestinal Tract (MetaHIT), the Human Microbiome Project (HMP), and Chinese diabetes consorts reported on microbiomes of healthy individuals (Lloyd-Price, Abu-Ali, & Huttenhower, 2016). With the availability of high throughput sequencing technologies and of wearable devices, multi-dimensional molecular pictures of normal patterns can be developed at the individual's level. Examples in this direction are a project by a Google spin-off that will track ten thousand healthy Americans for their genome, microbiome, physiological parameters captured by a wearable device, life style and well-being.

3.5. Digital Twins and the concept of enhancement

Future availability of Digital Twins for persons also has consequences for the notion of enhancement. One common definition of enhancement is the improvement of general abilities, "beyond the species-typical level or statistically-normal range of

functioning of a human being (President's council on bioethics, 2003; Allhoff, Lin, & Steinberg, 2009) (Menuz, Hurlimann, & Godard, 2013). The distinction between therapy and enhancement was proposed as a means to identify actions that require special moral consideration, because they change the constitutive aim of our medical interventions, which is to cure. (Daniels, 2000; President's council on bioethics, 2003).

At first sight, these definitions seemingly match the Digital Twin's engineering perspective. Engineering actions always aim at modifying a system. These modifications can be classified as either maintenance or improvement. Maintenance actions using Digital Twins focus on establishing a baseline of normal functioning for an individual artefact, to predict more accurately which maintenance interventions are needed. In repairs the modifications address a problem, and aim at restoring a system to the normal functioning. Improvement actions like 'souping up the engine of a motor' bring an existing functionality beyond the normal, or introduce a novel functionality.

Such clear-cut distinction depends though on the reference taken. In the engineering cases, it is 'the normal' as defined in the certification or classification (e.g. of a ship or the weight of a payload, stress, torque) which helps to define the boundary between systems maintenance and problem remediation versus improvement. In a similar way, the normal in the biological realm defines the boundaries between therapy and enhancement in "species typical normal functioning" accounts (Daniels, 2000). This definition of normal functioning is often based on population statistics. When taking the individual's normal patterns as reference in a Digital Twins approach, therapy entails the maintenance or restoration of this individualized normal state. It is well possible that an individual performs well in a certain trait when benchmarked against her individualized normal state, but underperforms vastly when compared to the rest of the population. In analogy with a wind turbine park, one can tune a poorly performing windmill towards the average mills in that park, instead of bringing it back to its twin's definition of regular performance. Or even more, one could decide to take measures to get it to the best performing mills in the park. So, even given the high-resolution picture on normal performance that can be derived from Digital Twins, the distinction between maintenance and upgrade crucially depends on the reference or baseline that is chosen, so that this distinction contains an important normative element. Likewise, as it has often pointed out (Hofmann, 2017), the distinction between therapy and enhancement will not result only from a detailed observation of the state-of-affairs but also from their interplay

with the realm of language and meaning. ADHD for instance has only been categorized as a diseased state in recent times (Lange, Reichl,, Lange, Tucha, & Tucha, 2010). Nature does not come with clear categories, and is often is characterized by gradients rather than by crisp clear joints at which one conceptually can cut. As pointed out along these lines by Bostrom and others, the concept of "disease" may not refer to any natural kind and depends on the perspective taken (Bostrom, 2008). In a promiscuous realism perspective, the human interest together with the patterns found in Nature will determine where one will cut (Dupré, 1993). Along these lines, categories like therapy and enhancement do not solely reflect patterns found in the data of patients. They also reflect our normative interests and conventions.

However, this does not imply that the (normative) choice of the baseline will be completely arbitrary or that the states of affairs are totally irrelevant for it. Some important moral distinctions are in fact rooted in, or depend on the state-of-affairs (Burms & Vergauwen, 1991). Speciesism for example is a moral distinction that attributes a special moral status to humans over other species. This distinction cannot be derived though from criteria like self-consciousness or capacity for suffering. Our moral intuitions concerning mentally handicapped thus are based on the fact that they belong to the human species, rather than on the fact that they lack standard cognitive abilities. Another example is incest. If Oedipus would have had a genetic testing kit, he would have been horrified when sleeping with his mother and killing his father. But in the absence of this knowledge of his hereditary links to both people, he committed both transgressions. Equally, eating a dish that afterwards is revealed to contain meat can be an unpleasant surprise for a vegetarian, though she enjoyed the dish while eating it. These cases exemplify that some important moral distinctions are grafted on structures deeply embedded in Nature.

If this is the case, then it is reasonable to expect that in a hypothetical scenario in which high resolution data on genetics, metabolism, life style, etc. is available for persons, and their individualized high-resolution pictures are offered by Digital Twins, we may witness some relevant shifts in the normative baselines to assess what counts as health, disease, therapy and enhancement. Consider, as a first example of a novel classification that already is the result of accumulating personal data, the emerging class of 'asymptomatic ill'. This class consists of healthy people with molecular patterns indicative of a high susceptibility to a disease, though they did not develop that disease yet (Plümecke, 2016). Now, assuming one takes some (medical) steps to prevent the disease to develop, one may wonder whether this intervention would qualify as therapy. Conceptually, it seems unwarranted to define therapy an

intervention done on a healthy individual. In this respect, such preventive care interventions resemble more what from an engineering perspective would be called a maintenance intervention. However, this wouldn't be simple maintenance, due to the specific goal for which it is done. This goal is to prevent one very specific and statistically uncommon malfunctioning or disease to occur via a targeted (medical) intervention, where the occurrence of this specific potential disease has been predicted based on high resolution picture of the individual subject. As it were, the subject is fine, but her digital twin is not OK. On the other hand, defining these interventions as forms of enhancement due to them being done on a (currently) healthy individual and/or due to them being based on information in a digital representation of the subject rather than on her actual conditions and/or being done via complex and costly interventions would not sound convincing either. After all, it is a disease that we are fighting. It may therefore well be that personalized medicine and Digital Twins will force us to stretch or revise our categories. For instance, by accepting the idea of something being a therapy, even if done on a healthy individual based on a critical condition of her Digital Twin, insofar as the intervention is done in order to address a potential illness of the individual which is highly probable to occur. In fact, it is to strike this balance that some already use the apparently paradoxical label of "preventive medicine". Needless to say, this is not only a conceptual but also a moral issue. Depending on whether these interventions are considered as daily care, therapy, or enhancement, different conclusions may be drawn on the question as to what extent and under which conditions they should be provided and their costs covered by a public healthcare system.

A second example of a possible normative shift caused by digital twins would be life extension via anti-ageing medicine. There is a high interest to develop ways to prolong the human life span, as in Google's spinoff Calico LLC or Venter's Human Longevity Inc. The rationale that is often used to support this type of research is a therapeutic one. Preventing diseases by making people growing old in a healthy way is better than curing diseases only when they happen to arise. Lifestyle and genetics already result in considerable differences in life span among people, so one can expect that there are mechanisms that can be engineered in order to extend people's life-span. Some people seem to have a constitution or habits that result in a long and healthy life. With the availability of Digital Twins, such naturally occurring people with an extremely long lifespan might end up in a dedicated medically salient category. If a combination of certain features in genetic makeup and lifestyle as displayed in someone's digital twin would allow us to reasonably predict their life-

span, this would lead to new medically relevant distinctions between healthy persons, even without the presence of enhancement technologies. One would be able to classify a set of people as prone to lead a long and healthy live, and sets of people with normal or with short life expectancies. This medically relevant distinction between persons, again, will be grafted on top of the (statistical) patterns that are found in the population of Digital Twins.[3] Now, let's imagine that, thanks to Digital Twins we come to discover with some precision which life-styles are typical of people in the class of long-livers, for instance a certain diet or a certain regime of physical activity. Let's also assume that based on this knowledge one would gradually manage to move more people into this class. This could be done for instance via the advertisement, possibly the nudge or any other set of psychological or economic incentives to live according to these healthier lifestyles. Again, the question arises as to whether a life extension achieved in this way would count as therapy or enhancement. On the one hand, one may not categorize this as human enhancement. The deviation from the norm can be for the individual, and still be in the normal life expectancy range of the human species as a whole. Moreover, if a group of people on another island starts to live whatever happens to be the life-extending lifestyle and thereby lives longer, this would hardly be considered enhancement. Living a healthy life is the paradigm of a health improvement that does not qualify as an enhancement (or therapy, for that matter). However, one may argue that there is a crucial difference between this scenario and the scenario that involves Digital Twins and an explicit policy of incentives. Here it can be said that a certain individual's or group's life extension has been achieved by design; because of the kind of knowledge provided by the data of the Digital Twins (high resolution, etc.), and because of the systematic, deliberate targeted policy that this knowledge has allowed for. In other words, whereas the means used to achieve life extension – food, physical activity – clearly fall into the field of natural remedies, the broader process of scientific acquisition of data and of (social) design of which they are part may turn the process into a form of engineering, and therefore, arguably, of human enhancement. In fact, if the same group of people would obtain the same life extension effect, but this time because they have the financial means to access some complex biotechnological interventions, intuition would probably lead us to classify this as enhancement. The reason is not merely that such a radical intervention surpasses a normal range derived from the distribution over the entire population. The reason to categorize this as en-

[3] Some ethical implications of such scenarios are discussed in the last section

hancement has to be, first of all, with the explicitly engineering nature of this intervention.

Certainly, the fact that such life extension would be achieved via costly technologies, would also have a symbolical boundary surpassed. It would impact the way we think about humans and ageing in general. It is a vastly rooted principle in human societies that the wealthy and the poor face the same facts of life: they grow old and die. Access to health care, nutrition, housing, etc. evidently can contribute to a longer life. But biologically speaking mortality per se is indifferent from human action. This biological fact is rooted in culture and society since the dawn of mankind. Technological modification of this process would not only result in a biological quantum leap, but also in a quantum leap in meaning. The concept of what it is to be human may fundamentally change (Temkin, 2011). The premise that "all humans are mortal" then will not hold true for all men to an equal extend anymore. Some will be less mortal than others due to technical means, eventually because of their financial means. In this case, the transgression that determines whether a modification is an enhancement therefore is also a transgression in the domain of meaning, that is grafted on a technological modification of biology. This fact holds true whether or not it concerns radical transformations, although radical transformations probably carry a higher likelihood to affect existing symbolical distinctions more harshly. This last point takes us to the ethical dimension of Digital Twins.

3.6. Digital Twins and the ethics of human enhancement

So far, we have used human Digital Twins - the assumption that one is in the possession of a data magnifying glass, that gives a detailed account of the molecular, phenotypic and life-style history of persons - as a conceptual tool to understand an existing trend in medicine, and to start a reflection on the potential conceptual implication of this trend on our understanding of the categories of health, disease, and enhancement. In this last section, we use Digital Twins to explore some possible ethical and societal implications of this trend.

A popular line of argumentation in favor of the prima facie moral acceptability of human enhancements starts from the observation that humans already use enhancement techniques, albeit low-tech ones. Athletes for instance improve their performance via physical exercise, a special diet, and a regular lifestyle. With the introduction of wearable health monitoring devices this type of improvements becomes supported by real time data from the individual athlete. The improvement

obtained by training and dietary schemes might be the same as the improvements obtainable via pharmaceutical means, both based on these early-stage Digital Twins. The aims and the factual outputs are similar, maybe even at the molecular level, which might lead to the welfarist position that therapy and enhancement are equally acceptable means to increase welfare (Giubilini & Sanyal). As outlined above though, the acceptability of the approach is not merely rooted in the data, but in the distinctions made at the level of meaning. Human enhancement achieved via technological means or programs based on Digital Twins may be seen as specifically problematic because of this. By using pharmaceutical means, an athlete will transgress a certain symbolical boundary that is institutionalized in her sport for a long time. It is exactly the transgression of this symbolic boundary that makes the athletes act problematic, not merely the result in performance. Let's assume that a society rethinks a marathon, now entailing the usage of pharmaceutical means to boost runner performance. One might consider the resulting contest as morally acceptable if no transgression at the level of meaning would be involved. But the participants of this activity would engage in something that is different from what we now call a marathon. The constitutive rules are changed. We could also think about introducing a rule in chess (and leave other rules unchanged), that allowed a night to jump twice in one turn. Since many human activities are defined by their point and meaning and embedding in a practice that is governed by formal or informal rules, they would engage in a very different type of activity (Whitehouse, Juengst, Mehlman, & Murray, 1997) (Santoni de Sio, Robichaud, & Vincent, 2014). This is a general point that goes beyond the sport example.

Egalitarian concerns constitute one of the main bioconservative arguments to caution enhancement. The fear is that human enhancement technologies might lead to different classes of people, and therefore have a disruptive effect on our democratic institutions (Fukuyama, 2002). Along these lines, human enhancement technologies can be thought of as increasing the already existing diversity among human beings. People already differ in strength, health, intelligence or longevity. When such differences would be available as quantified properties in a person's digital representation and available to the entire community for consulting, that evidently in itself carries the danger of discrimination and of the constitution of novel classes. This may create a crucial complication for the realisation of the ideal of human enhancement as a social equaliser. Consider, for example, cognitive enhancement. Enhancers, unlike natural talent and capacities, would be at least in principle be available to everybody in the same way. One therefore can argue that

enhancers are potential social equalizers, counterbalancing the individual differences that are randomly assigned by the natural and social lottery (Savulescu, Foddy, & Clayton, 2004). However, it turned out that individual differences matter also for the functioning of enhancers (Husain & Mehta, 2011). This doesn't necessarily mean that cognitive enhancers may not work for certain category of people (though it may well be the case). But it certainly means that a big quantity of individual data is needed to fine-tune the treatment or the enhancement. Digital Twins have therefore great potential to make enhancements more precise and effective, if the assumptions behind personal medicine prove to be correct. This holds true not only for cognitive enhancement, but for all sorts of therapy and enhancement. This necessity of acquiring a massive amount of data about the individuals may introduce new issues of equality that may counterbalance the desired equalising effect.

It hints at the fact that not the enhancements themselves, but rather the sheer availability of a vast amount of data like those of Digital Twins coupled with the human tendency to attribute meaning to patterns in data may make give more concerns for equality.

This immediately brings up the importance of privacy as an instrument to mitigate these effects. Privacy concerns that were raised in the context of genomics will be even more relevant in the case of Digital Twins, since the combination of multiple layers of biological and behavioral data will be much more telling about a person than genomics data alone. Given also the engineering analogy that is closely related to Digital Twins, privacy will be instrumental in avoiding that persons will be on a same par as designed objects, vis a vis their twins. In other words, privacy will avoid blunt comparison of human digital twins and therefore the grafting of symbolical distinctions on top of these data. However, this may create a trade-off or even dilemma between equality of capabilities versus equality of privacy. In order to grant everyone access to medical treatments, distributing pills or medical devices may not be enough. In this 'virtual patient' scenario is a prerequisite to collect everybody's data and to create a Digital Twin for everybody. Personalized medicine will arguably increase the cost at the individual level, when compared to off the shelf pills. Next to that, there will be differences in people's capacity to protect their data, due to differences in information about the risks, and differences in their contractual position in the "negotiation" about the use of their data. This is a concern, for the standard reasons about (medical) data protection (van den Hoven M. , 2008). But it also raises a new, specific, issue. Bioconservative fear of a class of biologically privileged persons might realize without any technological intervention; the mere

existence and knowledge of one's Digital Twin may create discrimination of the real people of which the twins are a digital representation. Self-fulfilling prophecy mechanisms similar to the ones active in the financial sector can come into play: the mere fact that other people or institutions think that you are going to be sick or weak or short-lived may make you sick, weak or short-lived. Much in the same way in which the mere fact that you are thought to be insolvent may eventually leave you broke. This marks an important difference between the use of Digital Twin in engineering and in medicine. The social and symbolic dimension in the human realm create a new layer of complication and potential ethical issues. A Digital Twin for a human maybe not only a powerful tool to improve one's physical condition. It may also be a second self who can – metaphorically speaking – rise up against its biological counterpart; or, more prosaically and realistically, being the source of serious moral damage for the real person. In this way, it may be the case that the only way to achieve equality of capabilities would be by creating data which may in turn be used to penalize some groups or to create new forms of discrimination. Governance mechanisms for safeguarding the rights of persons that have digital doubles therefore will be crucial. Such governance mechanisms can draw from how for instance biobanks or medical databases are designed, regulated, inspected, etc. The governance structures should for instance ensure transparency on how the digital doubles are used, protection of the data, and a fair distribution of the benefits derived from people's personal biological information.

The engineering approach that is inherent to Digital Twins also sheds a new light on current health care values, and opens the route to a whole new range of values. In current health care, where in most cases only a low resolution picture of the patient is available, regular health care values that apply are autonomy, beneficence, non-malfeasance and justice (Timmermans, Zhao, & van den Hoven, 2011). All these values will face different concretizations in case Digital Twins become available. Distributive justice for instance will be challenged due to the high resolution with which one can suddenly identify differences in constitution and capabilities among people. It will sharply raise the question on which conditions are to be treated in order to compensate for bad luck in the natural lottery. The value of autonomy will have to be implemented in view of a strong dependency of a digital model, and patients will have to develop a proper relation towards their personal Digital Twin.

With the availability of detailed molecular data of novel engineering methods to impact biological systems (e.g. engineering germlines or somatic cells via Crispr-

CAS), a whole range of values need to be decided upon. Examples are the efficiency of the engineering actions, the effectiveness of the design, the competitiveness of the design versus other designs. The question is then which enhancements to favour, and how to make the engineering decisions. Engineering in general requires decisions on which values to include in the design or the optimization of a system, and which values to maximize (van den Hoven, Lokhorst, & Van de Poel, 2012). Value-Sensitive Design approaches in engineering make explicit which values are implied in the technical development of an artefact, and try to overcome moral dilemmas by design. Given the analogies with engineering, this approach can also provide relevant insights in the field of personalized medicine and Personal Digital Twins. The trade-off between equality of access to (personalised) medicine and risks of data-based discrimination is one example of a challenge that Value-Sensitive Design may face in this domain.

Next to this, the results of medical engineering actions are intrinsically positional, as they are in the economic context of engineering artefacts. Individuals can aim at enhancements with personal flourishing as underpinning motif (e.g. ability to even more enjoy their swimming experience), but more likely will be driven by competitive motifs (outperform others that score less on the swimming property). Digital Twins may also lead to an impoverishment, by focusing on certain traits and neglecting others.

Rationality has limits, and this point is often pivotal in bioconservative perspectives on human enhancement (Giubilini & Sagar, 2015). Reason proves to be an instrument with very limited capabilities when it boils down to predicting the future. Predicting the consequences of radical enhancements is therefore merely impossible. It even proved to be difficult to assess the demographic effect of simple and uninvasive technologies like the prenatal determination of a child's sex (Fukuyama, 2002). The Brussels' philosopher Hottois stressed the point that our complex bio-physical world brings about the future, and that these dynamics can only be captured to an extremely limited extend via reason and via our systems of language and meaning (Hottois, 1996). In this perspective, one cannot fully anticipate the future impact of current human enhancements, whether they are disruptive or gradual. This lack of long-term predictability not necessarily implies that enhancement actions should be banned. One can accompany the process of making bio-physical modifications with deliberation about meaning, value, risks, etc. Since Digital Twins are in between the bio-physical world and this world of language and meaning, they constitute an important technical platform for enabling such techno-scientific

accompaniment. The data in Personal Digital Twins reflect the operational character of reality. These data are read-outs of the metabolic composition of the blood at a given point in time, the genomic code, the history of blood pressure and of physical movements of the body, and so on. As such, these data are an intermediate stage between the operational realm of the biophysical reality, and the realm of symbols, language and meaning. Availability of these data provides us with a substrate to graft symbolical distinctions and meaning on structures that are present in the bio-physical world. Digital Twins, be it as conceptual tool or as emerging technology, can therefore be a tool for moral accompaniment of technological evolutions. They can be one element, among many others, in an effort to realise a Responsible Innovation in this domain. They can aid both in understanding and in shaping the continuous interactions between engineering actions in the bio-physical world, and the world of values and meaning.

