SIMPORT

Sovereign and Intuitive Management of Personal Location Information

The SIMPORT Ethics Primer: What does it mean to do ethics in software development?

Gert Goeminne
Rainer Mühlhoff

Bundesministerium für Bildung und Forschung

Funded by the Federal Ministry of Education and Research in the area
Human-Technology Interaction for Digital Sovereignty

WWU MÜNSTER · FH MÜNSTER University of Applied Sciences · UNIVERSITÄT OSNABRÜCK · re:edu Reengineering Education · here

Deliverable:	Ethics by Design V2
Title:	SIMPORT Ethics Primer
Submitted:	12/12/2023
Version:	Version 2
Authors:	Gert Goeminne, Rainer Mühlhoff

Full title: The SIMPORT Ethics Primer: What does it mean to do ethics in software development?

Acknowledgements: the co-authors of this report extend their gratitude to SIMPORT colleagues and partners (*alphabetically)*: Aleksandra Kovacevic, Artur Kim Shim, Chris Kray, Corinna Balkow, Felix Erdmann, Gernot Bauer, Gina Buchwald-Chassée, Jerome Dreyer, Johannes Braese, Matthias Pfeil, Per Gülzow, Simge Oktay, Stefano Bennati, Sven Heitmann, Thomas Bartoschek. We also thank Jan Siebold and Corinna Balkow for critical reading and constructive feedback.

Contents

Introduction: ethics beyond guidelines

Every so often, we are asked, in ways as casual as they are devious, to consent to the collection and processing of our data. The frequency with which we are confronted with this indicates the extent to which our lives have become datafied and digitalised. What we do, think and desire takes shape in a profound way through our interweaving with digital devices. At the same time, the repeated request for our consent suggests that 'something is going on' with the digital traces we leave behind in the process. Others very much want to collect them, they are worth something and, apparently, things may happen with them that we might prefer not to. The consent form makes us raise our eyebrows, but at the same time, one simple click ('I agree') removes it from our minds, and we can go on living our digital lives in peace. Or not quite?

Where the public debate in the past decade mainly focused on privacy in the wake of a number of large-scale scandals such as Snowden's NSA revelations (2013) and the Facebook-Cambridge Analytica Scandal (2018), the recent controversies around language models like chatGPT and Dall.e have broadened it to ask how big data and Artificial Intelligence (AI) are already and will continue impacting the private and public sphere. For better and for worse. Within that debate, which is conducted emphatically in all types of media, the term ethics is falling more and more frequently, to the extent that it has penetrated everyday conversations between concerned citizens. Never before, it seems, ethics of emerging technologies have featured so prominently. If then, as concerned citizens, individuals throw their light online and query Google for the terms ‘ethics' and ‘big data' or 'ethics' and 'AI', they soon find themselves overwhelmed by a tidal wave of ethics guidelines, principles and codes of conduct. For a moment, it seems as if our concerns were unwarranted. Governments, NGOs and companies, all have their version of ethical principles that they put forward in developing ‘ethical’ big data applications. Everything is under control, it seems. Or not quite?

Delving a little deeper and reading some of these guidelines, it soon becomes frustratingly clear that one is going in circles. Time and again, one reads about the need for normative principles and recommendations to harness the 'disruptive potential' of these new technologies and the 'crucial ethical issues' that are thereby raised. Time and again, roughly similar principles are listed that can hardly be reasonably disputed. Who is not in favour of fairness, transparency, accountability, privacy, safety, beneficial impact and human autonomy? All this is typically phrased in an insidiously itemised fashion and neutral language, suggesting that simply applying these principles will resolve the ethical issues raised earlier. And time and again, such guidelines provide little or no practical guidance, and it remains altogether unclear whether these principles have already been applied and if so, what the outcome has been. In a recent controlled survey study, engaging with 63 software engineering students and 105 professional software developers, McNamara et al. (2018) critically reviewed the idea that ethical guidelines serve as a basis for ethical decision-making for software engineers. The painful conclusion was that their effectiveness is almost zero and that they do not change the behaviour of professionals from the tech community. Apparently, we should not be so reassured after all.