3.7. Chapter conclusions

The Digital Twins concept provides a solid thought instrument to analyse conceptual and ethical aspects of human enhancement. It does so by putting enhancement against the backdrop of individualized high-resolution data of people's molecular constitution, physiology, lifestyle and dietary habits. Next to that, Digital Twins are an emerging field in medicine, that has the potential to become the playfield where therapy and enhancement are explored. Comparison between Digital Twins in entire populations allows to get a much sharper idea on health versus disease, and by consequence sharpen the debate on therapy versus enhancement. Digital Twins also have the potential to be a rich source for identifying novel and effective engineering routes, both for therapy and enhancement. The engineering paradigm inherent to a Digital Twins based health care will raise novel ethical, legal and social issues for therapy and enhancement. Digital Twins for instance can challenge equality, even without the application of enhancement technologies. The differences between persons can be sharply defined and made extremely transparent based on the differences in their compiled information, leading potentially to segmentation and discrimination. Personal Digital Twins are an asymptotically data-intense scenario that clarifies the importance of governance concerning the production and use of personal biological and lifestyle data.

4 When Nature Goes Digital: Routes for Responsible Innovation

4.1. Chapter abstract

Digitalization of biological populations and ecosystems changes our relation towards them. *In silico* representations of biological systems make them available as resources that allow for novel ways of deriving economic value. These extracted data and models also open novel routes for responsible innovation based on biological systems and derived biological data. Responsible innovation based on natural resources is explored using the common pool resource framework and using the emerging field of biodiversity sequencing as an example. Biological systems that have a vast digital representation which is shared by a community have aspects from both a natural resource commons and from a knowledge commons, but differ in their structure and dynamics. We therefore propose the concept of "Twin Commons": the institutional arrangement of natural resources that have a tightly linked digital component which is shared and governed by a community, and that have research and innovation as important outlets.

4.2. Chapter introduction

Digitalization opens up biological populations and ecosystems for human usage in novel ways. As indicated by Hess and Ostrom, the mere fact that digital technologies allow capturing aspects of biological systems that were previously uncapturable, "creates a fundamental change in the nature of the resource, with the resource being converted from a nonrivalrous, nonexclusionary public good into a common-pool resource that needs to be managed, monitored, and protected, to ensure sustainability and preservation" (Hess & Ostrom, 2007).

Commons are the institutions that govern the creation and use of common pool resources – resources that are shared by multiple stakeholders and subject to social dilemmas (Ostrom, 1990). In this paper, we ask what type of commons can result from the digitalization of biological systems, and how these commons can contribute to responsible innovation based on (common pool) digitalized natural resources. The concept of common pool resources originated from the study of natural resources that are managed by a group of stakeholders and that are prone to social dilemmas

(Ostrom, 1999). Fishing grounds and forests are archetypical substrates of such 'natural resource commons'. The members of a community extract, for instance, fish or logs from the common pool resource, which can lead to depletion of the resource if not managed. Institutes around biophysical resources that have a digital correlate have been characterized as research commons (Dedeurwaerdere, Melindi-Ghidi, & Broggiato, 2016), or as knowledge commons (Strandburg, Frischmann, & Madison, 2017). The first perspective targets biological data and related samples in a mainly academic research environment. The knowledge commons perspective focuses on biological data as the resource that is prone to social dilemmas.

As an alternative analysis, we propose to introduce the concept of Twin Commons: commons in which the natural resources consist of a (bio)physical reality and of a digital extract that represents aspects of this reality. At the conceptual level, this concept helps to gain more clarity about the nature and the dynamics of digitalized natural resources and their related communities, and their difference from already existing categories of commons (i.e. natural resource commons, knowledge commons, digital commons) (Section 2). Twin commons build on the concept of 'innovation commons'. As such they can shed light on innovation in natural resource digitalization initiatives (Section 3). A twin commons framework is proposed to cover the main aspects in public natural resource digitalization initiatives (Section 4). Venues for responsible innovation in twin commons are explored in Section 5. In section 6, we discuss twin commons in the context of the biodiversity sequencing field, more specifically in the context of the setup of the recently initiated Earth Biogenome Project (Lewin, et al., 2018)[4].

4.3. Digitalized Natural Resources as Common Pool Resources

Natural resources, biological populations and ecosystems increasingly have a digital correlate. Molecular readout technologies now allow for characterizing the genomes of large populations, be it of humans, cattle, crops, wild plant species, up to microbes and fungi. Aerial imaging techniques allow for building high resolution and dynamic

[4] This project was launched in November 2018 and aims at sequencing all eukaryotic biodiversity within a decade. This endeavour requires a massive sample collection and sequencing initiative, encompassing multiple institutions and research communities from across the globe. The data will be made available as a global resource for research and innovation. The project therefore will need to develop institutions to deal with commons dilemmas.

pictures of fishing grounds, geological structures, and land usage. The results are large quantities of geospatial, genetic, biochemical and phenotypic data, residing in a plethora of data repositories. Access to these data not only allows studying these biological systems in innovative ways, it can fundamentally change the nature of these systems, since they become available as 'resources' that can be mined (*in silico*, and as a consequence also in vivo) and utilized for human purposes. In biological systems that are already used as resources, digitalization can trigger a fundamental repurposing. Digitalization thereby generates substantial risks. It facilitates biopiracy, a situation in which information on natural compounds is used to file patents or develop products, without consent or a proper compensation of the communities that have the traditional knowledge about the related species, nor a contribution to the biotope's sustainability. Digitalization also paves the way for potentially disruptive technologies, such as for instance the application of gene drive technologies for engineering ecosystems. Digitalization of biological systems also creates substantial opportunities. Biodiversity conservation can for instance benefit from a deeper knowledge of biological diversity at the genetic level. Economic and societal opportunities lie in innovations that can be derived from the bio-data.

Shared natural resources, whether digitalized or not, are subject to problems like congestion, free riding, conflict, overuse, pollution and degradation, commodification or enclosure, and non-sustainability (Hess & Ostrom, 2007). Classical economic theory predicts a 'tragedy of the commons' (Hardin 1968) as the bleak outcome of these social dilemmas. In this scenario, stakeholders collectively overuse and eventually deplete the resource in their rational pursuit to maximize individual interests. In practice however, local communities seem to be able to circumvent this grim prediction without having to revert to privatization of the resource. They do so by installing rules that organize the interactions among the stakeholders, and between the stakeholders and the resource. Ostrom termed these shared resources "Common Pool Resources" – and defined them as "natural or man-made resource system that are sufficiently large to make it costly (but not impossible) to exclude potential beneficiaries from obtaining benefits from its use" (Ostrom, 1990). Natural resources have been framed as common pool resources when a multitude of stakeholders is involved in their management and usage. The term 'commons' originally refers to the common grounds in medieval Europe, where villagers were allowed to harvest fruits, wood, and graze their herds. Commons imply social dilemmas: conflicts between individual's rational behavior and the group's optimal outcomes (Poteete, 2010). Ostrom and colleagues studied a variety of natural resource commons. Eight

institutional design principles were derived that are associated with long-during commons: clearly defined boundaries, rules that are tuned to local circumstances, participation in the rule setting and modification, monitoring of the appropriators and graduated sanctions in case of violations, conflict-resolution mechanisms, recognition of the right to self-organize by external institutions, and the organization of key activities in the commons as multiple layers of nested enterprises (Ostrom, 1990).

Digitalization can substantially affect the dynamics of natural resources and the related commons by creating a second digital common pool resource that is interlinked with the natural resource. The common pool resource concept was also applied to this "new shared territory of global distributed information" (Hess & Ostrom, 2007). The main entry point was that knowledge is increasingly co-produced, co-managed and co-used by a multitude of stakeholders. Especially with the advent of the Internet it became obvious that a digital equivalent of the known biophysical commons was in the making. Digitally stored information for instance provides the commonly shared knowledge grounds for scientists and technologists. These "Knowledge Commons" or "Cultural Commons" revolve around common pool resources consisting of pieces of information, rather than of matter (Madison, Frischmann, & Strandburg, 2010). They were defined as "the institutionalized community governance of the sharing and, in many cases, creation of information, science, knowledge, data, and other types of intellectual and cultural resources" (Strandburg, Frischmann, & Madison, 2017). Knowledge commons have some distinct features that set them apart from natural commons. Firstly, they are inherently global, since asymptotically information and knowledge are not constrained by the boundaries of a specific local community. Secondly, they concern non-subtractable resources. In contrast to for instance a fishing ground, a knowledge common pool resource cannot be depleted since knowledge does not decrease when being consumed. Nevertheless knowledge commons require a joint effort by the community of stakeholders in order to ensure production and availability of data and knowledge, to manage access rules, foster correct usage, etc. This non-subtractability makes knowledge also non-rivalrous, though the production of knowledge may depend on rivalrous input like time and money, and lead to rivalrous output like money or fame (Strandburg, Frischmann, & Madison, 2017). In contrast to natural resource commons, knowledge commons need to be created before they can be shared (Strandburg, Frischmann, & Madison, 2017).

The knowledge commons concept was also applied to biological systems that obtained a digital representation: genetic data (Pálsson and Prainsack 2011), microbiology data (Uhlir, 2011) and to various types of medical data (Strandburg, Frischmann, & Madison, 2017). This is not surprising since biomedical and biological sciences have become data intensive disciplines. Powerful digitization techniques such as next generation sequencing and medical imaging allow for massive readout of data from biological systems. These technologies result in large amounts of biological data, residing in repositories and made accessible via Internet technologies. Framing these systems as knowledge common pool resources provides an alternative to schemes where the data or the knowledge is considered as either public or proprietary. This perspective also implies an institutional component, rather than exclusively focusing on the individual actor, invention, or piece of information. Reframing the analysis in this broader context provides a more comprehensive view when analyzing processes like innovation and decision making in these complex environments. Knowledge commons have been broadly defined in this context, to cover not only knowledge but also the sustainable management of information and data (Strandburg, Frischmann, & Madison, 2017).

Importantly, digitalized natural resources must be analyzed neither as pure natural resources nor as pure digital resources. Rather, the commons revolving around these resources constitute hybrids made of a knowledge commons and a natural resource commons, having distinct dynamics and requiring a specific type of analysis. For instance, the production or read-out of biological data directly and intimately relates both to the biological samples as well as to the populations or ecosystems they originated from. Biological systems have been characterized as data repositories of sorts, in which massive amounts of genomic and molecular data are 'stored' and available for 'readout'. The human genome for instance has been termed in this context a natural resource, and medical genetics an extractive industry that extracts information and uses this to develop health care products (Evans, 2014) (Strandburg, Frischmann, & Madison, 2017). Analyzing these systems as mere knowledge commons would incorrectly imply that only the knowledge aspects are to be considered and are of value. Such framing runs the risk of reducing digitized biological systems to their knowledge components, thereby undervaluing and obscuring the underpinning populations or ecosystems.

4.4. Natural Resources as Substrate for Innovation

Digitalization of natural resources aims at knowledge generation and at deriving innovations with market value and/or societal value. Often large-scale data generating endeavors aim at deeper scientific understanding of the targeted biological systems. Digitalized natural resources therefore have been characterized as the subject of research commons. The research community dealing with the exchange of information and samples on microbial strains has been for instance described as a microbial commons (Dedeurwaerdere, Melindi-Ghidi, & Broggiato, 2016). Next to research, many large-scale initiatives also explicitly seek to foster a flourishing innovation landscape around the generated scientific data. For example, the genomic data gathered in the Genomics England initiative is intended to be used to develop predictive medicine practices (Marx, 2015). The creation of a silicon-valley like cluster around Amazonian biodiversity has been proposed as a route to a new biodiversity based economy (Nobre, et al., 2016). Such initiatives therefore aim at the fostering of high levels of data-driven innovation. Data-driven research differs from settings where the data are generated in order to investigate dedicated research hypotheses. This type of research focuses on exploratory rather than theory-driven experimentation (Pietsch, 2015).

Innovation commons were recently proposed as a special case of knowledge commons (Potts, 2018). Innovation commons were defined as institutions 'to facilitate cooperation and supply governance among a group of technology enthusiasts in order to create, under high uncertainty, a pooled resource from which the individual members of the community might seek to discover and develop entrepreneurial opportunities for innovation' (Potts, 2018). The focus thus is on the peer production of information in order to derive business opportunities. Innovation commons bring the data, tacit knowledge, technologies and stakeholders together, thereby providing the conditions to spot entrepreneurial opportunities. In the innovation commons outlined by Potts, the commoners gather around the question how to transform an idea into an innovation. The preconditions for such innovation commons thus are a new idea, invention or technology; distributed information and tacit knowledge that are related to the idea; and uncertainty about whether that idea provides a concrete entrepreneurial opportunity – i.e. the knowledge problem that is inherent to this early innovation stage (Potts, 2018). Examples of innovation commons are open source software initiatives and hacker spaces that focus on novel technologies like blockchain and synthetic biology. Innovation commons are efficient in minimizing the transaction costs of discovering entrepreneurial opportunities present in the

community and its resources (tacit knowledge of its members, data, technologies). They thereby solve a collective action problem: given the uncertainty that investments in this early phase effectively will yield viable innovations, the innovation commons thrive on the rules and control mechanisms that are set by the community itself. More formal organizations are claimed to be less efficient in managing the transactions in this early stage of innovation.

Communities around digitalized natural resources meet the preconditions for an innovation commons. They concern expert knowledge about the generated data: how the data are structured, what already is known about the data, what potentially can be derived from the data. This tacit knowledge is distributed across various experts and proto-entrepreneurs, that form a community around the given topic and datasets. Digitalized natural resources, however, provide a different flavor of innovation commons. Innovation communities centered around biological data are often long-standing and imply professional roles, as for instance in microbial research commons (Dedeurwaerdere, Melindi-Ghidi, & Broggiato, 2016) or the genome commons (Contreras, 2014). This in contrast to the hacker spaces that provided the archetype for Potts' innovation commons. Hackers spaces are short lived and populated with technology enthusiasts. In contrast, the Human Genome project for example ran for over a decade, steadily releasing new data in the genome commons. The commons related to digitalized natural resources are not purely pre-entrepreneurial, pre-firm nor pre-market. Data are generated over years and thereby give rise to a continuous valorization stream. The shared resources in innovation commons are the technologies and the knowledge that are required to derive innovations from them: knowledge about the market, about how to set up a successful enterprise around an innovation, about regulatory boundaries, etc. The shared resources in digitalized natural resource commons also encompass the natural resource and the data resources that are derived from it. The microbial research commons for instance consist not only of shared microbial sequence data, but also of microbial collections (Dedeurwaerdere, Melindi-Ghidi, & Broggiato, 2016). There are therefore clear differences with the innovation commons model as described by Potts. The question thus is how to describe the commons that result from the digitalization of biological systems in a way that accurately captures their innovation dynamics. In the next section we attempt to do this by building on the previously described commons concepts.

4.5. A hybrid concept: Twin Commons

Digitalization of biological systems concerns three interconnected resources. At the basis is the digitalized biological system, which becomes available as a natural resource. The derived biological data constitutes a knowledge resource. The conglomerate of entrepreneurial ideas, technologies and related knowledge about technical possibilities, and tacit knowledge about the data in general can function as an innovation resource from which opportunities can be identified. Each of these resources are prone to collective action problems since they are shared among multiple stakeholders. Neither markets nor state interventions alone are able to solve these issues of free riding, excessive resource usage, etc. For that reason, one can see that many natural resource digitalization initiatives in practice have the characteristics of commons (Strandburg, Frischmann, & Madison, 2017) (Dedeurwaerdere, Melindi-Ghidi, & Broggiato, 2016). As indicated in previous section, digitalized natural resource commons have characteristics of natural resource commons, knowledge commons, research commons and innovation commons. A commons concept that contains elements from all aforementioned types is thus required.

We propose 'Twin Commons' as a framework to denote the institutional arrangement of natural resources that have a tightly linked digital component which is shared and governed by a community, and that have research and innovation as important outlets. Twin commons comprise a twin resource (a natural resource and its digital representation, linked via technologies) that is managed by a group of stakeholders as a common resource (see Fig. 1). These elements provide an environment where innovation can take place. Twin commons consist of the following five components:

(1) (Bio)physical system: the natural resource, the particular ecosystem, or the population of biological individuals that are in scope as resource to be managed by the group of stakeholders.

(2) Digital and knowledge counterpart: the digital representation of the (bio)physical system. This can be the collection of genomic sequences derived from samples and related inferences about their relations to phenotypes, models of biological mechanisms and pathways, theories about potential applications, etc. This digital model represents certain molecular and phenotypic aspects of the biophysical system.

(3) Bridging technologies: technologies that provide a bridge between the (bio)physical system and its digital representation. This comprises the set of already available technologies that allow for deriving data from the bio-physical system (e.g.

genome sequencing technologies, sensor data, computational methods to analyse the data), and the set of technologies that allow to change the bio-physical system based on the digital representations (e.g. genomic editing technologies like CRISPR-Cas).

(4) Social system: the system that manages the natural, digital, and innovation resources as a commons. Along the lines of the Institutional Analysis and Development framework (Ostrom, 2005), one can further distinguish the rules, the action arena and the attributes of the community. Concerning the attributes of the community, its boundaries are defined by topic, rather than by physical location as is the case for natural resource commons. Often the data are produced, managed and used by a variety of stakeholders. Submission, management and usage of the data are subject to some of the social dilemmas that are characteristic of a commons. Researchers that contribute data to the resource want to avoid the free-rider problem, and ensure that they obtain what they see as the proper benefits from their efforts. Successful commons often develop rules to efficiently identify and sanction free-riders (Poteete, 2010). Rules, be them formal or informal, have an impact on the interactions between the actors. Data submission and access rules for instance vary across data repositories and demarcate the boundaries of the commons.

(5) Innovation common resources. Twin commons provide the raw material and the social structures that help enable innovation: the pool of shared or tacit knowledge about technologies, data, and related innovation opportunities, together with the access to the natural and digital resources.

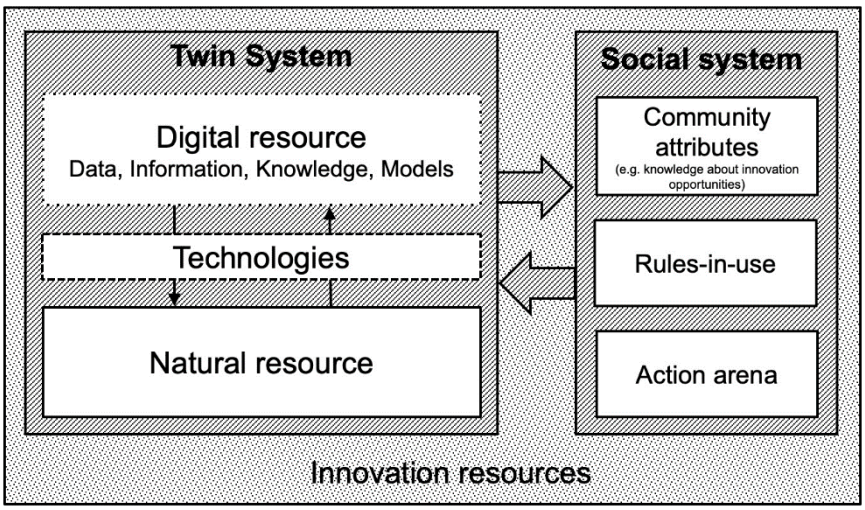


Fig.1. Components of a Twin Commons. The components in the Institutional Analysis and Development framework (Ostrom, 2005) are presented in an aggregated way under 'Social system', and put in relation to the resource component in the commons (the biophysical system; its digital representation; and the technologies that mediate the relation between both). Innovation resources comprise all the components in the twin commons.

The innovation aspect of twin commons has characteristics of the innovation commons as described by Potts (Potts, 2018). Twin commons bring together a heterogeneous community of funders, researchers and entrepreneurs that together construct innovation opportunities (including both technical and business opportunities). Twin commons do this specifically by bringing the stakeholders together around a scientific data gathering initiative. Access to a natural resource and to digitalization technologies is a prerequisite for constructing the pool of shared biological data. This interplay between the biophysical, the informational, and the societal aspects of natural resources is key to understanding the innovation dynamics in twin commons. Given the fact that biological systems evolved over billions of years, biological datasets can only shed light on a very tiny fraction of the complex biological reality. The primary source of innovation opportunities is in the biological system. Focusing solely on the knowledge component therefore would miss out on the importance and the inherent value of the biological system in the innovation process. Twin commons tend to be long standing. Construction of the shared pool of

biological data and knowledge takes time and resources. Twin commons therefore are not entirely pre-market. Tech transfer departments and public-private partnerships are often involved. The commoners can include amateur enthusiasts that look for business opportunities, but often consist of professional scientists, business developers, program managers, etc.

In short, twin commons in biology comprise digital information (for instance sequences and their annotation); the related biological populations, ecosystems, and the derived biological material; and a community that uses these resources in the identification of innovation opportunities and that applies self-governance to sustainably manage the commons pool resources. Examples are ubiquitous in biological sciences, though the focus is often put on the knowledge commons aspect. Human genomic data was for instance framed as a resource for knowledge commons (Evans, 2017). The tight link between human populations, derived genomic data, and innovation potential calls though for a setting where all elements are taken into account, in order to avoid usage of genomic data without a contribution to those populations (Hardy, Séguin, Ramesar, Singer, & Daar, 2008). A similar situation can be observed with respect to biodiversity data, where a social production model for both the sharable goods and the derived information was suggested. In the studied case, the microbial data commons hinge on the materials commons, though the impact of regulators on sharing behavior substantially differs in both domains, with a higher reluctance to share data than materials (Dedeurwaerdere, Melindi-Ghidi, & Broggiato, 2016).

4.6. Responsible Innovation in the Twin Commons

Innovation constitutes an intrinsic part of many large-scale biological data generating efforts. The vast impact these innovations can have on society and on the natural world sharply raises the question of responsibility. What are the venues for responsible innovation in a twin commons? Responsibility is a rich concept that goes beyond the attribution of blame for undesired events, and also encompasses a virtue ethics component. Interpreted as such, responsibility is a character trait of members in an innovators community (van de Poel & Fahlquist, 2012). Moreover, it has been argued that responsibility is not something that needs to be added to innovation, but provides both the very foundation and motivation for innovation (Bergen, 2017). These foundational and motivational relationships between responsibility and innovation are, on the one hand, already visible as inherent aspects of twin com-

mons. Responsibility plays an instrumental role in the management of every common pool resource. It provides a necessary complement to rational self-interest of individual stakeholders and thereby is instrumental in avoiding a tragedy of the commons. Experimental research showed that participants in social dilemma games are willing to give up part of their monetary benefits, in order to get the capability to punish players that misbehave. In successful commons, irresponsible behavior towards the commons or towards the community of stakeholders is carefully monitored and sanctioned (Poteete, 2010). In the case of twin commons, collective responsibility may take the form of implicit or explicit rules on for instance data release policies, data quality, extraction of intellectual property. Collective responsibility may also play an important role in constructing the innovation commons. These projects often aim at contributing to solutions for wicked problems, providing an overall focus on for instance sustainability and biodiversity conservation, or the feeding a growing world population.

On the other hand, tragedies of the commons nevertheless occur. Bergen thus points to a third relationship between responsibility and innovation in which "the outcomes of innovation form the structures through which we can actually work towards justice" (Bergen 2017: 362). The question is how one can better ensure societal value by the deliberate inclusion of values in the innovation process, so that technologies evolve in a direction that is beneficial for society as a whole, and negative outcomes are avoided. Responsible innovation has been proposed as a conceptual framework to enable this. In a broad sense, it is defined as "taking care of the future through collective stewardship of science and innovation in the present." (Stilgoe, Owen, & Macnaghten, 2013). The emphasis of collective stewardship can also be found in the definition of von Schomberg (von Schomberg 2013). Other definitions focus on the translation of value considerations into functional requirements in the innovation process. Moral evaluation of the consequences of the possible actions during innovation should lead to requirements that are included the innovation process (van den Hoven, et al., 2013). In short, the rules, interaction mechanisms, and values that characterize a twin commons, can provide the context for innovation to occur in a responsible way.

Theories of action in commons (Poteete, 2010) can provide a framework to spell out the notion of collective stewardship in responsible innovation. The Institutional Analysis and Development developed by Ostrom and others (Ostrom, 2005) allows analyzing stakeholders and their relations in a commons and has already been linked to responsible innovation (Kuzma, et al., 2017). Design principles for a sustainable

management of common pool resources also apply to commons where innovation assets (biological samples, data, knowledge, etc.) are a shared resource. Innovation commons for instance require mechanisms to avoid that free riders file intellectual property based on shared resources. Collective stewardship in a responsible innovation approach often encompasses the involvement of a broader set of stakeholders in the innovation process. In industrial settings, information constitutes a competitive advantage, and stakeholders strongly differ in the power they can exert (Blok & Lemmens, 2015) (Bergen, 2017). In a twin commons, all stakeholders participate in one way or another in the governing of the shared resources. Stakeholder involvement in a commons is therefore not something that is injected in an already existing innovation process, but an intrinsic and structural part of the innovation process.

Alongside technical innovations, twin commons also allow for framing social innovations. Social innovations have not been a primary focus in the responsible innovation literature (Lubberink, Blok, Van Ophem, & Omta, 2017) (Blok & Lemmens, 2015) (Bergen, 2017). Common pool resources, with the focus on solving social dilemmas, provide an established framework to structurally include social innovations in the sustainable management of a natural resource. The interaction with digitalized biological systems challenges the status quo and calls not only for technical but also for social innovations. The national implementations of the Nagoya Protocol for instance can be regarded partially as social innovations. Mechanisms for a fair sharing of benefits needed to be fundamentally rethought – in light of both the natural resources that could be distributed globally, potentially providing significant income from innovation, and the global community that could share information via digital routes.

Twin commons also provide a conceptual instrument to pinpoint more precisely which areas of responsibility to consider in relation to the innovation process. One has responsibilities towards the biological systems in scope, for instance the responsibility to not negatively impact the sustainable management of the resource. Another example is the responsibility to respect the inherent value of the resource during the innovation process. One also has responsibilities towards the derived data and knowledge, for instance the responsibility to contribute to sustainable management of the data resource, the responsibility to guarantee privacy in case of personal data, or the responsibility to contribute information to the data resource. Responsibilities regarding the mediating technologies can include innovating in fields that contribute to the sustainability of the natural resource, for instance improved management of a biotope via smart applications based on sensor data. Responsibilities towards the

community of stakeholders include a fair distribution of the value derived from innovation, especially when commercial initiatives are derived from publicly funded initiatives.

Last but not least, the different components of a twin commons provide an instrument to make explicit what values are at stake in the commons and its innovation processes. Innovations related to twin commons often concern emerging technologies, disruptive innovation, dual use technologies and/or control dilemmas (Collingridge, 1980), (Owen, et al., 2013), all of which can lead to value contestation or neglect. Mapping values at stake in twin commons is an important step toward achieving responsible innovation. Values influence which innovations are pursued when applying data-derived insights. They constitute the moral context in which innovations occur including innovations related to digitalized natural resources. For example, a synthetic biology route was used to produce the antimalarial agent artemisinin using a novel route with the aim to make antimalarial treatment more available for patients in developing countries (Paddon & Keasling, 2014). Similar technologies are also used for the production of enzymes for washing powder, or for the conversion of crop residues into biofuel, aiming at economic benefits but also on a decreased impact on the environment. In some cases, the impact on global natural commons is potentially very high, as in the case of gene drives technologies. Such cases show that values – whether embedded in governance measures or in research ethics – and public deliberation are pivotal (Oye, et al., 2014). The components of a twin commons constitute areas where values impact the innovation process. As such, they can facilitate responsible innovation in cases of epistemic insufficiency where it is difficult to anticipate the impact of innovations (Blok & Lemmens, 2015). For example, embedding of values like the preservation of biodiversity (Lewin, et al., 2018) or more instrumentally data ethics values like correctness and openness of data (Floridi & Taddeo, 2016) provides the context in which innovation occurs.

4.7. Biodiversity Sequencing initiatives – venues for responsible innovation

In this section we apply the twin commons concept and illustrate its relation to responsible innovation by considering the emerging field of biodiversity sequencing, as exemplified by the recently initiated Earth Biogenome Project. The Earth Biogenome Project aims at the sequencing a vast representation of eukaryotes, comprising all known plants and animals (Lewin, et al., 2018). The project is analogous to the Earth Microbiome Project (The Earth Microbiome Project Consortium, 2017),

another large-scale biodiversity sequencing project that charts microbial communities in habitats spread across the globe. Biodiversity constitutes an important source of innovation opportunities (Lewin, et al., 2018) (Nobre, et al., 2016). Efforts to map genetic diversity in all different kingdoms of life, including human genetic diversity, aim at this innovative potential, including the development of novel drugs, molecular biology tools, bioprocesses, and improved agricultural production. Sequencing of varieties of important crops like rice and corn for instance provides a rich resource for breeding (Varshney, Terauchi, & McCouch, 2014). Twin commons can shed light on challenges and opportunities in biodiversity sequencing based responsible innovation. The tension between local ecosystems and biological sample collections versus globally distributed digital sequence information testifies to the importance of the relation between material commons and data commons. Resources (samples, sequence data) in these projects often constitute shared resources that are subject to collective action problems. Next to this, they provide the material for (bio)data-driven innovations that can vastly impact the natural world and society.

Biodiversity sequencing project resources are twin resource systems. The Earth Biogenome Project concerns a natural resource (plant, animal and fungal species from various environments across the globe) and a digital resource (the metadata that describe the sample and measured parameters, the genomic sequence reads, and derived data). The digital resource is linked to the organisms in their ecosystems with technologies (sampling technologies, DNA sequencing technologies, genome editing technologies). Genomic data are not fully analogous to the data found in digital commons (Pálsson & Prainsack, 2011). Genomics data are embedded in biochemical molecules in biological beings, and partly available as representations of nucleotide sequences in data repositories. Sample acquisition provides the bridge between biotopes and digital sequence information. Organization of sample acquisition at a global scale builds upon multiple twin commons in case of the Earth Biogenome Project, by relying on a conglomerate of dedicated initiatives that organize their own sample collection. The Darwin Tree of Life project for example collaborates with multiple organizations, like botanical gardens and academic institutions, to gather samples from the British islands. Nesting of smaller locally funded commons (for instance the Darwin Tree of Life project, focusing on specific biotopes) into larger commons (the Earth Biogenome Project in this case) provides opportunities for a distributed funding of this large initiative, by breaking it down to individually funded local initiatives.

Samples tie data to a certain physical location or specific biological entity: the biotope or the organism from which the data were derived from. This provides a tension with the nature of electronic data, which can be easily distributed globally. Taking locality into account however is necessary to ensure fair access to the data and related benefits, for instance the benefits derived from innovations. Access and benefit sharing is the focus of the Nagoya protocol, which is a global protocol that is implemented on the national level. The protocol's objective is to foster a fair distribution of the benefits derived from biodiversity genetic resources and from traditional knowledge about these genetic resources. This entails the regulation of the access to genetic resources, access to technologies, benefit sharing obligations, and compliance with local regulation (Convention on Biological Diversity, 2010). Brazil implemented its own framework to regulate access to genetic heritage and associated traditional knowledge for purposes of scientific research, bioprospecting, and technological development. A national system of genetic resource management and associated traditional knowledge (SisGen) was put in place to facilitate compliance with the legislation, by supporting the registration of access, shipment or exploitation of genetic material and associated traditional knowledge. The federal government is the recipient of the benefit sharing via the National Fund for Benefit Sharing. The money in this fund is intended for a multitude of purposes, amongst which the sustainable management and conservation via support for indigenous people and traditional farmers. How exactly this fund will contribute to biotope preservation (for instance by preventing deforestation in the Amazon basin), however, needs to be seen.

Considering these as twin resource systems provides a way to bring this in the scope of management. The Amazon Bank of Codes and the Amazon Third Way Initiative (Nobre, et al., 2016) aim to tightly connect the Amazon biotope to its derived data and to the innovation community in order to support the sustainable management of these resources. Blockchain technologies will be used to track usage of genetic data derived from the Amazon biotopes, together with the generated IP assets, and record the provenance, rights and obligations related to the usage of these twin resources (Lewin, et al., 2018).

The twin resources in public biodiversity sequencing projects are in practice managed as commons. Managing the construction and usage of a biodiversity sequencing project implies heterogeneous stakeholders and a proneness to social dilemmas. Biodiversity sequencing initiatives have sometimes been framed as commons. For instance, the communities around microbial databases and microbial biodiversity have been

studied as commons (Dedeurwaerdere, Melindi-Ghidi, & Broggiato, 2016) (Hess & Ostrom, 2007), moreover commons were proposed as a strategy to design this research field (Uhlir, 2011). In the medical field, the knowledge commons framework was applied to human genomics data, population biobanks, cancer biology, neuroscience data, and data on rare diseases (Strandburg, Frischmann, & Madison, 2017). The Earth Biogenome Project (Lewin, et al., 2018) shows that multiple stakeholders are involved in the rule setting and management of both the natural resource and its derived genomic data. Its governance structure will include representatives from government, private industry, civil society, international organizations, private foundations, and participating research communities and organizations.

The stakeholders in a twin commons engage in a complex web of interactions to manage the resource and distribute the benefits. Framing biodiversity sequencing projects as a twin commons emphasizes the (technology-mediated) connection between the biological resources (be it botanical gardens, natural history museums, microbial collections, or real biotopes like the Amazon rain forest) and the derived data and knowledge. Natural resource commons tend to be local and have rather clearly defined community boundaries, while knowledge commons tend to be global. Biodiversity sequencing programs have characteristics of both. In case of the Earth Biogenome Project (Lewin, et al., 2018) a global community of academics and companies will have access to the information. But local communities in for instance the Amazon basin should also profit from the revenue streams that these global parties generate, and from local development of scientific know-how.

Framing biodiversity sequencing initiatives as twin commons also embeds the notion that interactions between the actors in the commons, and between these actors and the natural resource, are mediated by technologies. For instance, in the Amazon Bank of Codes project, blockchain technologies are proposed as a technology to track the relations between the biological origins and their derived data, and to the track the interactions of stakeholders with these data. Such technologies would allow registering both the rights and the duties of stakeholders in a distributed setting, providing an instrument in the practical implementation of the Nagoya protocol. Values embedded in blockchain technologies (the fact that one cannot tamper with the information, and that the information can be globally accessible) thus support the fair sharing of benefits derived from innovations with the communities from which the data were derived. Properly designed technology can support common formation by supporting design principles in Ostrom's commons framework, for instance by lowering the cost of implementing and managing clearly

defined community boundaries, monitoring of the community members, and collective choice arrangements that allow most participants to contribute to decisions (Williams & Hall, 2015).

Digitalization allows for quantification and transparency and can thereby directly support better management (although it could also lead to more intensive exploitation) of a natural resource. In non-digitalized natural resources it is often hard to know exactly how much is extracted from the commons. With the aid of sensors, databases and dashboards one can much better quantify resource usage and make this transparent to the entire community of stakeholders. Next to this, the addition of social software components can provide a binding factor for an otherwise anonymous and global community. Norms and values can be embedded in the structure of such social software environments. Tools can for instance be embedded for monitoring and punishing free-riders, for the appraisal of contributors, for a fair distribution of benefits, and for an open forum for deliberation about rules and standards. The types of commons that result from sequencing initiatives vastly differ. This opens up multiple venues for responsible innovation through attention to funding regimes, the definition of the boundaries of the commons, and the ways participants and other stakeholders are involved in the commons.

Biodiversity sequencing projects target innovation. Biodiversity sequencing projects are typically structured as research communities with valorization and technology transfer outlets. Biodiversity sequence data constitutes a rich source for innovative applications in a broad range of human endeavors: health care, production of fine chemicals and biofuels, bioremediation, agriculture, biomaterials, etc. About half of the FDA-approved drugs that are on the market are derived directly or indirectly from natural products (Katz, 2011). Most of the genetic engineering tools that are available to biotechnology are derived from Nature. Enzymes like reverse transcriptases and restriction enzymes have given rise to entire industries in health care, agriculture and white biotechnology. Human biodiversity at the genetic and the metagenomic level is also a rich substrate for developing innovative diagnostic and therapeutic tools. Population genomics initiatives are mapping biodiversity in entire human populations, in order to develop novel health care approaches.

And yet, the rapid loss in biodiversity constitutes one of the so-called 'wicked problems' of this century. The rate of extinction of species is estimated to increase about five times in the near future as compared to the recent past (Johnson, et al., 2017). One of the main causes is overexploitation by natural resource-intensive industries (Johnson, et al., 2017). Significant loss of natural habitats and related

biodiversity will lead to a definitive loss in natural capital, and eventually to ecosystem instability. Such loss in natural capital would vastly reduce the ability to meet the grand challenges related to the growing world population.

Some claim that developing new economic models that are based on biodiversity could provide a major opportunity to make biodiversity conservation compatible with economic growth (Rodríguez & Sotomayor, 2019) (Nobre, et al., 2016). Such models would rely heavily on biodiversity-based innovation. This strategy is pursued for instance in the "Third Way", proposing to "aggressively research, develop, and scale a new high-tech innovation approach that sees the Amazon as a global public good of biological assets and biomimetic designs that can enable the creation of innovative high-value products, services, and platforms for current and for entirely new markets by applying a combination of advanced digital, material, and biological technology breakthroughs to their privileged biological and biomimetic assets." (Nobre, et al., 2016). The construction of a biodiversity database at the DNA sequence level in this case is an explicit strategy to value the related natural resources in a fundamentally different way, and to transform the old resource-intensive industries and agricultural practices into a new bio-based economy that preserves the Amazon's ecosystem. Commons can facilitate the enablement of such new economic models, by providing alternatives to the classical dichotomy between state control and the market, by clustering the resources that are needed for innovation (biological materials, data, knowledge, ideas), and by providing a context to frame not only technical innovation but also social innovation (via the development of institutions that allow for collective self-governance).