In discussing the value of such principle-based approach to ethics, Gogoll et al. (2021) argue that "without sufficient concretization, reference, contextualization and explanation, software engineers are left to themselves in juggling different values and compliance with each other and every one of them." More often than not, the implementation of ethics principles requires a trade-off, for instance between privacy and transparency, as exemplified by Gogoll et al. (2021) with the case of an app that enables geo-tracking. The trade-off between privacy and transparency varies significantly depending on whether this app is used to implement, for instance, an anti-doping regime in professional sports or as a navigation app for finding the optimal route to one's destination. In the former scenario, it might be deemed acceptable for professional athletes to sacrifice more privacy in the battle against illegal doping, whereas we find it disturbing that ordinary users of a navigation app are subjected to continuous tracking and monitoring. Yet, this is exactly the point where ethical deliberation comes into play by asking morally, economically, socially and culturally charged questions about the why and the how of geo-tracking. And here it is important to note that data applications, as well as the data that fuel them, are actually developed in pre-existing infrastructures and scenarios, where large companies are very powerful, have engaged in problematic conduct in the past and at present, and are not always well regulated by governments (Daly 2021). However, these contextual aspects are completely left out of ethical guidelines, which see technical artefacts as isolated entities that can be optimised by experts to find technical solutions to technical problems.

It is our conviction that the omnipresent principle-based approach to ethics is prompted and perpetuated from a simplistic and male-dominated view on innovation as a linear process, in which an 'in-control' agent, undisturbed by social logics and influences, develops the perfect technological solution to a societal problem he has identified. Rather than fundamentally questioning this outdated 'linear innovation model', we argue with Hagendorff (2020) that ethics guidelines amend it with an equally male-dominated, deontological approach to ethics, giving it a schizophrenic twist. Adhering to the idea that the 'perfect' solution exists, it is believed that 'unintended consequences' can be 'designed out' by implementing the 'right' principles. While we will elaborate on this argument in more detail from both theoretical reflections on technology and innovation (Part 1) and from an empirical case study of the SIMPORT project (Part 2), it suffices here to point out some key aspects where ethics guidelines and the linear innovation model mutually align.

- First, the focus on prescriptive rules can be understood from the linear model of innovation in the sense that they are meant to make the 'in-control' agent who is at the helm of the innovation process even a little more 'in-control', now also on an ethical level. In doing so, the focus on rule adherence as the way to success also stands out. In contrast, almost no guideline talks about technology in contexts of care, nurture, help, humility, social responsibility or ecological network (Hagendorff 2020).
- Second, ethics guidelines and the linear model share a quasi-complete neglect for contextual factors, leaving out the complex interplay between innovation and society. Almost none of the guidelines talk about broader power dynamics, or take into account the organisational context such as culturally embedded background assumptions and values that influence the design process (Krijger 2021).

- Third, ethics guidelines and the linear model share a male-dominated solution based approach. Indeed, just as the linear model assumes that innovation is the *a priori* solution, it is noteworthy that almost all guidelines suggest that technical solutions exist for many of the problems described (Hagendorff 2020).
- Fourth and finally, the shared linear-teleological logic stands out. Within the linear innovation model, ethics guidelines are considered an upstream input, leaving no room for a self-reflexive attitude in responding to emerging ethical and social challenges that typically arise over the course of an innovation process.

This report will argue that the superficial deontological approach to ethics, as reflected in the numerous guidelines in circulation, needs to be refuted and replaced by an ethics that goes 'deeper'. Ethics should go deeper by firstly embedding ethical deliberation right into the heart of the development process, where instead of applying ethical principles, critical reflection and cultivated sensitivity on the relation between innovation and society is brought to bear on technical decisions. Second, and relatedly, ethics needs to go deeper by addressing head-on entrenched contextual aspects of technology development such as power dynamics and design values. To this end, this report revisits how technology is, from its most early conception on, always already interwoven with the societal and cultural context. In Part 1, it does so from the outside, descriptively, by replacing the linear innovation model with a Social Sciences and Humanities (SSH) inspired account of technology and innovation as inherently social. This foregrounds the responsibility of the developer, who is called on to respond to the irreducible ambivalence of technology. Part 2 argues that empowering developers to take up this responsibility must start from a renewed self-understanding of their practice as intrinsically linked to the social context. To this end, this report engages, from the inside this time, with the SIMPORT project and the way it has pursued an ethics by design approach. Discussing the collaborator's experiences and probing their attitudes and responsibilities, it is brought to the surface what enables and constrains them in taking up a self-reflexive ethos of 'response-ability'. Part 3, finally, presents inspiring and thought-provoking key recommendations for response-able software development.