Analysing biodiversity sequencing projects as twin commons offers routes towards responsible innovation. Responsibility during the innovation process has very different flavors across the different biodiversity sequencing related activities. Responsible ways of dealing with innovation in the Earth Biogenome Project focus on fair and broad access to the data, and fair sharing of the benefits derived from innovations. In the synthetic biology community, a strong emphasis is put on openness in the data and the approaches, on a copy-with-pride mentality rather than patenting strategies (Torrance, 2017), and on safety - since this potentially concerns dual-use technologies (Wang & Zhang, 2019).

Inclusion of stakeholders in the innovation process is one of the conceptual dimensions of responsible innovation (Burget, Bardone, & Pedaste, 2017). In biodiversity sequencing projects it can be challenging to install stakeholder involvement in the early innovation steps to clarify the norms and values at stake and to align them

with what is societally preferable. Innovation is inherently characterized by information asymmetries and power imbalances, since innovation is pursued to create a competitive advantage and generate intellectual property (Blok & Lemmens, 2015). Stakeholder involvement during early innovation steps (especially when related to corporate environments) can benefit from the collective self-regulation characteristic to commons. Responsible behaviour can be fostered by deliberate inclusion of values in the structure and workings of the commons. This can be done via the values that underpin the creation of the common resources, for example the values of sustainability and of biodiversity conservation in the case of biodiversity sequencing initiatives.

Values are also inherent to the rules that apply when valuable innovations are derived from the data. Since heterogeneous stakeholders together construct an environment for the identification of opportunities, stakeholder involvement is at play in setting the rules and in the values that are pursued in the innovation commons, rather than in each individual innovation step. The so-called pharmaceutical commons testify to these dynamics (Lezaun & Montgomery, 2015). Deliberation of the values at stake in the commons (via the commoner's community), and inclusion of values in the rules that govern the common, provide two clear venues for responsible innovation in public biodiversity sequencing initiatives. Next to this, Value Sensitive Design (van den Hoven J. , 2013) provides a way to embed values in the technologies used in the twin commons. Value sensitive design comprises the deliberate inclusion of values in artefacts (van den Hoven J. , 2013). One can target the inclusion of values in the artefacts that are the output of the innovation process, as for instance in the development of washing powder enzymes that work at lower temperatures and thereby contribute to sustainability. Another route is to consider the values that are embedded in the technologies that mediate the relation among the natural resource, its data representation, and the community of stakeholders. Values can for instance determine the access to the resources and the boundaries of the community. Blockchain technologies can ensure traceability and validity of the contracts and interactions.

The twin commons framework can be used to map out the community of stakeholders involved in commons-based innovation. When analyzing genomic data repositories as genome commons, the different stakeholders have been described in terms of funders, data generators, data intermediaries, data subjects, the public, data users and scientific leaders (Contreras, 2014). The latter two groups are formally tasked with the scientific aspects of the innovation, but it is clear that all stakeholder

groups contribute to the innovation dynamics. Stakeholder heterogeneity poses a challenge in managing common pool resources (Poteete, 2010) given the asymmetry in investments and benefits.

The twin commons framework can also be instrumental in the clarification of the values to be pursued during responsible innovation. Framing biodiversity sequencing projects simply as knowledge commons carries the risk that social dilemmas amongst the actors are considered to be merely about the information, not also about the ecosystems that underpin the information. Such framing also misses out on the fact that innovation happens not only in the conceptual space, but requires the interaction with the biological system. A twin commons implies that both the natural and the informational components are valued as intricate parts of the resource system. It avoids the virtualization and thereby depreciation of the biological system when dealing with it.

4.8. Chapter conclusions

Biological systems are being increasingly digitalized by wireless sensors, imaging technologies, molecular readouts, and other means. This digitalization makes these biological systems available in unprecedented ways. While the digital resource provides fertile ground for research insights and innovations, digitalized natural resources pose specific opportunities and challenges. The main question explored here is how to provide the right environment for research and innovations—one that develops innovations in a societally beneficial ways while mitigating risks.

In this paper we propose twin commons as framework and related it to venues for responsible innovation. We define twin commons as digitalized biological systems that constitute a common resource managed by a group of stakeholders, have a proneness to social dilemmas, and have research and innovation as their main focus. The twin commons framework builds on natural resource commons (Ostrom, 1990) and on knowledge commons (Hess & Ostrom, 2007). Twin commons display characteristics of both types of commons, since they concern a twin resource: a digital resource that is derived from a natural resource. For instance, ecosystems are geographically local, but the derived data have (tangentially) the character of a global resource. The biological system is a given, while the digital resource is produced. Bringing both together in one framework makes this tension explicit. The twin commons framework also builds on the innovation commons (Potts, 2018). Twin commons are formed in order to provide a fruitful environment for research

and for marketable innovations. Many of the large biodiversity sequencing initiatives for instance aim at the generation of knowledge, but also on the fostering of a biotech-economy. Twin commoners, however, tend to be professionals who gather in long term initiatives, in contrast to innovation commons (modelled on hacker-space communities) which tend to be short lived and composed of enthusiasts pre-market (Potts, 2018). More empirical research is needed to analyze the variety of public natural resource digitalization projects, for their stakeholder composition, community structure, spatial and time component, and innovation dynamics. As an outlook, cross learning over the diverse biodiversity sequencing projects can be highly valuable. Along the lines of common pool resource analysis executed on natural resource commons by Ostrom and others, one can expect to derive characteristics of successful and responsibly innovative commons.

The twin commons framework can be instrumental when pursuing the responsible digitalization of natural resources. One aim of responsible innovation is to align innovations with the values that are preferred by stakeholders. Twin commons provide multiple venues to assist in reaching that goal. They provide an instrument to map the stakeholders and the interactions and rules that are set amongst them. They also provide a structure to map the values that are implicitly or explicitly present in the innovation environment. Finally, they provide a tool to engineer the innovation environment, for instance by embedding values, or by making deliberation about values a design component of the community. This helps in guiding innovation in environments where early-stage stakeholder involvement is difficult to foster, for instance in industrial settings where information asymmetries can be key to competitiveness. A responsible innovation approach in digitalized natural resource commons can combine risk analysis and the alignment of innovations with transparently deliberated values. Further research in concrete natural resource digitalization projects will be needed to derive practical formulas for responsible innovation in these systems.

Interactions among stakeholders in twin commons are digitally supported, given the often global character of these communities. This provides interesting venues for a Value Sensitive Design approach, in which values are embedded in the systems that support electronic interaction among stakeholders and between stakeholders and the data. Tracking of the usage and of the derived benefits can for instance be done using blockchain technologies, as was proposed for the Amazon Bank of Codes. Here also, further conceptual and empirical research is needed to identify

principles for organizing fair and sustainable commons that provide innovations that are aligned with societally preferred directions and that mitigate the related risks.

5 Responsible Innovation in Synthetic Biology in response to COVID-19: The role of data positionality

5.1. Chapter abstract

Synthetic biology, as an engineering approach to biological systems, has the potential to disruptively innovate the development of vaccines, therapeutics, and diagnostics. Data accessibility and differences in data-usage capabilities are important factors in shaping this innovation landscape. In this paper, the data that underpin synthetic biology responses to the COVID-19 pandemic are analyzed as positional information goods—goods whose value depends on exclusivity. The positionality of biological data impacts the ability to guide innovations toward societally preferred goals. From both an ethical and economic point of view, positionality can lead to suboptimal as well as beneficial situations. When aiming for responsible innovation (i.e. embedding societal deliberation in the innovation process), it is important to consider hurdles and facilitators in data access and use. Central governance and knowledge commons provide routes to mitigate the negative effects of data positionality.

5.2. Introduction

Synthetic biology is a bio-engineering field that pursues the data-driven design of biological systems (Freemont, 2019). It combines molecular biology and lab automation with *in silico* design techniques that are fueled by biological data. *In silico* design refers to the computer-aided design of biological molecules and biological processes, for example, the modeling of proteins or the modeling of pathways that allow for the biochemical synthesis of compounds. Synthetic biology was highlighted in a report from the European Parliament as one of the emerging technologies that can fight the COVID-19 pandemic (Kritikos, 2020). The National Institute of Health in the USA also identified synthetic biology as one way to speed up vaccine development (Begley, 2020). Its potential to revolutionize the development and production of vaccines, therapeutics, and diagnostics underpins this hope. The techniques developed in the synthetic biology community open up radical new possibilities and allow

for a more rapid exploration of such possibilities than with established processes. Synthetic biology labs and firms actively started applying their technologies to contribute solutions for the COVID-19 pandemic. Although probably not part of the first wave of drugs and vaccines, such innovations can shape future responses to this and all future pandemics. For example, DNA- and mRNA-based vaccine technologies can ease the development and production of vaccines. These vaccines consist of synthetic nucleotide strands that trigger the formation of proteins via the individual's own cells, thereby inducing an immune response. The availability of viral sequence data can thus be rapidly translated into vaccine candidates. This allowed ventures such as Moderna and Inovio to move into clinical development in just a few months following the public release of the genetic code of the virus (Thanh Le, et al., 2020). Synthetic biology techniques are also used to construct antigen-carrying nanoparticles. Such nanoparticles have been shown to effectively trigger immune responses in mice and nonhuman primates (Marcandalli, Fiala, Ols, & al., 2019). Nanoparticles can potentially reduce the need for adjuvants and facilitate scalable production. They also show high stability at room temperature, which would ease their distribution in low-income countries (Shin, et al., 2020). This evolution testifies to the disruptive potential of data-driven "plug and play" platforms that aim at the modular design of vaccines against new viruses (CEPI, 2020). Synthetic biology techniques have also been applied in drug discovery and development. For example, cell-free systems were used to design biosynthetic pathways for the antiviral agent valinomycin (Zhuang, et al., 2020). Cell-free systems are free from the complexity and constraints that come with intact cells, containing only the biological components that support the process of interest. Such systems therefore have the potential to further widen the range of engineering possibilities. Synthetic biology techniques have also been applied to develop diagnostic tests for SARS-CoV-2 (Broughton, et al., 2020).

The previous examples are indicative of the potential of synthetic biology techniques to disruptively transform how society can respond to viral outbreaks. Given the devastating impact of the COVID-19 pandemic on people, societies, and economies, rapid responses based on innovations in vaccine development, therapeutics, and diagnostics can be very beneficial. Innovations, however, need to be aligned with societal values to realize this potential. Besides biosafety and biosecurity, innovations need to align with values, such as privacy, access to good healthcare, and a fair distribution of derived benefits. Guiding innovation toward such societally preferred goals is highly relevant in view of the deluge of innovations in synthetic

biology and the strong moral load of data-driven innovations in healthcare (Bruynseels, Santoni de Sio, & van den Hoven, 2018). Responsible research and innovation (RRI) was proposed as a way to align technological innovation with values preferred by society. RRI has been explored in both synthetic biology (Macnaghten, Owen, & Jackson, 2016) and healthcare settings (Silva, Lehoux, Miller, & Denis, 2018) (Douglas & Stemerding, 2013). By including social and ethical aspects in the innovation process, RRI provides a concrete approach for a moral accompaniment of technoscientific developments. RRI has been defined as "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)" (Von Schomberg, 2011) or as "taking care of the future through collective stewardship of science and innovation in the present" (Stilgoe, Owen, & Macnaghten, 2013). How, then, can the collective stewardship of innovations be organized in the case of synthetic biology?

Access to data is pivotal when pursuing synthetic biology innovations and is therefore important when pursuing RRI. As a bio-engineering practice, synthetic biology requires a close intertwinement of *in silico* discovery and modeling and automated lab experiments (Freemont, 2019). Without access to genomic sequence data, high-quality sequence annotations, metabolic models, and so on, it is not possible to achieve much. Capabilities are also required to enable the data to be put to use: computational power, cutting-edge algorithms, and access to know-how (Sachsenmeier, 2016). For COVID-19, excellent public resources are available. Full viral genome sequence data were published in the Global Initiative on Sharing All Influenza Data (GISAID, 2020) and in Genbank's SARS-CoV-2 data hub (GenBank SARS-CoV-2, 2020) starting in early January 2020 (Holmes, 2020) (NHC, 2020). Researchers swiftly used this information, for instance, to synthesize substitutes of the actual viral genome, thereby speeding up global research. Currently, hundreds of variants are available from locations across the globe. The COVID-19 Genomics UK Consortium (COG-UK) aims at sequencing SARS-CoV-2 viruses from up to 230 000 UK COVID-19 patients, with an underpinning commitment to open science and FAIR data principles (COVID-19 Genomics UK (COG-UK) consortium, 2020).

Access to data can be a prerequisite for innovation. On the other hand, data frictions can hamper this ability to innovate. Data frictions (Edwards, 2010) (Bates,

2018) are defined as "socio-material factors that coalesce to slow down and restrict data generation, movement and use." Data frictions have a "politics"; they influence what is known by whom and therefore how future knowledge and social relations are shaped (Bates, 2018). They are an important factor in shaping innovation because they impact which parties are involved. Data frictions relate to the kinds of data and the repositories they reside in, data standards, data-transfer mechanisms and policies, or the lack thereof. They comprise technical and societal hurdles that impair data access, as well as being catalysts that foster data access. During the outbreak of avian influenza A virus (H5N1), a stop placed on data sharing led to significant controversy. Indonesia stopped sharing clinical specimens to international laboratories participating in the World Health Organization Global Influenza Surveillance Network (Sedyaningsih, Isfandari, Soendoro, & Supari, 2008). The rationale behind this was that sharing materials enables international companies to develop vaccines, but that the Indonesian population would not benefit from these developments. The demand for access to drugs and vaccines, for agreements on intellectual property rights, and for capabilities built up via technology transfer and scientific collaborations resulted in an international agreement on a "pandemic influenza preparedness framework." Along the same lines, the MERS coronavirus was isolated in Saudi Arabia, but intellectual property rights on products based on the MERS genomic sequence were owned by a Dutch institute. This situation led to a significant dispute and questions about data sharing (Butler, 2013). Data sharing was also hampered during the Ebola 2013–2015 outbreak due to a variety of hurdles (GRCIDP, 2018). In some cases, viral genomic sequence data were swiftly uploaded via the public platform GenBank, but no standard method existed to disseminate the data. Most of the samples provided for genetic sequencing never resulted in publicly released data (Yozwiak, Schaffner, & Sabeti, 2015). International initiatives recognized the need for improved data sharing, resulting in initiatives such as the establishment of GISAID, a platform for sharing influenza virus sequences and related epidemiological data (Bogner, Capua, Lipman, & Cox, 2006). These cases testify to the fact that it is crucial to organize access to data in such a way that innovation toward societally preferred goals is stimulated while risks are mitigated.

The data-driven innovation response to the COVID-19 pandemic exemplifies this point. Data positionality is put forward as a useful lens through which to analyze innovation dynamics in relation to data. Data positionality refers to situations where the value of data depends on the extent that others do not have access to that data. Positional goods' theory was developed to describe a category of certain marketable

goods whose value depends on externalities; namely, on how they compare with things owned by others (Hirsch, 1977) (Frank, The Demand for Unobservable and Other Nonpositional Goods, 1985) (Pagano, 1999) (Vatiero, 2009) (Zinnbauer, 2018). Positionality implies exclusivity: Scarcity needs to be guaranteed and parties need to be able to benefit from the resulting exclusivity. The question that arises is how data positionality can impact the capability to steer innovation in synthetic biology in societally preferred directions. Positionality is related to the concept of data frictions (Edwards, 2010) (Bates, 2018). Hurdles in accessing biological data determine whether and how data-driven innovation can be steered. Data access thus needs to be considered when aiming at responsible innovation. In this paper, the role of data in COVID-19-related innovations in synthetic biology is used to illustrate data positionality and its repercussions for responsible innovation.

5.3. Synthetic biology data as a positional good

The rush for innovations in the wake of the COVID-19 pandemic is driven by both the pursuit of societal benefits and economic rationales. The current pursuit of drastically shortening vaccine development timelines (Thanh Le, et al., 2020) testifies to this. COVID-19 vaccine development is embedded in a significant economic reality, where high investments are required to bring a vaccine to the market (Gouglas, et al., 2018). CEPI is an organization that invests in vaccine development programs; part of its funding goes to synthetic biology companies (CEPI, 2020).

Given this context, it is insightful to analyze the biological data used in synthetic biology in terms of information goods that have a market value. "Information goods" refers to commodities whose market value is determined by their information content and not by their material properties. Engineering approaches in synthetic biology can depend on a variety of information goods (genomic sequences, sequence annotations, enzyme properties, metabolic models, etc.), often from a variety of species. These goods also depend on lab protocols, algorithms, scientific knowledge, and technical know-how. The information goods that fuel innovations in synthetic biology are therefore very heterogeneous. In terms of the response to COVID-19, viral genomic sequence data and annotations, human genomic sequence data, and clinical and epidemiological data are all crucial inputs for innovations in prevention, diagnosis, and therapy.

What type of goods are information goods? Economic theory has various ways of classifying goods. Commonly, goods are categorized along two axes: according to

their excludability and rivalry. Goods are excludable if parties can be denied access to them. Goods are subtractable (or rivalrous) if consumption by one party reduces the possible consumption by another party. Biological data are non-subtractable goods since consumption of the data by one party will not make the data unavailable to other parties. However, their production requires subtractable goods, such as time, money, and biological systems (e.g. ecosystems or populations). And the biological data themselves can result in subtractable goods (Strandburg, Frischmann, & Madison, 2017), such as new medical treatments or washing powder enzymes. Biological information goods have often been categorized as public goods, which are non-excludable and non-subtractable. The rationale for this categorization is that they are a form of scientific knowledge, which is the archetypical example of a global public good. Scientific theories, such as Einstein's relativity theory, are available to all, and usage of the theories does not diminish their value for others. This typecasting as a global public good can be used as a strategy to instill ethically preferred dynamics by stimulating the sharing of data across national and international boundaries (Chadwick & Wilson, 2004). Moreover, in principle, well-oiled online markets that allow for efficient price-setting should result in information goods that cost virtually nothing. Digital artifacts can be copied at high speed and low cost as soon as the first artifact is made (Quah, 2002). The synthetic biology community has tended to promote an ethos of open innovation (Torrance, 2017), which is, at first sight, in conflict with the positional character of synthetic biology data. Along those lines, open-source software development is often used as inspiration when shaping the field (Urquiza-Garcia, Zielinski, & Millar, 2019). The open-source software movement proved to be a very viable complement to proprietary software schemes, and it vastly stimulated innovation (Boyle, 2008). Translated to the field of synthetic biology, this finds its analogy in schemes that allow for building freely on the genetically encoded functions shared by the community, such as through the BioBricks Public AgreementTM (BioBricks Public Agreement(TM), 2020). Similarly, the open-science movement provides a model where scientific findings and related datasets are made publicly available without the hurdle of subscription costs (Levin, Leonelli, Weckowska, Castle, & Dupré, 2016) (Burgelman, et al., 2019). The synthetic biology community's response to the corona pandemic showed that publicly available information can vastly speed up innovation.

In practice, the categories of biological information goods are much more colorful. Rather than being a public good, many datasets are not publicly available but reside in proprietary databases, experience delays or incompleteness in data

release, or are only accessible given the right membership. Intellectual property regimes further complicate the picture by providing temporal monopolies over the concrete applications of patented knowledge. The ambiguity of synthetic biology in this respect is indicative. Synthetic biology is often defined as an engineering discipline, next to it being a scientific discipline. This hints at the fact that synthetic biology does not only result in scientific theories that are asymptotically available to all. In many cases, the outputs concern designs and engineered systems that provide a competitive advantage in a market, and thus are inherently related to information asymmetries. In economic terms, these observations are indicative of a market failure that leads to a tendency for some information to become exclusive rather than being free and open (Zinnbauer, 2018). The value of some information goods depends on whether others do not own them; thus, on their exclusivity. Having priority access to biological data puts one in the position to mine the data first and produce derived goods, such as scientific papers, new pharmaceuticals, medical treatments, etc. Such information goods are more valuable if others do not have equal access to them or are less capable of putting these data to use. Digital information goods therefore do not, by definition, result in open data or in a market-clearing price that is close to zero. They can experience scarcities that are either artificially constructed or are the result of socio-technical constraints in data movement.

5.4. Drives behind positional effects in biological information goods

Positionality has recently been described as an overlooked property of information goods in general, which can explain certain failures in the data market (Zinnbauer, 2018). Instead of everyone enjoying a world of free and open data, many data-holders benefit from constructing an artificial scarcity in information so that a much higher premium can be gained. And there are buyers who are willing to pay these high premiums as long as the scarcity remains guaranteed, and they can benefit from exclusivity. This holds true for premium political, business, and legal information, and also for certain forms of scientific information (Zinnbauer, 2018). Many biological data and much derived knowledge reside in databases with tightly controlled access—often proprietary—and they are sometimes subject to intellectual property rights or expensive subscriptions. Thus, the introduction of a "manufactured scarcity" (Zinnbauer, 2018) counteracts the fact that the data themselves are, in principle, infinitely sharable. In economic terms, such effects are called "positional." Positional goods (Hirsch 1976) have a value that is determined in relative terms by

their externalities. Their value does not merely depend on the quantity of the goods, but on their exclusivity—on the extent to which others have no access to them.

The positionality of information goods comes in various forms. At an abstract level, one can distinguish between horizontal and vertical positionality, depending on the types of externalities that impact the value of the information good. Vertical positionality refers to goods whose value is inversely related to the degree to which others have access to them. Horizontal positionality refers to goods whose value depends on the accessibility of other goods (van den Hoven, Helbing, & Domingo-Ferrer, FuturICT - The road towards ethical ICT, 2013). Different forms of positionality can be distinguished in the case of information goods (Zinnbauer, 2018). These can either be the result of strategies to create artificial scarcities to obtain a positional advantage or the result of practical constraints that hamper the fluent distribution of information goods. Multiple forces that shape data frictions can be distinguished (Bates, 2018): (1) data-sharing infrastructures, (2) socio-cultural factors, and (3) regulatory factors. Such frictions are claimed to have a "politics" because they shape the interactions between parties involved in data handling and exchanges (Bates, 2018). Data-sharing infrastructures can introduce friction because of the complexity of the data representations needed to capture biological data, the lack of generally accepted data standards and ontologies, the anonymization and encryption methods required to guarantee genomic privacy, a variety of technical constraints such as bandwidth or computational power when dealing with big amounts of sequence data, and so on. A lack of data standards and data interoperability, for instance, was put forward as a challenge to open science in the Organisation for Economic Co-operation and Development's (OECD's) policy response to COVID-19 (OECD, 2020). Data friction can also arise because of a lack of time or skills to cleanse, prepare, and submit the data, or a lack of time for scientists to document their experiments and annotate the data. Socio-cultural factors can legitimize data frictions, for instance, by guaranteeing the data privacy of research subjects or patients or avoiding a misinterpretation of the data. Socio-cultural data frictions also arise in highly competitive environments where data are retained in an explicit attempt to retain a competitive advantage or are shielded from scrutinization by other researchers (Bates, 2018). Synthetic biology, in this regard, combines an open-source ethos and an intertwinement with commercialization activities. Often, the core members of the synthetic biology community are systematically in close proximity to commercial activities (Raimbault, Cointet, & Joly, 2016). A culture of data sharing is deeply interwoven with this scientific field, while, on the other hand,

information asymmetries are implied by the competitive publication and innovation landscape.

Data positionality provides a lens through which to interpret the effects of data frictions in both highly competitive and highly collaborative research and development triggered by the COVID-19 pandemic. Various forms of data positionality can be distinguished. Temporal positionality refers to data goods in which the time component drives the differences in data accessibility. For instance, being able to run a speed-trading algorithm on servers next to the stock market can provide a few milliseconds of earlier access that an algorithm needs to outperform competitors. An analogous situation holds true for biological data. For instance, access to the SARS-CoV-2 genomic sequence and to epidemiological information proved to be crucial in effective policy and technological responses to the pandemic, for instance, in terms of the ability to rapidly develop diagnostic tests (Peeri, et al., 2020). Hence data-release policies are important in shaping open environments where optimal use is made of research data. Building on the experiences from previous viral outbreaks, data-sharing platforms such as GISAID (GISAID, 2020) and data-sharing guidelines (RDA COVID-19 Working Group, 2020) were put in place. Geographical positionality refers to the competitive advantage that results from proximity to the location where the data are generated. It is easier to derive value from a dataset if one has direct insight into how the data were generated and processed, and if one has personal connections with the researchers who were involved in the process and one can tap into their tacit knowledge. The tight link between the data and the human population from which they were derived also ties data to a specific region. Biobanks, for instance, constitute a key resource in the fight against pandemics (Vaught, 2020) and these repositories have physical locations. Population genomics and electronic health records are often bound to local populations and national initiatives. National borders (as a proxy for national regulations and political assessments) can result in data frictions and the related positional effects. For instance, for human genomic material and derived information, genomic sovereignty was proposed to ensure "a nation's ability to capture the value of its investments in the field of genomic medicine" (Hardy, Séguin, Ramesar, Singer, & Daar, 2008). Getting access to data can require personal connections to scientists in heavily affected regions, for instance, when pursuing association studies (Olena, 2020). Association studies—relating the genetic profile of individuals to their disease outcome—will be a key tool in answering the question of why SARS-CoV-2 hits patients with varying severity. Data-sharing initiatives, such as the COVID-19 Host Genetics Initiative,

can help in reducing geographical positional effects in this case (The COVID-19 Host Genetics Initiative, 2020). It is important to note that technical abilities and know-how are as crucial as the datasets that fuel the innovation process—and these assets can be equally positional. These innovation capabilities are geographically unequally distributed. For instance, investments in synthetic biology ventures in general in the second quarter of 2019 amounted to 1.2 billion USD, only 12% of which occurred outside of the USA (SynBioBeta, 2019).

Owning positionality refers to situations in which ownership of the information good results in positionality. Next to keeping data in databases with restricted access, layered access and delays in data release are inherent to the biomedical field. This creates sub-domains that span a range from closed and proprietary, to knowledge commons that are managed by a community, up to databases that are geared toward open access. The transparency of proprietary data services (and of the quality of data in general) came under close scrutiny in the context of COVID-19 research, with the retraction of two high-profile papers (Piller & Travis, 2020). One of the publications impacted trials with the drug hydroxychloroquine and the other led to increased demand for the drug ivermectin. In the case of monopolized positionality, single parties own the key data assets in a certain market. Such a situation also relates to new frontier positionality, referring to situations where parties gain a competitive advantage by entering into novel data fields. Synthetic biology is a field par excellence where dedicated technologies are developed and related information is gathered around cutting-edge fields of research. For instance, the extremely rapid development of mRNA- and DNA-based vaccine candidates for SARS-CoV-2 hinged on disruptive technologies that had already been explored in the context of other diseases.

Horizontal positionality occurs when the value of an information good depends on access to other information goods (van den Hoven, Helbing, & Domingo-Ferrer, FuturICT - The road towards ethical ICT, 2013). This is very often the case for biological data since the combined analysis of multiple biological datasets is often needed in order to derive value from them. Mining a population's genomes in combination with medical records can provide powerful insights into disease trajectories and tailored treatments for specific sectors of the population (Boeck Jensen, et al., 2014). For instance, the UK Biobank is going to add COVID-19 health-related data to its records, providing an integrated dataset for researchers to study the relationship between a person's genetic makeup and disease susceptibility (UK Biobank, 2020). Heterogeneous datasets often end up in separate, specialized data

repositories with their own specific data-release schemes, access policies, and technical accessibility, which results in frictions when connecting the data. Big-tech positionality (Zinnbauer, 2018) refers to the positional advantage that big companies and institutes can have when integrating data because of their access to significant amounts of proprietary data and to their data-analysis capabilities.

5.5. Organizing responsible innovation in view of data positionality

To organize responsible innovation, one needs to be "response-able"—to be able to respond to the novel opportunities and risks that emerge. Previous viral outbreaks with pandemic potential proved that the level of data sharing significantly impacts this ability to respond. Responsiveness was highlighted as one of the dimensions of RRI (Stilgoe, Owen, & Macnaghten, 2013). It requires the ability to swiftly steer an innovation process if deemed appropriate. The COVID-19 pandemic demanded a quick innovation response to deliver therapeutics, vaccines, and diagnostics. This speed of response must go hand in hand in hand with mechanisms to ensure ethical correctness and societal desirability. Aspects such as patient safety and data privacy, as well as dialogues around the desirability of novel therapeutic and preventative technologies, need to be interwoven with the entire innovation process. On the positive side, data positionality can foster competition in a Schumpeterian scheme, and thereby it can become instrumental in pursuing innovations that match societal preferences. However, leaving everything up to a market dynamic can also result in unequal data distribution and an unequal capability in terms of innovating and steering innovations. The COVID-19 pandemic led to a strong international push to mitigate temporal and geographical data positionality by strengthening rapid data-release and data-sharing mechanisms across national and institutional boundaries. For instance, the initial sharing of viral sequences via existing data-sharing mechanisms, such as GISAID (GISAID, 2020) and GenBank (GenBank SARS-CoV-2, 2020), provided the necessary information for academic labs and companies to synthesize parts of the viral hereditary material, thereby vastly speeding up innovation processes globally. Synthetic biology approaches to COVID-19 aim at dramatically shortening the development of vaccines, therapeutics, and diagnostics. The hope is that the sheer diversity of innovations (Thanh Le, et al., 2020) will provide room for responsiveness in terms of shaping the overall solution space.

Public debates on new technologies also often revolve around anticipation, reflexivity, and inclusion. Next to responsiveness, these perspectives together reflect

societal concern and interest in technological innovation and can be used as dimensions in a responsible innovation approach (Owen, Macnaghten, & Stilgoe, Responsible research andinnovation: From science in societyto science for society, with society, 2012) (Stilgoe, Owen, & Macnaghten, 2013). Along these lines, the anticipate, reflect, engage, act (AREA) framework from the 2014 Rome Declaration (Italian Presidency of the Council of the European , 2014) was recently proposed as a step toward RRI in COVID-19-related data research (Leslie, 2020) (Braun, Blok, Loeber, & Wunderle, 2020). These dimensions have been considered in the context of synthetic biology (Macnaghten, Owen, & Jackson, 2016) and healthcare (Silva, Lehoux, Miller, & Denis, 2018). Next to the aforementioned dimensions, value domains specific to RRI in healthcare were proposed. These include, for instance, health equity, the level of care, frugality (if more can be achieved with fewer means), and the values that are embedded in the business model of the innovators (Silva, Lehoux, Miller, & Denis, 2018). Anticipation is about considering possible outcomes of new technologies. Reflexivity means taking a step back and considering the innovation activities from a broader perspective. Data positionality should be included in assessments about anticipation and reflexivity, given the potential contribution to unintended consequences, and the effect on the desirability of the possible futures that the innovations will contribute to. Synthetic biology solutions, such as universal vaccines, for instance, have a "plug and play" character, and their effectiveness is thus tied to data availability. Geographical and temporal positionality will therefore be at play when such solutions will require tailoring to new variants of viruses. Owning positionality and big-tech positionality will relate to the question of how the landscape of providers and beneficiaries should be organized. Inclusion is a central theme in RRI approaches (Burget, Bardone, & Pedaste, 2017) (Bogner & Torgersen, 2018). Many stakeholders do not often have a say in the development of new technologies, although they need to bear the consequences later on. Healthcare innovations, for instance, have an impact on many of us; nevertheless, these innovations often take place in the confined labs of academic and corporate research institutes. For this reason, broad stakeholder involvement in the early innovation steps is often targeted in responsible innovation approaches (Owen, Bessant, & Heintz, 2013). This is in contrast to risk-assessment methods where the technical experts are the main driving forces and where assessments of the novel technologies are mainly done when approaching the market. The various types of data positionality can negatively impact this ability to include stakeholders. Competitive advantage is central to the very notion of positionality. Positionality therefore implies a topology

in which the innovation step is shielded from external parties. Traditional models of drug innovation took place in the well-shielded environment of pharmaceutical companies, building on public data, but also deriving competitive advantage from proprietary data. These models are increasingly opened up in public–private partnerships that allow for deeper involvement of the public and the many stakeholders (biobanks, researchers, public healthcare funders, etc.), and in open innovation models. Such openness in the early innovation step implies a reduction in data friction.

The high sense of urgency related to the COVID-19 pandemic highlighted the importance of data access, data quality, and capabilities to put data to use. The high speed at which data-driven innovation occurs stresses existing processes. Central governance provides one route to shape data-driven innovation by modulating the effects of data positionality. Governance instills reciprocity among stakeholders via rules that constrain the options each individual rational party can select. Rules and regulations partially constrain the freedom of the individual players; nevertheless, they can be to their overall benefit by avoiding resource-wasting situations (Frank, 2012). Data-privacy rules, for instance, constrain the space of innovations, thereby avoiding innovations that do not adhere to the imposed privacy values. Data-release policies, such as the open-science movement (Levin, Leonelli, Weckowska, Castle, & Dupré, 2016) (Burgelman, et al., 2019), can result in a broadening of the innovation space. Intellectual property arrangements provide an institutionalized way to create positional assets. Patenting of naturally occurring genetic sequences is no longer allowed by the United States Patent and Trademark Office, but it is still allowed in Europe (Cole, 2015). Whether patenting genomic sequences has a positive or negative impact on healthcare innovation remains under debate (Liddicoat, et al., 2019). Central governance may also be needed to contain data positionality to the sphere of research and innovation. For instance, the race for prime access to personal biological data should not negatively impact a person's level of healthcare, career opportunities, or family. Avoiding "informational injustice" (Manders-Huits & van den Hoven, 2008) requires rules and regulations that install data frictions at the boundaries of societal spheres.

In light of the ongoing pandemic, innovators were confronted with both aggressive timelines and high technological uncertainty. This situation triggered a vast increase in collaborations across highly diverse parties, leading to "a culture of collaboration across government, industry and academia" (Ledford, 2020). This situation hints at future routes for responsible innovation. Fostering self-regulation

and self-governance in innovation communities can be a venue for guiding innovation toward societally preferred goals. Research and innovation communities centered around synthetic biology can provide an entry point. A growing body of literature analyzes biological data as a common-pool resource—more precisely, as information goods in a knowledge common (Strandburg, Frischmann, & Madison, 2017). This common's perspective is helpful in clarifying the complex rules and the self-regulating properties of biological data-driven research and innovation communities. It also provides a framework to develop communities that mitigate negative positional effects and foster positive effects. The exclusion of certain stakeholders in an information common has been brought forward as an important element in the shaping of power asymmetries (Prainsack, 2019). In pharmaceutical commons, positionality can be a driving force for gathering stakeholders around an innovation topic. Intellectual property and privately owned assets, when not used to fence off competitors, can function as a magnet to attract parties into collaboration (Lezaun & Montgomery, 2015). For an individual party, it constitutes a "ticket of admission" to pharmaceutical product-development partnerships. Carefully managing access to data can also be a strategy in mitigating cybersecurity risks related to pathogen databases (Vinatzer, et al., 2019) and confining the data to an innovation community adhering to strong research ethics. In these settings, certain types of positionality can be transformed into a creative source rather than a wasteful situation.

5.6. Chapter conclusions

The pace of synthetic biology innovations in response to the COVID-19 pandemic is unprecedented. This dynamic is driven by significant progress in synthetic biology as well as by improvements in data sharing. The potential societal impact of synthetic biology innovations calls for ways in which to foster beneficial outcomes that resonate with societal values while avoiding potential negative effects. Responsible innovation has been proposed as a framework to achieve this goal. When applied to synthetic biology innovations related to the COVID-19 vaccine, therapeutic, and diagnostic developments, it is clear that the role of data is pivotal. In this paper, biological data used in synthetic biology are analyzed as positional information goods. Positionality refers to the observation that the value of some pieces of information is related to their exclusivity. Various flavors of positionality can be identified relating to different types of data-access hurdles. Data positionality is Janus-faced—it can hamper responsible innovation but can also be a stimulating force. Measures to

shape the data topology in terms of positionality are therefore an important instrument in steering synthetic biology innovations toward societally preferred goals. Central governance and self-governance in commons-like settings provide venues to mitigate the negative effects of data positionality.

6 Outlook for further research

Innovations that are inspired by biology hit the market at an increased rate, in medicine, synthetic biology, industrial biotechnology, agriculture, and many other societal sectors. Biological data increasingly fuel these innovations, by providing the inspiration for designs of molecules, materials, production mechanisms, and so on. These high levels of innovativeness will be needed to meet the grand challenges of today's world.

The avalanche of innovations though also sharply raises fundamental questions about the Baconian project. Bacon's utopian vision brought a lot of benefits, but also proved to be disastrous in its exploitation of ecosystems. The key question is whether the utopian vision can be re-thought in a direction that re-establishes a healthy and sustainable connection with our natural environment and with ourselves, while retaining the original inspiration that innovations can improve life. When embarking on such a rethinking, it will be crucial to construct ways to align innovations with societal needs and values. This is not a trivial task. Innovations shape our world, for the better and the worse. But steering the innovation tsunami into beneficial directions is a tantalizing endeavor, for many reasons. To name a few: innovation is often strongly driven by a Schumpeterian dynamic, oriented toward profitability in the market. The innovation process is often shielded away from society, because of the competitive advantage it can provide. Innovation happens in very heterogeneous contexts, including settings as diverse as hacker spaces, academic groups and corporate labs. Moreover, the future is radically opaque for our conceptual instruments. And especially when it also concerns disruptive innovations, predictions on which innovations do contribute to a desired future are by definition difficult to make.

The question of how to structurally include ethical and societal perspectives in the innovation process in order to bring about the desired futures is therefore highly relevant, but equally hard to answer. The research field of Responsible Innovation started to take up this challenge. Different to previous frameworks, RI's ambitions are to bring in the ethical and societal aspects in the entire innovation process. A main driver for the development of RI was the observation that significant technological innovations sometimes are rejected by society when entering the market. By including broader segments of society in the innovation process, the hope was to develop a shared responsibility for the innovation, thereby ensuring societal acceptance. Next to this focus on societal acceptance of innovations, the main focus of the

RI and Responsible Research and Innovation (RRI) has been the governance of publicly funded research, mainly in the context of the European Union's framework programs.