Part 1: Ethical and social aspects of emerging technologies

Part 1 presents a selected SSH-vocabulary to articulate, assess and discuss the social nature of technology and what this implies for the responsibility of the designer. In Chapter 1, we take a small tour through the related fields of philosophy and sociology of technology to show how technology cannot possibly be separated from its social dimension. We are convinced that an adequate ethics of software development must be based on a better, clearer and more nuanced understanding of the many ways in which technologies can be described as 'social' or even 'political'. In Chapter 2, we start from this understanding of the social nature of technology to foreground the issue of the social responsibility of technology design. In doing so, we look at the design process as a necessary act of abstraction. In this sense, it is true that technological design can only take place by narrowly focusing on a specific technological affordance. But this does not mean that such an abstraction process is socially neutral, as the linear innovation model implies, quite the contrary. What a design process does or does not take into account, consciously or unconsciously, makes up its radical responsibility.

1. The social nature of technology

"Do artefacts have politics?" It is the title of a 1980 essay by philosopher Langdon Winner in which he argues that this question can at best be rhetorical in nature: of course artefacts have politics (Winner 1980). After all, time and again, technologies, once launched into the world and people interact with them, turn out to be much more than a neutral means to an end. They change our habits, shape our desires and aspirations, open up new worlds and ways of being in the world. And in this sense, they are genuinely political. In what follows, we present a small vocabulary of inspiring SSH concepts that allow us to articulate and discuss the non-neutral, social nature of technologies. In doing so, we will discuss three SSH perspectives on technology that draw attention to three forms of social embeddedness, respectively 'Technological Mediation Theory', 'Science and Technology Studies', and 'Critical Theory of Technology'. In doing so, we gradually zoom out from the micro-level of the context of use in which a technology is always contained, over the meso-level of the societal context in which an innovation has to materialise to the macro-level of cultural values and power relations that largely determine the fate of a technology. Together, they paint a picture of the thoroughly social nature of technology as being characterised by a reciprocal relationship: On the one hand, technology is shaped by societal needs, values and aspirations; on the other, technology itself has the power to shape society by influencing the way people think, act and communicate.

1.1. What things do: Technological Mediation Theory

Technological Mediation Theory (TMT)[1] focuses on the relationship between humans and the world and how, in our modern day, it is almost always mediated by technological artefacts. From the knife we use to butter our morning sandwich to the navigation app that guides us through congested traffic to the solar panels that generate the electricity on which our computer runs: our lives are thoroughly technological. TMT seeks to understand how technology shapes the way we as contemporary humans navigate the world. In this mediating relationship, TMT distinguishes two basic modes, respectively the ‘embodiment' and the 'hermeneutic' relation. We briefly explain these below and link them to human-computer interaction (HCI) terminology of user experiences and user interfaces (UX/UI).

The classic example of an **embodiment relation** that TMT takes from Heidegger is that of the hammer. As long as a hammer is at our disposal, we see the artefact as an object detached from us. However, when we pick up the hammer and use it to drive a nail into the wall, the hammer becomes one with our hand. Focused on hammering, we no longer realise that we are holding a hammer, we experience the world through the hammer. It is only at the moment when, say, the head flies off the handle or we hit our thumb that the hammer returns to our consciousness as a mere - albeit damned - thing. The hammering hammer, to put it in a Heideggerian way, illustrates the startling ability of humans to incorporate as it were artefacts into their body schema, to extend their corporeality through artefacts. In the context of HCI, one may think of

[1] For an extensive introduction see Verbeek (2005).

scrolling and swiping on a tactile screen, where the UI becomes an almost transparent part of UX as a means to achieve some action. Analogous to the hammering hammer, this transparency turns into an opaqueness the second that, say, a raindrop falls on the screen: suddenly, it is all but evident that the image changes when swiping a finger across a glass screen.