The scope of RI and RRI therefore is currently rather limited, especially in view of the world's challenges. This calls for a broadening of both the ambitions and the concepts in RI. One major transition is to shift the focus from matching individual innovations with societal values, toward matching entire bioeconomy segments with societal values. Grand challenges require concerted action, spanning entire domains of technological innovations. The ability to act therefore needs to go beyond the assessment of individual technological innovations. For instance, when considering responsible innovation in agriculture, is has been argued that priority should be given to a well-informed debate about which type of agriculture our society wants to foster, rather than to debates about individual technologies (Bogner and Torgersen 2018). A similar argument can be mode for health care (Lehoux, Silva, Sabio, & Roncarolo, 2018) and for synthetic biology (Macnaghten, Owen, & Jackson, 2016). In order to meet today's challenges, it is necessary to shape the bigger pictures, and to foster innovations that support them. An example in this direction is provided in a recent report of the European Union. In this foresight report, an inventory of hundred radical innovations was made. These radical innovations were linked to Global Value Networks, defined as "networks of actors connected by relationships that create value" (Warnke, et al., 2019). Global Value Networks are driven by a global value promise – a promise to meet a global demand (for instance good health care) or a collective need. The authors of the report expect that value creation in the future will be shaped by global value promises, in particular by the UN Sustainable Development Goals.

Bio-based innovation plays an increasingly important role in meeting such global value promises. Especially the connection of information technologies (in particular AI) with health and environmental technologies was seen as a fruitful area for speeding up positive technological transformation, while controlling negative side effects (Warnke, et al., 2019). The development of bio-based economies is seen as a route towards an innovation landscape that allows for economic progress and societal benefits, while avoiding negative effects on the environment and replacing current resource-exhausting practices. For example, synthetic biology is anticipated to contribute to the development of a sustainable bio-based economy (Bueso & Tangney, 2017) (Loeffler, et al., 2018) (French, 2019), defined as "an economic model where knowledge-based utilization of biological resources and processes can

be applied to the sustainable production and manufacturing of goods, and the provision of services across all economic sectors" (Freemont, 2019). These bio-based economies heavily rely on bio-inspired design, and therefore on biodata and the related reality of biological populations and ecosystems.

Research will be required to determine how biodata-based innovation can be made transformative for entire economies, and how such transformations can be fostered and steered according to certain values. The tight intertwinement of biodata with the innovation process and with values makes biodata a key element to include in such analysis. This dissertation started to explore this space, and showed the potential of biodata as an entry point for bio-based responsible innovation. The Genomics England project (Marx, 2015) and the Amazon Bank of Codes (Nobre, et al., 2016) provide example of such a direction in the field of health care and of bio-economy, respectively. Both initiatives, in very different ways, foster the development of a responsible community of innovators centered around a body of biodata. These biodata in both cases are linked to a concrete biological reality: a human population in the case of Genomics England, and the ecosystem of the Amazon basin in the case of the Amazon Bank of Codes. Further conceptual and empirical research is needed to identify principles for organizing R&D environments that foster responsible innovation. As highlighted by Ostrom, communities that are successful in managing a certain resource tend to share certain characteristics (Ostrom, 1990). Along these lines, research communities that manage a biodata-resource also have characteristics of knowledge commons (Pálsson and Prainsack 2011) (Uhlir, 2011) (Strandburg, Frischmann, & Madison, 2017). Exploration of successful innovation commons can be a fruitful way to identify principles for designing settings for responsible innovation. Such an analysis can vastly benefit from encompassing research settings beyond academic research, and include corporate research and development and public-private partnerships. Inclusion of these is important given the sheer impact of their innovations on the market, and thereby on society.

Some of these elements that are important in fostering biodata-based responsible innovations were explored in this dissertation. In order to transform them into design principles for RI-fostering commons, further research is required. For example, information asymmetries were highlighted as being impactful when pursuing responsible innovations. Information asymmetries are inherent to innovation, but when unfettered they give rise to an innovation dynamics that cannot be steered by society. Such asymptotic technological determinism can be counteracted by explicitly installing and removing information asymmetries with the focus on the fostering

of responsible innovations. Along these lines, research is needed on how flows of biological materials, flows of biodata, and flows of derived value interrelate and can be managed in such a way that they foster a fair and sustainable innovation environment, and innovations that are aligned with societal values. Another design principle for bio-based responsible innovation might revolve around ways to include societal values in the innovation process. One might consider embedding societally deliberated values in the socio-technical environments in which the innovations take place, for instance in the form of innovation commons. Such strategies would avoid applying responsible innovation as if it were a regular innovation process that includes more stakeholders. Instead of deliberating about values that need to be taken into account in a particular innovation, the focus would shift to providing an environment with optimal conditions for fostering responsible innovations. This also implies that RI includes social innovations.

The way biodata are put to use in innovations also transforms the way we think about - and the way we treat - our natural environment, our fellow human beings and ourselves. Further philosophical research is needed on the tight interplay between natural processes and culture that is at play in bio-inspired technologies. The tension between these heterologous domains (operational processes versus symbolic representation) is made explicit in biodata and computational models, since they constitute interfaces between our world of language, meaning and aspirations, and the world of natural evolution. Application of biodata, computational models, and asymptotically also digital twins, requires taking a position on how we relate to our natural environment. A re-thinking of Bacon's utopian vision therefore can benefit from including a focus on biodata and computational models. Philosophical research is required to analyze how biodata-based innovations impact how human beings relate to the natural world. When considered solely as instruments towards and increased efficiency, they further underpin a purely anthropocentric Baconian vision in which human engineering actions rationally mold the natural world towards human needs. Such vision can lead to a far reaching instrumentalization of the natural world. Biodata-based innovations though also open perspectives on an inclusiveness that goes beyond the purely anthropocentric, opening up a 'more-than-human world'. Biological data have the potential to make heard what otherwise remains hidden. They thereby provide venues to broaden the group of stakeholders that is involved in a responsible innovation process, and also include what is of interests to the natural world. After all, shaping the world via innovations is a joint enterprise.

References

Allhoff, F., Lin, P., & Steinberg, J. (2009). Ethics of Human Enhancement: An Executive. *Science and Engineering Ethics*, 16 (2): 1-12 .

Ashley, E. B., Wheeler, M., Chen, R., Klein, T., Dewey, F. D., Ormond, K., . . . Altman, R. (2010). Clinical assessment incorporating a personal genome. *The Lancet, 375*(9725), 1525 – 1535.

Ayday, E., De Cristofaro, E., Hubaux, J.-P., & Tsudik, G. (2013). The chills and thrills of whole genome sequencing. IEEE Computer Society Digital Library.

Bacon, F. (1626). *New Atlantis.*

Baldi, P., Baronio, R., De Cristofaro, E., Gasti, P., & Tsudik, G. (2011). Countering GATTACA: Efficient and secure testing of fully-sequenced human genomes. *CCS '11 Proceedings of the 18th ACM conference on computer and communications security.*

Barnes, B., & Dupré, J. (2008). Genomes and what to make of them. Chicago and London: Univ. Chicago Press.

Bates, J. (2018). The politics of data friction. *Journal of Documentation, 74*(2), 412-429.

Begley, S. (2020). To develop a coronavirus vaccine, synthetic biologists try to outdo nature. *Stat news*.

Bergen, J. (2017). Responsible Innovation in light of Levinas: rethinking the relation between responsibility and innovation. *J. Resp. Innov., 4*(3), 354-370.

BioBricks Public Agreement(TM). (2020). *BioBricks Public Agreement(TM)*. Retrieved 2020, from https://biobricks.org/bpa/users/agreement/

Blok, V., & Lemmens, P. (2015). The emerging concept of responsible innovation. Three reasons why it is questionable and calls for a radical transformation of the concept of innovation. In B.-J. O. Koops, *Responsible innovation 2: Concepts, approaches and applications* (pp. 19-35). Berlin: Springer international.

Boeck Jensen, A., Moseley, P., Oprea, T., Ellesøe, S., Eriksson, R., Schmock, H., . . . Brunak, S. (2014). Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nature Communications*, 5: 4022.

Bogner, A., & Torgersen, H. (2018). Precaution, Responsible Innovation and Beyond – In Search of a Sustainable Agricultural Biotechnology Policy. *Frontiers in Plant Science*.

Bogner, P., Capua, I., Lipman, D., & Cox, N. e. (2006). A global initiative on sharing avian flu data. *Nature, 442*, 981.

Bostrom, N. (2008). Ethical issues in human enhancement. In J. Ryberg, T. Petersen, & C. Wolf, *New waves in applied ethics* (pp. 120-152). Pelgrave Macmillan.

Boyle, J. (2008). *The public domain: enclosing the commons of the mind.* New Haven & London: Yale University Press.

Braun, R., Blok, V., Loeber, A., & Wunderle, U. (2020). COVID-19 and the onlineification of research: kick- starting a dialogue on responsible online research and innovation (RoRI). *J. Resp. Innov.*

Brenner, M. (2012). Personalized Medicine: Words that mean just what you choose? *Mo. Ther.*, 20(2): 241-242.

Broughton, J., Deng, X., Yu, G., Fasching, C., Servellita, V., Singh, J., . . . Chiu, C. (2020). CRISPR–Cas12-based detection of SARS-CoV-2. *Nature Biotechnology*.

Bruynseels, K., Santoni de Sio, F., & van den Hoven, J. (2018). Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm. *Front. Genet.*

Bueso, Y., & Tangney, M. (2017). Synthetic biology in the driving seat of the bioeconomy. *Trends in Biotechnology, 35*(5), 373-378.

Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., & Schouppe, M. (2019). Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. *Frontiers in Big Data*.

Burget, M., Bardone, E., & Pedaste, M. (2017). Definitions and Conceptual Dimensions of Responsible Research and Innovation: A Literature Review . *Sci Eng Ethics, 23*, 1–19 .

Burms, A., & Vergauwen, R. (1991). Natural kinds and moral distinctions. *Philosophia, 21*(1-2), 101-105.

Butler, D. (2013). Tensions linger over discovery of coronavirus. *Nature*.

CEPI. (2020). *Coalition for Epidemic Preparedness Innovation*. Retrieved 2020, from https://cepi.net

Chadwick, R., & Wilson, R. (2004). Genomic databases as global public goods. *Res Publica*, 10: 123-134.

Choudhury, S., Fishman, J., McGowan, M., & Juengst, E. (2014). Big data, open science and the brain: lessons learned from genomics. *Frontiers in Human Neuroscience, 8*(239).

Cole, P. (2015). Patentability of genes: a European Union perspective. *Cold Spring Harb. Perspect. Med., 5*, a020891.

Collingridge, D. (1980). The social control of technology.

Contreras, J. L. (2014). Constructing the Genome Commons. In B. M. Frishcmann, *Governing Knowledge Commons.* Available at SSRN: http://ssrn.com/abstract=2474405.

Convention on Biological Diversity. (2010). *Nagoya Protocol.* Retrieved 2019, from https://www.cbd.int/abs/about/

COVID-19 Genomics UK (COG-UK) consortium. (2020). An integrated national scale SARS-CoV-2 genomic surveillance network. *The Lancet Microbe*.

Cyranoski, D. (2019). Russian 'CRISPR-baby' scientist has started editing genes in human eggs with goal of altering deaf gene. *Nature*.

Daniels, N. (2000). Normal functioning and the treatment-enhancement distinction. *Cambridge Quarterly of Healthcare Ethics, 9(3)*, 309-322.

Dedeurwaerdere, T., Melindi-Ghidi, P., & Broggiato, A. (2016). Global scientific research commons under the Nagoya Protocol: Towards a collaborative economy model for the sharing of basic research assets. *Environ. Sci. Policy, 55*, 1-10.

Douglas, C., & Stemerding, D. (2013). Governing synthetic biology for global health through responsible research and innovation. *Sys. Synth. Biol., 7*, 139-150.

Dupré, J. (1993). In J. Dupré, *The Disorder of Things.* Cambridge, MA: Harvard University Press.

Edwards, P. (2010). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming.* Cambridge, MA.: MIT Press.

El Emam, K. (2011). Methods for the de-identification of electronic health records for genomic research. *Genome Medicine, 3*, 25.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*, 57–74.

Endy, D. (2005). Foundations for engineering biology. *Nature Reviews, 438*(24), 449-453.

Erlich, Y., & Narayanan, A. (2014). Routes for breaching and protecting genetic privacy. *Nat. Rev. Genet., 15*(6), 409-421.

European Political Strategy Center. (2016). *Towards an innovation principle endorsed by better regulation.* European Commissino.

Evans, B. (2014). Mining the human genome after association for molecular pathology v Myriad Genetics . *Genetics in Medicine*, 504-508.

Evans, B. (2017). Genomic Data Commons. In K. Strandburg, B. Frischmann, & M. Madison, *Governing Medical Knowledge Commons* (pp. 74-101). Cambridge University Press.

Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2083).

Framingham Heart Study. (2017). Retrieved from www.framinghamheartstudy.org

Frank, R. (1985). The Demand for Unobservable and Other Nonpositional Goods. *The American Economic Review, 75*(1), 101-116.

Frank, R. (2012). The Darwin Economy: Liberty, Competition, and the Common Good. Princeton University.

Freemont, P. (2019). Synthetic biology industry: data-driven design is creating new opportunities in biotechnology. *Emerging Topics in Life Sciences, 3*, 651-657.

French, K. (2019). Harnessing synthetic biology to meet the 2030 UN Sustainable Development Goals. *Nature Sustainability*, 250-252.

Fukuyama, F. (2002). *Our posthuman future.* Douglas & McIntyre, Ltd.

GE. (2017). *Predix technology brief - Digital twin.* GE.

GenBank SARS-CoV-2. (2020). *GenBank SARS-CoV-2*. Retrieved 2020, from https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/

GIANT Consortium. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature, 467*, 832–838.

GISAID. (2020). *GISAID*. Retrieved 2020, from https://www.gisaid.org/

Giubilini, A., & Sagar, S. (2015). The ethics of human enhancement. *Philosophy Compass*, 10(4): 233-243.

Glaessgen, E. H., & Stargel, D. (2012). The Digital Twin Paradigm for F uture NASA and U.S. Air Force Vehicles. *53rd Structures, Structural Dynamics, and Materials Conference: Special Session on the Digital Twin.* American Institute of Aeronautics and Astronautics.

Gottweis, H., Kaye, J., Bignami, F., Rial-Sebbag, E., Lattanzi, R., & Macek, M. J. (2012). *Biobanks for Europe, a challenge for governance.* European Commission.

Gouglas, D., Thanh Le, T. H., Kaloudis, A., Danielsen, T., Hammersland, N., Robinson, J., . . . Røttingen, J.-A. (2018). Estimating the cost of vaccine development against epidemic infectious diseases: a cost minimisation study. *Lancet Glob. Health, 6*, e1386–1396.

GRCIDP. (2018). *Data Sharing during the West Africa Ebola Public Health Emergency: Case Study Report.* https://www.glopid-r.org: Global Research Collaboration for Infectious Disease Preparedness. Retrieved from https://www.glopid-r.org

Green, M., & Botkin, J. (2003). Genetic exceptionalism' in medicine: clarifying the differences between genetic and nongenetic tests. *Ann. Intern. Med., 138*(7), 571-575.

Green, R., & Thomas, A. (1998). DNA: five distinguishing features for policy analysis. *Harvard Journal of Law & Technology, 11*(3), 571-591.

Gymrek, M., McGuire, A., Golan, D., Halperin, E., & Erlich, Y. (2013). Identifying personal genomes by surname inference. *Science, 339*(6117), 321-324.

Hardy, B.-J., Séguin, B., Ramesar, R., Singer, P., & Daar, A. (2008). South Africa: from species cradle to genomic applications. *Nature Reviews Genetics* , 9, S19–S23.

Hess, C., & Ostrom, E. (2006). *A framework for analyzing the microbiological commons, pp 335-349.* ISSJ 188, UNESCO, Blackwell Publishing Ltd.

Hess, C., & Ostrom, E. (2007). Introduction: An overview of the knowledge commons. In *Understanding knowledge as a commons* (pp. 3-26). Cambridge, MA: MIT Press.

HIPAA Privacy Rule, .. C. (2007). Retrieved from U.S.Government Publishing Office: http://www.gpo.gov

Hirsch, F. (1977). *The Social Limits to Growth.* London: Routledge.

Hofmann, B. (2017). Limits to human enhancement: nature, disease, therapy or betterment? *BMC Medical Ethics*, 18 (56).

Holmes, E. (2020). *Initial genome release of novel coronavirus 2020 [14 January 2020].* Retrieved from http://virological.org/t/initial-genome-release-of-novel-coronavirus/319

Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., . . . Craig, D. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genetics, 4*(8), e1000167.

Hottois, G. (1988). Evaluer la technique – Aspects éthiques de la philosophie de la technique,. In G. Hottois, *Liberté, humanisme, évolution.* Paris: Librairie Philosophique J. Vrin,.

Hottois, G. (1995). *Philosophie et technosciences.* Kampen: Kok Agora.

Hottois, G. (1996). Entre symboles et technosciences. Un itinéraire philosophique. In G. Hottois, *Philosophie et technosciences.* Seyssel (Paris): Champ Vallon (PUF).

Husain, M., & Mehta, M. (2011). Cognitive enhancement by drugs in health and disease. *Cell*, (15): 28-36.

Ilkilic, I. (2009). Coming to grips with genetic exceptionalism: roots and reach of an explanatory model. *Medicine Studies, 1*, 131-142.

Infosys Insights. (2016). *The future for industrial services: the digital twin.* Infosys Ltd.

Italian Presidency of the Council of the European . (2014). *Rome Declaration on Responsible Research and Innovation.*

Jasanoff, S. (2016). *The ethics of invention: technology and the human future.* New York: W.W. Norton & Company.

Johnson, C., Balmford, A., Brook, B., Buettel, J., Galetti, M., Guangchun, L., & Wilmshurst, J. (2017). Biodiversity losses and conservation responses in the Anthropocene. *356*, 270-275.

Kahane, G., & Savulescu, J. (2015). Normal human variation: refocusing the enhancement debate. *Bioethics*, 29(2): 133-143.

Katz, F. (2011). Proposal for a microbial semi-commons:perspectives from the international cooperative biodiversity groups. In P. Uhlir, *Designing the microbial research commons.* Washington D.C.: National Academic Press.

Keeling, C., Bacastow, R., Bainbridge, A., Ekdahl, C. J., Guenther, P., Waterman, L., & Chin, J. (1976). Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii. *Tellus, 28*(6), 538-551.

Knoppens, B. (2009). Genomics and policymaking: from static models to complex systems? *Hum. Genet., 125*, 375-379.

Kritikos, M. (2020). *Ten technologies to fight coronavirus.* European Parliamentary Research Service (EPRS) - Scientific Foresight Unit STOA.

Kummar, S., Williams, P., Lih, C.-J., Polley, E., Chen, A. P., Rubinstein, L. V., . . . Doroshow, J. (2015). Application of Molecular Profiling in Clinical Trials for Advanced Metastatic Cancers. *JNCI J Natl Cancer Inst*, 107(4): djv003.