The second basic mode of technological mediation TMT distinguishes is the **hermeneutic relation**. In the hermeneutic relation, we are also engaged with the world by means of an artefact, but the artefact is not transparent, for instance when we use a thermometer to find out how warm it is. Here, the artefact does not withdraw from our relationship with the world, but provides a representation of it, which requires interpretation to communicate something about the world. Because interpretation is required in this relation (the thermometer must be 'read'), it is called hermeneutic. Thus, in the hermeneutic relation, the world is not perceived through an artefact but rather by means of it. An example of a hermeneutic relation from HCI is the use of a navigation app. When we 'read' a navigation app, we are not concerned with the app but with the world of which the app's UI reveals certain aspects, namely specific features of all kinds of locations. By interpreting these, the world, by means of the UI, unfolds itself as a collection of interesting and less interesting destinations connected in efficient and less efficient ways. Here, the UI is not integrated 'transparently' into the UX, but represents certain features of the technology through which we interact with the world: buttons represent actions; fonts and their sizes may express a hierarchy; emoticons suggest emotions.

Besides the embodiment relation and the hermeneutic relation, TMT further distinguishes the alterity relation and the background relation, which are, as it were, extreme forms of the hermeneutic and the embodiment relation, respectively. In the **alterity relation**, we as humans are engaged with the technology much like a 'quasi-other'. Think, for example, of setting the address in a navigation app. At that point, the UI acts as a quasi-other with which we have to interact. The inverse happens in the **background relation**, where the technology completely disappears from our immediate perception and becomes part of the background of our experience. As an example in a HCI context, one can think of algorithmic manipulations that imperceptibly co-shape the UI. In this way, the four modes of technological mediation allow themselves to be arranged on a spectrum that extends from the alterity relation via the hermeneutic and the embodiment relation to the background relation. It is interesting to note how the two ends of this spectrum rely on each other; they are, in a sense, each other's condition of possibility. What comes to the fore in the embodied relation, such as the act of hammering, disappears into the background in the hermeneutic relation, where, for example, it is the components of the hammer that come to the fore when repairing a broken hammer. And vice versa.

The importance of the TMT view of technology is that it shows how, as Verbeek (2011) puts it, "while fulfilling their function, technologies do more [than nothing]: they give shape to what we do and how we experience the world". The critical point here is that the mediating relationship transforms both our perception and our ability to act according to a regular pattern of amplification and reduction. Binoculars, for example, simultaneously amplify and reduce my vision: the detailed view of the colour pattern of an eagle's wings simultaneously deprives me of

the view of its mighty position, circling high in the sky as it prepares to roost on a cliff. A similar pattern of enabling and constraining occurs with regard to our ability to act. For instance, writing with a crown pen, its built-in slower writing pace invites longer contemplation of sentences as one writes them down. With a keyboard, the writing pace is faster, giving rise to a writing language that is closer to colloquial language. A word processor, on the other hand, offers the possibility to engage with the composition of the text, which in itself gives rise to a very particular style of writing. Technologies are thus not neutral tools, but play an active, always ambivalent role in the relationship between humans and the world. In this way, TMT also allows to conceive of intentionality, freedom and agency as being the result of sophisticated connections and interactions between humans with and through their technological artefacts. By means of TMT, a concept like 'intuitive', for example, allows itself to be critically scrutinised as something highly ambivalent. Precisely because artefacts 'withdraw' from our conscious experience when we use them 'intuitively', i.e. we we look and act 'through' artefacts, the non-neutral character of the underlying human-technology-world relationship fades into the background. It is the ambivalent nature of technology, this recurring pattern of foreground/background, amplification/reduction and enabling/constraining that TMT draws attention to.

1.2. The social fabric of technology: Science and Technology Studies

Whereas TMT focuses on technologies-in-use, Science and Technology Studies (STS) focuses on what makes these functioning technologies possible. After all, when we say a technology functions, this 'functioning' is always predicated on an extensive socio-technical network of interwoven material infrastructure and social norms and rules. A car, for instance, 'functions' only thanks to an extensive network of paved roads, filling stations, traffic rules, oil refineries, car repair shops, car manufacturers, insurance schemes, etcetera. In turn, this socio-technical fabric knows an extensive history, in which various interested actors and social groups have competed to push through 'their' definition of technology. It is this societal meso-level of technology that is studied within STS. We present two key STS perspectives, respectively Actor Network Theory which focuses on the networked nature of a functioning technology, and Social Construction of Technology which reveals that what a technology is and what it serves are relative concepts subject to social negotiation and struggle.