Kupferschmidt, K. (2011). Danish archipelago launches mass sequencing plan. *Science*.

Kuzma, J., Gould, F., Brown, Z., Collins, J., Delborne, J., Frow, E., . . . Stauffer, S. (2017). A roadmap for gene drives: using institutional analysis and development to frame research needs and governance in a systems context. *J. Resp. Innov., 5(sup1)*, S13-S39.

Lange, K., R. S., Lange, K., Tucha, L., & Tucha, O. (2010). The history of attention deficit hyperactivity disorder. *Atten Defic Hyperact Disord.*, 2(4): 241–255.

Le Hellard, S., & Steen, V. (2014). Genetic architecture of cognitive traits. *Scand. J. Psychol., 55*(3), 255-262.

Ledford, H. (2020). The coronavirus outbreak could make it quicker and easier to trial drugs. *Nature, 582*, 172.

Lehoux, P., Silva, H., Sabio, R., & Roncarolo, F. (2018). The unexplored contribution of responsible innovation in health to sustainable development goals. *Sustainability, 10*(4015).

Lehrach, H., Ionescu, A., & Benhabiles, N. (2016). *The future of health care: deep data, smart sensors, virtual patients and the Internet-of-humans.*

Leslie, D. (2020). Tackling COVID-19 through responsible AI innovation: five steps in the right direction. *Harvard Data Science Review*.

Levin, N., Leonelli, S., Weckowska, D., Castle, D., & Dupré, J. (2016). How do scientists define openness? Exploring the relationship between open science policies and research practice. *Bulletin of Science, Technology & Society, 36*(2), 128 –141.

Lewin, H., Robinson, G., Kress, W., Baker, W., Coddington, J., Crandall, K., . . . Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. *PNAS*.

Lewis, S., & Maslin, M. (2015). Defining the Anthropocene. *Nature, 519*, 171–180.

Lezaun, J., & Montgomery, C. (2015). The pharmaceutical commons: Sharing and exclusion in global health drug development. *Sci. Technol. Human Values, 40*(1), 3-29.

Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., . . . Xie, X. (2015). CRISPR/Cas9-mediated gene editing in human. *Protein Cell*, 6(5):363–372.

Liddicoat, J., Liddell, K., McCarthy, A., Hogarth, S., Aboy, M., Nicol, D., . . . Hopkins, M. (2019). Continental drift? Do European clinical genetic testing laboratories have a patent problem? *European Journal of Human Genetics*, 27 : 997–1007.

Living Factories. (2017). *Synthetic Biology as an enabler of sustainable bioeconomy - A roadmap for Finland.* Living Factories.

Lloyd-Price, J., Abu-Ali, G., & Huttenhower, C. (2016). The healthy human microbiome. *Genome Medicine, 8:51*.

Loeffler, M., Hinrichs, J., Moß, K., Henkel, M., Hausmann, R., Kruse, A., . . . Wodarz, S. (2018). Processing of Biobased Resources. In L. I., *Bioeconomy*. Cham.: Springer.

Lubberink, R., Blok, V., Van Ophem, J., & Omta, O. (2017). A framework for responsible innovation in the business context: Lessons from responsible-, social- and sustainable innovation. In L. Asveld, R. Dam-Mieras, T. Swierstra, S. Lavrijssen, K. Linse, J. van den Hoven, L. Asveld, R. van Dam-Mieras, T. Swierstra, S. Lavrijssen, K. Linse, & J. van den Hoven (Eds.), *Responsible Innovation 3: A European agenda?* Springer International Publishing AG.

Lunshof, J., Bobe, J., Aach, J., Angrist, M., Thakuria, J., Vorhaus, D., . . . Church, G. (2010). Personal genomes in progress: from the Human Genome Project to the Personal Genome Project. *Dialogues Clin. Neurosci., 12*(1), 47–60.

Lunshof, J., Chadwick, R., Vorhaus, D., & Church, G. (2008). From genetic privacy to open consent. *Nature, 9*, 406-411.

Macnaghten, P., Owen, R., & Jackson, R. (2016). Synthetic biology and the prospects for responsibleinnovation. *Essays in Biochemistry*, 60347–60355.

Madison, M. J., Frischmann, B. M., & Strandburg, K. J. (2010). Constructing Commons in the Cultural Environment. *Cornell Law Review, 95*, 657-710.

Malin, B. (2006). Re-identification of familial database records. *AMIA Annu. Symp. Proc.*, (pp. 524–528).

Malin, B., Loukides, G., Benitez, K., & Clayton, E. (2011). Identifiability in biobanks: models, measures, and mitigation strategies. *Hum. Genet., 130*(3), 383-392.

Manders-Huits, N. (2010). Practical versus moral identities in identity management. *Ethics Inf. Technol., 12*, 43-55.

Manders-Huits, N., & van den Hoven, J. (2008). Moral identification in identity management systems. In S. Fischer-Hübner, P. Duquenoy, A. Zuccato, & L. Martucci, *The future of identity in information society (IFIP International Federation for Information Processing)* (pp. 262, 77-91). Boston: Springer.

Marcandalli, J., Fiala, B., Ols, S., & al., e. (2019). Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. *Cell, 176*(6), 1420-1431.

Marx, V. (2015). The DNA of a nation. *Nature, 524*, 503-505.

Menuz, V., Hurlimann, T., & Godard, B. (2013). Is human enhancement also a personal matter? *Sci. Eng. Ethics*, 19:161-177.

Meyer, M. (2013). Assembling, governing, and debating an emerging science: the rise of synthetic biology in France. *Biology in History, 63*(5), 373-379.

Miller, W., & Jaffe, A. (2016). Biomarkers in heart failure: the importance of inconvenient details. *ESC Heart Failure*, 3: 3-10.

NHC. (2020). *National Health Commission of the People's Republic of China*. Retrieved 2020, from http://en.nhc.gov.cn/2020-04/06/c_78861.htm

Nobre, C., Sampaio, G., Borma, L., Castilla-Rubio, J., Silva, J., & Cardoso, M. (2016). Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. *PNAS, 113*(39), 10759–10768.

Novell, J. (1990). From Da Vinci to Harvey: the development of mechanical analogy in medicine from 1500 to 1650. *J. Royal Soc. Med.* , 396-398.

OECD. (2020). *Why open science is critical to combatting COVID-19.* OECD.

Olena, A. (2020). Two Genetic Regions Linked with Severe COVID-19. *The Scientist*.

Ostrom, E. (1990). *Governing the Commons - The evolution of institutions for collective action.* Cambridge: Cambridge University Press.

Ostrom, E. (1999). Coping with the tragedy of the commons. *Annu. Rev. Polit. Sci., 2*, 493–535.

Ostrom, E. (2005). *Understanding institutional diversity.* Princeton, NJ: Princeton University Press.

Owen, R., Bessant, J., & Heintz, M. (2013). *Responsible innovation: managing the responsible emergence of science and innovation in society.* Chichester: John Wiley & Sons, Ltd.

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research andinnovation: From science in societyto science for society, with society. *Science and Public Policy, 39*, 751–760.

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013). A framework for responsible innovation. In R. Owen, J. Bessant, & M. Heintz, *Responsible innovation: Managing the responsible emergence of science and innovation in society* (pp. 27-50). John Wiley & Sons, Ltd.

Oye, K., Esvelt, K., Appleton, E., Catteruccia, F., Church, G., Kuiken, T., . . . Collins, J. (2014). Regulating gene drives. *Science, 345*(6197), 626-628.

Pálsson, G., & Prainsack, B. (2011). Genomic stuff: governing the (im)matter of life. *International Journal of the Commons*, 5, 259-283.

Paddon, C., & Keasling, J. (2014). Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nature Reviews Biotechnology*, (12): 355.

Padmanabhan, A., & Luna, B. (2014). Developmental imaging genetics: linking dopamine function to adolescent behavior. *Brain and Cognition, 89*, 27-38.

Pagano, U. (1999). Is Power an Economic Good? Notes on Social Scarcity and the Economics of Positional Goods. In F. M. Bowles S., *The Politics and the Economics of Power* (pp. 116-45). London: Routledge.

Peeri, N., Shrestha, N., Rahman, S., Zaki, R., Tan, Z., Bibi, S., . . . Haque, U. (2020). The SARS, MERS and novel coronavirus(COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *International Journal of Epidemiology*, 1–10.

PGP Consent Form. (2014). Retrieved from Personal Genome Project: http://www.personalgenomes.org/harvard/sign-up

Pietsch, W. (2015). Aspects of theory-ladenness in data-intensive science. *Philosophy of Science, 82*(5), 905-916.

Piller, C., & Travis, J. (2020, June 12). Authors, elite journals under fire after major retractions. *Science, 368*(6496), 1167-1168.

Plümecke, T. (2016). Genes, symptoms and the "asymptomatic ill": towards a broader understanding of genetic discrimination. *New Genetics and Society*, 35(2): 124-148.

PMI working group. (2015). *The precision medicine initiative cohort program - Buildnig a research foundation for the 21st century medicine.*

Poteete, A. J. (2010). *Working together - Collective action, the commons, and multiple methods in practice.* Princeton NJ: Princeton University Press.

Potts, J. (2018). Governing the innovation commons. *Journal of Institutional Economics, 14*(6), 1025–1047 .

Prainsack, B. (2019). Logged out: Ownership, exclusion and public value in the digital data and information commons . *Big Data & Society, 6*(1).

President's council on bioethics. (2003). *Beyond therapy: Biotechnologie and the pursuit of happiness.* Washington, DC: Dana Press.

Quah, D. (2002). Digital goods and the new economy. *CEP discussion paper*.

Raimbault, B., Cointet, J.-P., & Joly, P.-B. (2016). Mapping the Emergence of Synthetic Biology. *PLOS ONE*.

RDA COVID-19 Working Group. (2020). *Recommendations and Guidelines on data sharing.* Research Data Alliance.

Roche, P., & Annas, G. (2001). Protecting genetic privacy. *Nature Review Genetics, 2*, 392-396.

Rodríguez, A. R., & Sotomayor, O. (2019). *Towards a sustainable bioeconomy in Latin America and the Caribbean: elements for a regional vision.* Santiago: Natural Resources and Development series, N°191, Economic Commission for Latin America and the Carribean (ECLAC).

Sachsenmeier, P. (2016). Industry 5.0 - The relevance and implications of bionics and synthetic biology. *Engineering*, 225-229.

Santoni de Sio, F., Robichaud, P., & Vincent, N. (2014). Who Should Enhance? Conceptual and Normative Dimensions of Cognitive Enhancement. *Humana.Mente Journal of Philosophical Studies*, (26) 179-197.

Sarata, A. (2008). *Genetic exceptionalism: genetic information and public policy.* Congressional Research Service, HSDL.

Sarwate, A., Plis, S., Turner, J., Arbabshirani, M., & Calhoun, V. (2014). Sharing privacy-sensitive access to neuroimaging and genetics data: a review and preliminary validation. *Front. Neuroinform., 8*, 35.

Savulescu, J., Foddy, B., & Clayton, M. (2004). Why should we allow performance enhancing drugs in sport. *British Journal of Sports Medicine*, 38: 666-670.

Schumpeter, J. (1943). *Capitalism, socialism and democracy.* Taylor & Francis e-Library, 2003.

Scoles, S. (2016). A Digital Twin of Your Body Could Become a Critical Part of Your Health Care. *http://www.slate.com*.

Sedyaningsih, E., Isfandari, S., Soendoro, T., & Supari, S. (2008). Towards Mutual Trust, Transparency and Equity in Virus Sharing Mechanism:The Avian Influenza Case of Indonesia. *Ann Acad Med Singapore, 37*, 482-8.

Shin, M., Shukla, S., Chung, Y., Beiss, V., Khim Chan, S., Ortega-Rivera, O., . . . Steinmetz, N. (2020, Aug). COVID-19 vaccine development and a potential nanomaterial path forward. *Nature Nanotechnology, 15*, 646–655.

Si, T., & Zhao, H. (2016). A brief overview of synthetic biology research programs and roadmap studies in the United States. *Synthetic and Systems Biotechnology*, 1-7.

Siemens. (2015). *The digital twin - Digitalization in machine building.* Siemens.

Silva, H., Lehoux, P., Miller, F., & Denis, J.-L. (2018). Introducing responsible innovation in health: a policy-oriented framework. *Health Research Policy and Systems, 16*(90).

Soini, S. (2012). Genetic testing legislation in Western Europe - a fluctuating regulatory target. *3*(2), 143–153.

Steinhubl, S., Muse, E., Barrett, P., & Topol, E. (2016). Digital medicine. Off the cuff: rebooting blood pressure treatment. *The Lancet*, 749.

Stephens, Z., Lee, S., Faghri, F., Campbell, R., Zhai, C., & Efron, M. (2015). Big Data: Astronomical or Genomical? *PLOS Biology*, DOI:10.1371/journal.pbio.1002195.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy, 42*, 1568–1580.

Strandburg, K., Frischmann, B., & Madison, M. (2017). Governing Medical Knowledge Commons (Cambridge Studies on Governing Knowledge Commons). Cambridge: Cambridge University Press.

Sweeney, L., Abu, A., & Winn, J. (2013). *Identifying participants in the personal genome project by name.* Data Privacy Lab, IQSS, Harvard University.

SynBioBeta. (2019). *Synthetic Biology Investment Report 2019 Q2.* SynBioBeta.

Synthetic Biology Leadership Council. (2016). *Biodesign for the bioeconomy: UK synthetic biology strategic plan 2016.* SBLC.

Telenti, A., Pierce, L., Biggs, W., di Iulio, J., Wong, E., Fabani, M., . . . Venter, J. (2016). Deep sequencing of 10,000 human genomes. *PNAS*, 113 (42): 11901–11906.

Temkin, L. (2011). Is living longer living better? In J. Savulescu, R. ter Meulen, & G. Kahane, *Enhancing Human Capacities.* Hoboken (New Jersey): Wiley-Blackwell.

Thanh Le, T., Andreadakis, Z., Kumar, A., Román, R., Tollefsen, S., Saville, M., & Mayhew, S. (2020). The COVID-19 vaccine development landscape. *Nature Reviews - Drug Discovery*, 305-306.

The 100.000 Genomes Project. (2012). *Protocol v3.*

The COVID-19 Host Genetics Initiative. (2020). The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *European Journal of Human Genetics, 28*, 715–718.

The Earth Microbiome Project Consortium. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature, 551*, 457-465.

Timmermans, J., Zhao, Y., & van den Hoven, J. (2011). Ethics and Nanopharmacy: Value Sensitive Design of New Drugs . *Nanoethics, 5*, 269–283 .

Torrance, A. (2017). Better to give than to receive: an uncommon commons in synthetic biology. In B. Frischmann, K. Strandburg, & M. Madison, *Governing medical knowledge commons* (pp. 192-221). Cambridge: Cambridge University Press.

Uhlir, P. (2011). Designing the digital commons in microbiology - Moving from restrictive dissemination of publicly funded knowledge to open knowledge environments: A case study in microbiology. In P. Uhlir, *Designing the microbial research commons.* Washington DC: The National Academic Press.

UK Biobank. (2020). Retrieved 2020, from https://www.ukbiobank.ac.uk/2020/04/covid/

UNESCO. (2003). Retrieved from International Declaration on Human Genetic Data: http://portal.unesco.org/en/ev.php-URL_ID=17720&URL_DO=DO_TOPIC&URL_SECTION=201.html

United Nations. (1998). Retrieved from Universal Declaration on the Human Genome and Human Rights: www.ohchr.org/EN/ProfessionalInterest/Pages/HumanGenomeAndHumanRights.aspx

Urquiza-Garcia, U., Zielinski, T., & Millar, A. (2019). Better research by efficient sharing: evaluation of free management platforms for synthetic biology designs. *Synthetic Biology, 2019, 4*(1).

van de Poel, I., & Fahlquist, J. N. (2012). Risk and Responsibility. In S. Roeser, R. Hillerbrand, P. Sandin, & M. Peterson, *Essentials of Risk Theory* (pp. 107-143). Springer.

van den Hoven, J. (2008). Information technology, privacy, and the protection of personal data. In J. van den Hoven, & J. Weckert, *Information Technology and Moral Philosophy (Cambridge Studies in Philosophy and Public Policy)* (pp. 301-321). Cambridge.

van den Hoven, J. (2013). Value sensitive design and responsible innovation. In R. Ower, J. Bessant, & M. Heintz, *Responsible innovation: Managing the responsible emergence of science and innovation in society* (pp. 75-83). John Wiley & Sons, Ltd.

van den Hoven, J. (2014). Responsible Innovation: A New Look at Technology and Ethics. In J. van den Hoven, N. Doorn, T. Swierstra, B.-J. Koops, & H. Romijn, *Responsible Innovation 1: Innovative Solutions for Global Issues.* Dordrecht: Springer.

van den Hoven, J., Helbing, D., & Domingo-Ferrer, J. C. (2013). FuturICT - The road towards ethical ICT. *The European Physical Journal Special Topics*.

van den Hoven, J., Jacob, K., Nielsen, L., Roure, F., Rudze, L., Stilgoe, J., . . . Martinez Riera, C. (2013). *Options for strengthening responsible research and innovation - Report of the expert group on the state of art in Europe on responsible research and innovation.* Directorate-General for Research and Innovation Science in Society, European Commission. Luxembourg: Publications Office of the European Union.

van den Hoven, J., Lokhorst, G.-J., & Van de Poel, I. (2012). Engineering and the problem of moral overload. *Sci. Eng. Ethics*, 143-155.

van den Hoven, M. (2008). Information technology, privacy, and the protection of personal data. In M. van den Hoven, & J. Weckert, *Information technology and moral philosophy* (pp. 301-321). Cambridge University Press.

Varshney, R., Terauchi, R., & McCouch, S. (2014). Harvesting the promising fruits of genomics: Applying genome sequencing technologies to crop breeding. *PLoS Biol, 12*(6).

Vatiero, M. (2009). Positional goods: a diagrammic exposition. *Quaderni del Dipartimento di Economia Politica*, vol 575. Retrieved from http://ssrn.com/abstract=151904

Vaught, J. (2020). Biobanking During the COVID-19 Pandemic. *Biopreservation and Biobanking, 18*(3).

Velleman, D. (2001). The genesis of shame. *Philosophy & Public Affairs, 30*(1), 27-52.

Vinatzer, B., Heath, L., Almohri, H., Stulberg, M., Lowe, C., & Li, S. (2019). Cyberbiosecurity Challenges of Pathogen Genome Databases. *Front. Bioeng. Biotechnol.*

Von Schomberg, R. (2011). Prospects for Technology Assessment in a framework of responsible research and innovation. In *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methode* (pp. 39-61). Wiesbaden: Springer VS.

Wang, F., & Zhang, W. (2019). Synthetic biology: Recent progress, biosafety and biosecurity concerns, and possible solutions. *J. Biosafety and Biosecurity, 1*(1), 22-30.

Warnke, P., Cuhls, K., Schmoch, U., Daniel, L., Andreescu, L., Dragomir, B., . . . Kuusi, O. (2019). *European commission Foresight report: 100 radical innovation breakthroughs for the future.* Luxembourg: Publications Office of the European Union.

Whitehouse, P., Juengst, E., Mehlman, M., & Murray, T. (1997). Enhancing cognition in the intellectually intact. *Hastings Center Report*, 27(3): 14-22.