The genesis of the bicycle is the paradigmatic example in the **Social Construction of Technology (SCOT)**[2] to show that a technology is nothing but what users consider it to be. In the early days of the bicycle, it was mainly thought of as a 'macho machine'. It was a thing that tough guys would use to show off their skills to girls. Hence, the big difference in size between front and rear wheel was largely welcomed. After all, this made it quite a feat to ride it without falling, which is exactly what one wants from a macho machine. Later, however, people started considering the bicycle more as a means of transport, first mainly for women, who thus gained a new mobility and freedom, at a later stage for everyone. Bicycles with equal wheels then came on the market. For quite a while, there was a 'struggle' between these two models, the very

[2] For an extensive introduction see Pinch and Bijker (2009).

definition of a bicycle being at stake. So through the ages, a bicycle has always been what people wanted it to be and the look of the bicycle has adapted accordingly (e.g. Bijker et al. 1987).

SCOT thus shows how technology is shaped by a variety of social and political negotiations that simultaneously define an object's form, its meaning and the societal problems to which it is a solution. That a technology is a social construct is underlined in the argument that technologies exhibit 'interpretive flexibility': different social groups (engineers, designers, policy-makers, users etc.) have different needs and values and it is through a process of interaction between these groups that a particular form and meaning stabilises. This interpretive flexibility can still later be seen at work, when a technology has already acquired its function but is reconfigured - 'hacked' one might say - by certain users. The telephone, for example, was initially introduced as a tool for transmitting information in the business sphere. However, over time, its usage expanded beyond its intended purpose as individuals embraced it as a means for interpersonal communication, fostering social interaction and sociability. The telephone's initial design and intended use were eventually interpreted and 'misused' in a way that enhanced its role in facilitating personal connections and social engagement. An analogous but much more complex story can be told about the Internet. Originally created to support time sharing on a network of mainframe computers, today it serves entirely different purposes. This shift is not explained by technical reasons but by social ones. The Internet, one could argue, has been hacked and re-hacked by different actors with different and certainly not always beneficial interests who have tried to clothe it with ever new definitions and purposes. In this way, SCOT shows that the purpose of technology is indeterminate, making it impossible to treat the latest stage in a developmental sequence as its telos. Any technology, as a matter of principle, is always open to contestation and redefinition.

However, this principled possibility should immediately be balanced with the realisation that a functioning technology is always embedded in a socio-technical network. **Actor Network Theory (ANT)**[3] insists on this by focusing on the strategies and tactics that key actors employ in bringing together a stable network of people and devices in which a new technology will 'succeed'. ANT examines how networks are established, what connections exist, how they move, how actors are incorporated into a network, how nodes of a network in turn form a whole network and how networks achieve temporal stability. For example, an ANT study on the rise of the electric car would show a network of quite diverse nodes, each of which in itself being a network: car manufacturers, battery manufacturers, environmental policy, energy companies, traffic rules, parking regulations, electricity grid standards, charging terminals etcetera. A riding electric car thus depends on a vast network consisting of rules, devices and people that all have to be aligned, a strategic process in which the interests and concerns of social groups and actors play a crucial role.

ANT assumes that when an actor, regardless of his/her position, is removed from or added to the network, as is the case when a technology is introduced into an organisation or society, the

[3] For a concise introduction see Law (2007) and references therein.

functioning of the entire network is affected. The interdependencies of networks typically show up when things go wrong in a system; conversely, these interconnections are usually hidden when things go right. For instance, the failure of the WIFI connection can be the beginning of a long search through a network that includes, amongst many other things, your home rooter, the network provider you pay via a monthly payment order and that you are now trying to reach in all kinds of 'other' ways as well as the fibre-optic cable that you never realised runs one metre underground down your street where pavement works are now underway. Conversely, the failing WIFI connection, and the more the longer the interruption lasts, exposes the networks that depend on it: your partial work-from-home job, your financial administration and its (un-)paid invoices, your children's social lives etcetera.

Complementing each other, SCOT and ANT thus point to the deep network-like interweaving of interests and concerns that all come into play when we talk about a 'successful', i.e. functioning technology. Speaking of a technology as a means to an end is thus far from neutral.

1.3. The cultural bias of technology: Critical Theory of Technology

Critical Theory of Technology (CTT)[4] shifts attention from the social meso-level of constructivist technology studies to the cultural macro-level. CTT takes for granted that technologies are socially constructed. But while SCOT focuses on uncovering the social groups that influenced the design of a given technology, and ANT focuses on the strategies used by different actors in its deployment, CTT is interested in the broader cultural values and practices that surround it. In other words, CTT's focus is less on specific social groups or the strategies they employ, and more on the cultural resources brought into play in the design process.