Wilkerson, T. (1995). *Natural kinds.* Avebury: Ashgate Publishing Limited.

Williams, M., & Hall, J. (2015). Hackerspaces: a case study in the creation and management of a common pool resource. *Journal of Institutional Economics, 11*(4), 769–781.

Willour, V., Seifuddin, F., Mahon, P., Jancic, D., Pirooznia, M., Steele, J., . . . J.B., P. (2012). A genome-wide association study of attempted suicide. *Molecular Psychiatry, 17*, 433–444.

Yozwiak, N., Schaffner, S., & Sabeti, P. (2015). Data sharing: Make outbreak research open access. *Nature*, 477–479.

Zhuang, L., Huang, S., Liu, W.-Q., Karim, A., Jewett, M., & Li, J. (2020). Total in vitro biosynthesis of the nonribosomal macrolactone peptide valinomycin. *Metabolic Engineering Volume, 60*, 37-44.

Zinnbauer, D. (2018). *Information Wants to Be Expensive, Not Free! And this is Bad for Justice, Democracy, the Economy.* Retrieved from SSRN: https://ssrn.com/abstract=3125074 or http://dx.doi.org/10.2139/ssrn.3125074

Index

Summary

Innovations in biotechnology increasingly shape our societies and our planet. The stream of innovations that could be witnessed in the past few decades opens up new ways to do agriculture, to provide healthcare and to produce compounds and materials, amongst many other things. Many of these innovations rely on biological data. The ability to extract a plethora of biodata vastly increased over the past few decades. These biodata provide deeper insights into the workings of biological systems, thus constituting a fertile ground for bio-inspired innovations. For example, population genomics data provides the basis for a personalized health care. Biodiversity data derived from ecosystems provides the basis for the identification of novel drugs, high value chemicals and materials. Biodata is increasingly crucial when aiming for a flourishing bio-economy and biomedicine.

Biodata-based innovations thereby raise very substantial ethical questions. Pronounced cases like human genome editing, or the engineering of entire species via gene drive technologies, make clear that innovations need to go hand in hand with societal deliberation and an ethical accompaniment of technology development. The question though is how such responsible innovation can be organized. Extraction of biodata is done at a speed that surpasses Moore's law. And the resulting biodata-based innovations are fast-paced. It is therefore highly needed to consider how a responsible guidance of innovation in biotechnology can be accomplished, in view of the biodata-avalanche. This question provides the entry point for this dissertation. Central to this analysis is the special ontological and epistemological position of biodata. Biodata resides at the interface between the biophysical world and the realm of human language and meaning. This makes biodata a central locus when pursuing a value-driven accompaniment of innovation in the field of biotechnology.

In Chapter 2, the effect of this special position of biodata is analyzed in the case of personal genomic data. The increasing availability of genomic data – and more generally 'personal big biodata' (which comprises a wide variety of medical and health care data covering both medical images and a panoply of biomarkers) – combined with the computational power and analytical tools of bioinformatics calls for a rethinking of how to use these data responsibly. In particular, arguments are brought forward on why genomic privacy contributes to an ethical usage of genomic data. These arguments underpin genomic exceptionalism, the thesis that personal

genomic data have a special status compared to other types of medical data. Personal genomics data are intimately related to the biophysical reality of a person, and on the other hand provide a substrate for cultural interpretations and actions. This implies the responsible appraisal of the limits of what data allow us to know about individuals, and the acknowledgement that respect for human persons and their dignity requires taking the epistemic position that there is always more to know about them than even the most comprehensive set of data may offer in terms of knowledge. The ideal of acknowledging the limits of our knowledge of persons is referred to as 'epistemic modesty'. Privacy is instrumental in ensuring a person's ability for self-determination in view the acts of epistemic immodesty of others. In addition to ex ante approaches to data protection such as privacy enhancing technologies, attention is also drawn to the potential of ex post approaches, e.g. tools and mechanisms supporting proper and modest interpretations of genomic data, as well as the detection of prohibited use of genomic data for certain ends.

Chapter 3 provides an analysis of the impacts of personalized models of human biology on core notions in health care, like therapy, preventative care and human enhancement. Emerging data-driven health care practices increasingly build on large and sometimes continuous data streams derived from individual persons. Digital twins stand for an engineering paradigm in which individual physical artefacts are paired with a digital model that dynamically reflects the status of that artefact. When applied to persons, digital twins are used as the hypothesis that one would be in the possession of *in silico* representations that dynamically reflect aspects of an individual's molecular status, physiological status and life style over time. This perspective redefines the concept of 'normality' or 'health', as a set of patterns that are regular for a particular individual, against the backdrop of patterns observed in the population. This perspective can induce a normative shift in how therapy, preventative care and enhancement are distinguished. This normative shift relates to how meaning is derived from measurement data. A promiscuous realist account is used to clarify how moral distinctions based on digital twins are the result of both cuts made at Nature's joints, and the world of language and meaning that is grafted on these structures.

Chapter 4 focuses on responsible innovation in the case of biodiversity sequencing data. Large scale sequencing initiatives increasingly give access to biodiversity at the molecular level, providing a rich source for innovations. Biological data as an information resource can have characteristics of a Common Pool Resource, when used and managed by multiple stakeholders. Communities that are centered around

such digitalized ecosystems or biological populations have aspects of both natural resource commons and knowledge commons, but differ in their structure and dynamics. The concept of "Twin Commons" is proposed: the institutional arrangement of biological resources that have a tightly linked digital component which is shared and governed by a community, and that have research and innovation as important outlets. Twin commons can help in the identification and the inclusion of values in the socio-economical and techno-scientific systems that underpin the data-driven innovations. This can complement stakeholder involvement as a strategy in a responsible innovation approach.

Chapter 5 analyses which are the effects of hurdles in biodata accessibility on the ability to organize responsible innovation. Data accessibility and differences in data-usage capabilities can be important factors in shaping the innovation landscape. Data frictions can turn biodata into positional information goods — goods whose value depends on exclusivity – be it because of manufactured scarcity or of other types of data friction. From both an ethical and economic point of view, positionality can lead to suboptimal as well as beneficial situations. When aiming for responsible innovation (i.e. embedding societal deliberation in the innovation process), it is important to consider hurdles and facilitators in data access and use. These effects are made tangible in the case of synthetic biology innovations in response to the COVID-19 pandemic. Differences or delays in data access impact the ability to innovate, as well as differences in the ability to put data to use. The roles of governance and of collective self-regulation in commons is highlighted as venues towards responsible innovation, and opportunities to mitigate negative positional effects are identified.

Samenvatting

Biotechnologische innovaties hebben een groeiende impact op onze samenleving en op de planeet. Deze stroom aan innovaties opent nieuwe wegen voor belangrijke sectoren zoals landbouw, gezondheidszorg, en de productie van chemicalieën en materialen. In de voorbije decennia zijn technologieën ontwikkeld die grote hoeveelheden biologische data produceren. De data geven ons meer inzicht in biologische systemen. Dit vormt een rijke voedingsbodem voor bio-geïnspireerde innovaties. Gepersonaliseerde gezondheidszorg baseert zich bijvoorbeeld op genetische data van grote bevolkingsgroepen. Biodiversiteits-data van ecosystemen zijn een bron voor het ontwikkelen van nieuwe geneesmiddelen, complexe chemicaliën en materialen. Biologische gegevens worden steeds belangrijker bij het uitwerken van een bloeiende bio-economie en geneeskunde.

Innovatie op basis van biologische gegevens roept belangrijke ethische vragen op. Uitgesproken voorbeelden - zoals ingrepen in het menselijk genoom, of aanpassingen in een diersoort via gene drive-technologieën - maken duidelijk dat innovaties hand in hand moeten gaan met maatschappelijk overleg en ethische begeleiding. De vraag is hoe zo'n verantwoorde innovatie kan worden georganiseerd. Biologische gegevens worden gegenereerd tegen een snelheid die de wet van Moore overtreft, en ook de innovaties die hieruit resulteren evolueren snel. Nadenken over verantwoorde begeleiding van op data gebaseerde innovatie in de biotechnologie is daarom bijzonder belangrijk. Dit vormt het startpunt voor dit proefschrift. Centraal in de analyse staat de bijzondere ontologische en epistemologische positie van biodata. Biologische gegevens bevinden zich op het raakvlak tussen de biofysische wereld en het rijk van de betekenissen die we eraan toekennen. Biodata is dus een belangrijk element bij het organiseren van verantwoorde innovatie in de biotechnologie.

In hoofdstuk 2 worden de effecten van deze bijzondere positie van biodata geanalyseerd. De toenemende beschikbaarheid van persoonlijk genomische data maakt het nodig om te bekijken wat verantwoord gebruik van deze gegevens inhoudt. Dit wordt vooral belangrijk wanneer 'persoonlijke big biodata' gecombineerd worden met de rekenkracht en analytische instrumenten van de bioinformatica. 'Persoonlijke big biodata' omvatten een breed scala aan medische en gezondheidsgegevens, met zowel medische beelden als een scala aan biomarkers. Om een ethisch gebruik te garanderen is genomische privacy belangrijk. De reden ligt in genomisch

exceptionisme: de speciale status van persoonlijke genetische gegevens. Persoonlijke genetische gegevens zijn nauw verbonden met de fysieke persoon, én ze zijn een substraat voor culturele interpretaties en acties. Er is dus een verantwoorde beoordeling nodig van wat de data ons al dan niet over een individu kunnen vertellen. Respect voor de waardigheid van een persoon vereist het epistemische standpunt dat er altijd meer over die persoon te weten valt dan dat zelfs de meest uitgebreide dataset vertelt. Deze 'epistemische bescheidenheid' bestaat uit het erkennen van de grenzen van onze kennis. Privacy is hierbij essentieel, omdat dit het vermogen tot zelfbeschikking waarborgt, met het oog op de epistemische onbescheidenheid van anderen. Naast ex ante-benaderingen van gegevensbescherming, zoals privacybevorderende technologieën, wordt ook de aandacht gevestigd op het potentieel van ex-postbenaderingen, bv. hulpmiddelen en mechanismen die bescheiden interpretaties van genomische gegevens ondersteunen, evenals de detectie van het misbruik van genomische gegevens.

Hoofdstuk 3 bevat een analyse van de impact van gepersonaliseerde computermodellen (Digital Twins) op kernbegrippen in de gezondheidszorg, zoals therapie, preventieve zorg en menselijke verbetering. Nieuwe praktijken in de geneeskunde gebruiken in toenemende mate data over individuele personen. Digital twins staan voor een paradigma in de ingenieurstechnieken waarin een digitaal model heel precies de status van een individueel artefact weergeeft. Digital Twins in de geneeskunde staan voor de hypothetische situatie waarbij men computermodellen heeft van aspecten van de moleculaire status, fysiologische status en levensstijl van individuele personen. Wat dan 'normaal' of 'gezond' is wordt dan geherdefiniëerd als een reeks patronen die kenmerkend zijn *voor een bepaald individu*, tegen de achtergrond van patronen die in de populatie worden waargenomen. Dit kan een normatieve verschuiving teweegbrengen in de concepten van therapie, preventieve zorg en menselijk verbetering. Deze normatieve verschuiving heeft betrekking op hoe betekenis wordt ontleend aan data. Het concept 'promiscu realisme' wordt gebruikt om te verduidelijken hoe een moreel onderscheid kan ontstaan ten gevolge van Digital Twins. Ze vormen een manier om menselijke populaties op een bepaalde manier in categorieën op te delen, en deze van betekenis te voorzien.

Hoofdstuk 4 richt zich op de vraag hoe verantwoord te innoveren op basis van biodiversiteits-data. Grootschalige wetenschappelijke initiatieven geven een beeld van biodiversiteit op moleculair niveau. Dit vormt een rijke bron voor innovaties. Biodiversiteits-data hebben de kenmerken van een gemeenschappelijk goed, wanneer ze gebruikt en beheerd worden door meerdere belanghebbenden.

Maatschappelijke instituten die zich vormen rond dergelijke gedigitaliseerde ecosystemen of biologische populaties hebben aspecten van commons. Ze verschillen echter in hun structuur en dynamiek van commons die natuurlijke hulpbronnen beheren, en van kennis-commons. "Twin commons" wordt daarom voorgesteld als raamwerk. Twin commons zijn instituten die gemeenschappelijke gedigitaliseerde biologische hulpbronnen beheren, en die een focus hebben op onderzoek en innovatie. Twin commons kunnen helpen bij het identificeren van waarden, en het opnemen van waarden in de sociaaleconomische en technisch-wetenschappelijke context waarin de innovaties plaatsvinden. Dit kan – naast het betrekken van een brede groep van belanghebbenden - een belangrijk element vormen in een verantwoorde innovatieaanpak.

Hoofdstuk 5 analyseert hoe de graad van toegankelijkheid van biodata een effect heeft op het vermogen tot verantwoorde innovatie. De mate van toegankelijkheid van gegevens en de mogelijkheden om deze te analyseren, vormen in belangrijke mate het innovatielandschap. Fricties in de beweging van data kunnen deze data veranderen in positionele informatiegoederen. De waarde van positionele informatiegoederen hangt af van hun exclusiviteit – en die kan veroorzaakt worden door kunstmatig gecreëerde schaarste of door andere soorten data-wrijving. Zowel vanuit ethisch als economisch oogpunt kan positionaliteit leiden tot suboptimale, en soms ook tot gunstige situaties. Bij het streven naar verantwoorde innovatie (d.w.z. het verankeren van maatschappelijke afweging in het innovatieproces), is het dus belangrijk om rekening te houden met hindernissen en facilitatoren bij datatoegang en -gebruik. Innovaties op het gebied van synthetische biologie als reactie op de COVID-19-pandemie worden als voorbeeld gebruikt. Verschillen in datatoegankelijkheid en in analyse capaciteit hebben invloed op het vermogen om te innoveren. De rol van centraal bestuur en van collectieve zelfregulering in commons wordt geëvalueerd als manieren om verantwoorde innovatie vorm te geven. Er worden ook mogelijkheden aangegeven om eventuele negatieve positionele effecten te verzachten.

About the author

Koen Bruynseels (Belgium, °1970) received his education as a bio-engineer at the K.U.Leuven. He obtained a Ph.D. in Medical Sciences at the K.U.Leuven, with a dissertation on hepatic fructose metabolism studied with NMR spectroscopy. A postdoc at the Flemish institute of Biotechnology further got him on the track of biological data analysis. He joined a start-up of the University of Ghent as researcher, mining crop plant molecular data at the initial phase of the innovation process. Later on, he moved into various R&D positions in BASF, being responsible for international computational biology and systems biology groups and for R&D digitalization initiatives. The research focuses on the analysis of biological data and on the development of computational models, with applications in plant biotechnology and industrial biotechnology. In these roles, he established multiple public-private collaborations and got familiar with many facets of the innovation process in biotechnology.

Next to this scientific track, Koen obtained a bachelor degree in philosophy at the Higher Institute of Philosophy (K.U.Leuven). The dissertation work on the role of the future in an ethics of technology spurred the interest in the philosophical and ethical aspects of technology. He joined the 4TU.Centre for Ethics and Technology (at that point in time 3TU) as a Ph.D. candidate at the T.U.Delft, as external promovendus.

Simon Stevin Series in Ethics of Technology
Delft University of Technology, Eindhoven University of Technology, Univeristy of Twente & Wageningen University
Editors: Philip Brey, Anthonie Meijers, Sabine Roeser and Marcel Verweij

Books and Dissertations

Volume 1: Lotte Asveld, *'Respect for Autonomy and Technology Risks'*, 2008

Volume 2: Mechteld-Hanna Derksen, *'Engineering Flesh, Towards Professional Responsibility for 'Lived Bodies' in Tissue Engineering'*, 2008

Volume 3: Govert Valkenburg, *'Politics by All Means. An Enquiry into Technological Liberalism'*, 2009

Volume 4: Noëmi Manders-Huits, *'Designing for Moral Identity in Information Technology'*, 2010

Volume 5: Behnam Taebi, *'Nuclear Power and Justice between Generations. A Moral Analysis of Fuel Cycles'*, 2010

Volume 6: Daan Schuurbiers, *'Social Responsibility in Research Practice. Engaging Applied Scientists with the Socio-Ethical Context of their Work'*, 2010

Volume 7: Neelke Doorn, *'Moral Responsibility in R&D Networks. A Procedural Approach to Distributing Responsibilities'*, 2011

Volume 8: Ilse Oosterlaken, *'Taking a Capability Approach to Technology and Its Design. A Philosophical Exploration'*, 2013

Volume 9: Christine van Burken, *'Moral Decision Making in Network Enabled Operations'*, 2014

Volume 10: Faridun F. Sattarov, *'Technology and Power in a Globalising World, A Political Philosophical Analysis'*, 2015

Volume 11: Gwendolyn Bax, *'Safety in large-scale Socio-technological systems. Insights gained from a series of military system studies'*, 2016

Volume 12: Zoë Houda Robaey, *'Seeding Moral Responsibility in Ownership. How to Deal with Uncertain Risks of GMOs'*, 2016

Volume 13: Shannon Lydia Spruit, *'Managing the uncertain risks of nanoparticles. Aligning responsibility and relationships'*, 2017

Volume 14: Jan Peter Bergen, *Reflections on the Reversibility of Nuclear Energy Technologies*, 2017

Volume 15: Jilles Smids, *Persuasive Technology, Allocation of Control, and Mobility: An Ethical Analysis*, 2018

Volume 16: Taylor William Stone, *Designing for Darkness: Urban Nighttime Lighting and Environmental Values*, 2019

Volume 17: Cornelis Antonie Zweistra, *Closing the Empathy Gap: Technology, Ethics, and the Other*, 2019

Volume 18: Ching Hung, *Design for Green: Ethics and Politics for Behavior-Steering Technology*, 2019

Volume 19: Marjolein Lanzing, *The Transparent Self: a Normative Investigation of Changing Selves and Relationships in the Age of the Quantified Self*, 2019

Volume 20: Koen Bruynseels, *Responsible Innovation in Data-Driven Biotechnology*, 2021

Simon Stevin (1548-1620)

'Wonder en is gheen Wonder'

This series in the philosophy and ethics of technology is named after the Dutch / Flemish natural philosopher, scientist and engineer Simon Stevin. He was an extraordinary versatile person. He published, among other things, on arithmetic, accounting, geometry, mechanics, hydrostatics, astronomy, theory of measurement, civil engineering, the theory of music, and civil citizenship. He wrote the very first treatise on logic in Dutch, which he considered to be a superior language for scientific purposes. The relation between theory and practice is a main topic in his work. In addition to his theoretical publications, he held a large number of patents, and was actively involved as an engineer in the building of windmills, harbours, and fortifications for the Dutch prince Maurits. He is famous for having constructed large sailing carriages.

Little is known about his personal life. He was probably born in 1548 in Bruges (Flanders) and went to Leiden in 1581, where he took up his studies at the university two years later. His work was published between 1581 and 1617. He was an early defender of the Copernican worldview, which did not make him popular in religious circles. He died in 1620, but the exact date and the place of his burial are unknown. Philosophically he was a pragmatic rationalist for whom every phenomenon, however mysterious, ultimately had a scientific explanation. Hence his dictum 'Wonder is no Wonder', which he used on the cover of several of his own books.