Central to CTT is the analytical distinction between two ways of understanding a technology, on the one hand as pure function and on the other as pure meaning. Looking through the lens of 'primary instrumentalisation', we decontextualise a technology and reduce it to its functional essence: a knife serves to cut, a car to move and a navigation app to navigate. Looking through the lens of 'secondary instrumentalisation', we are mindful of the many ways in which a technology has acquired cultural significance. Cars, for instance, have played a decisive role in how we have shaped our living and working habits over the past decades, are vested by many with notions of aesthetics and prestige, and were for a long time a symbol of freedom and independence, and now all these cultural values come to crumble in the face of climate change and the quest for liveable cities. What a car is, is determined not only by its function, but also by what it means in a particular context.

Viewed from the perspective of technology development, **primary instrumentalisation** takes place by decontextualising objects and simplifying them to emphasise those qualities that grant them a function. When my chain comes off my bike and, because I don't want to get my hands dirty, I pull a leaf from a tree to put it back on, this can be considered a 'primary instrumentalisation': on the one hand, I 'abstract' what I intend to do to the function of 'keeping

[4] For an extensive introduction see Feenberg (2002).

my hands clean while manipulating a chain' and, in the same movement, I 'decontextualise' the tree leaf from its natural context identifying in it certain features, pliability and impermeability, for example. The starting point of this basic technical orientation, unique to humans, is imaginative and perceptual and consists, as the example of the tree leaf as protection against greasy hands shows, in the identification of ‘affordances’, i.e. useful properties of things.

Such technical insights appear little social in nature and affordances can be used in very different social contexts. It is in this limited sense of 'technology as a neutral means to an end' that we call it 'primary instrumentalisation'. But we do not have to dig deep to uncover at least a kind of minimal social contingency that governs the selection and application of an affordance, even in its simplest form. In the example cited, we can think, for example, of culturally embedded notions of hygiene and cleanliness and the role hands play in social intercourse that may have informed the identification of a particular affordance of a tree leaf.

It is this cultural embedding of affordances that we call **secondary instrumentalisation** and which becomes increasingly emphatic as we consider more advanced technologies. Consider, for instance, to stay in line with the example given, the increasing use of disposable gloves in the catering sector. Here, in addition to primary instrumentalisation and the affordance of 'protecting hands', an increasing complex of secondary instrumentalisations plays a role. The fact that we consider these originally surgical gloves as an efficient instrument in this context is telling of the cultural recontextualisations the affordance of 'protecting hands' has undergone since the late 19th century, first and foremost within a surgical context and then in the catering sector. In that whole process, a multitude of medical, economic, legal, ethical, technological and societal considerations crept in. One may think of considerations around medical requirements (whereby the first concern was to protect surgeons' hands from chemicals rather than to avoid the risk of infection), economic efficiency (understood as mass production), utilisation efficiency (understood as disposability), choice of materials (as laid down in ISO standards on permeation, penetration and degradation) and finally culturally embedded ideas on hygiene, whereby surgical safeguarding standards have been elevated to a quality label for a catering establishment. In all this, the so-called ‘technical heritage’ of a specific chemical sector also plays a role through the affordances of rubber, latex and nitrile, as do the economic interests and concerns that are hereby at play.

The crux of CTT lies in the premise that primary and secondary instrumentalisation are two sides of the same coin. Looking through the lens of 'primary instrumentalisation', we see a nitrile disposable glove as an efficient tool in handling food products. However, this view is inextricably linked to a host of culturally embedded secondary instrumentalisations as we suggested above. One more example from a design perspective, the creation of a refrigerator, may clarify this further. To make a refrigerator, engineers work with basic components such as electrical circuits and motors, insulation, special-type gases and so on, combining them in complex ways to generate and store cold. Each of these technologies can be broken down into even simpler decontextualised and simplified affordances borrowed from nature. This is the level where primary instrumentalisation predominates, in the form of pure technical understanding. But even though these technical matters have been so thoroughly simplified and extracted from all

contexts, the knowledge of the components is still insufficient to fully determine the design. There remain important questions, such as the size of the fridge, which are settled not on technical grounds but on social principles (e.g. in terms of the likely needs of a standard family). Even family size is not a decisive factor entirely: in countries where daily shopping is done on foot, refrigerators tend to be smaller than in countries where weekly shopping is done by car. At key points, then, the technical design of this artefact is directly linked to the cultural design of society. The refrigerator in our kitchen seamlessly combines both the technical and cultural register.

The choice of 'best' technology design is thus never a purely technical matter: designs are always ambivalent, and it is only through the application of the secondary instrumentalization that the actual form and function of a device is resolved. Technology design is not only a strategic contest between interested actors and social groups, as argued by STS, it is also a function of the way in which things appear to be 'natural', 'efficient' or 'intuitive' to the designer. It is this cultural bias to which CTT draws attention.

2. The social responsibility of technology design

The previous chapter elaborated a vocabulary to reflect and discuss technology as inextricably intertwined with the social and cultural context. Right from its conception in the design lab, technology has been shown to lead social life. Here we want to zoom in on the ethical implications such a view has for technology design. What does it mean for designers to see their own practice and its ensuing outcomes as socially embedded? Where and how do values become built into the design of a technology? In other words, where does the social responsibility of technology design reside?

2.1. Design and intentionality: reflecting on the power position of designers

Compared to the linear model of innovation, the previous chapter on the social embedding of technology marks a shift in the understanding of the 'power position of designers'. Within the linear innovation model the designer, as an in-control agent, purposefully designs an artefact that serves as a neutral means to an identified end. Society is seen here as a relatively separate domain of application, and it is entirely up to that same society to use this technology ‘well’ or ‘badly’. According to this view, a knife is nothing but a tool to cut, where cutting is understood as the affordance that was intentionally designed; it is up to people and society whether this affordance is employed to cut bread or to kill. It is in this sense that R&D actors may be heard saying: "we have merely created a technology". This statement could be noted, for example, from the lips of Prof Van Montagu, one of the founding fathers of genetic modification technology, on the occasion of a public debate (Ghent, 2014). With this, Van Montagu was putting words to the almost schizophrenic position assigned to the designer in the linear innovation model. On the one hand, he is considered all-powerful in the seclusion of the lab, where he intentionally designs a technological solution to a defined problem. But as soon as these solutions live their lives in the world, the designer washes his hands of the matter: it is then up to politics, regulation and education to ensure society conforms in order for his/her invention to reach its full potential (e.g. eradicate hunger in the world) and not to lead to unintended consequences (e.g. monopolisation by big companies).

Against this clearly demarcated power position of the designer in the linear innovation model, the conception of innovation as a socially embedded process yields a deeply ambivalent picture: on the one hand, more power comes into the designer's hands, on the other, less. In what follows, we expound on this ambivalent picture step by step, starting with the idea that a designer does much more than just create an artefact.

The designer has power …

TMT’s understanding of technology mediating the relationship between humans and the world grants designers great power as they are now seen as actively shaping the way humans perceive and act in the world. Thus, **a designer does much more than just create an artefact**, he also designs a user experience. For instance, architecture not only shapes a house or a building or a city, but also helps create the space in which people live their lives and in which certain actions

make more sense than others or become more obvious. Indeed, TMT shows us how, in a user experience, the technology becomes transparent, and how it fades into the background. Once we are habituated to technologies, we stop looking at them and instead look and live through them to the information and activities we use them to facilitate. So the architectural space becomes a backdrop against which we live our lives.

We need not delude ourselves here. If there is one group that has mastered this insight from TMT very well, it is technology companies that seek to maximise their profit margin. Feeding specific content and interaction-possibilities to social media-users can be seen as an example of designing not just an application but a specific user-experience that nudges users into specific – lucrative for the company – behaviour. The introduction of AI systems into peoples' everyday lives poses a particular challenge here. Indeed, the 'predictive performance' of an AI system, whose maximisation can be seen as steering the self-learning journey of such a system, is ultimately measured not in terms of some predictable truth, but in terms of an aggregate goal defined by the designer of the system such as overall profit or safety. **The invisible layer of algorithmic mediation, which structures and influences our thinking and acting, renders users radically susceptible to manipulation**. At an individual level it threatens our autonomy, and at a collective level, it stands to undermine social values, like democracy, which are premised on individuals being capable of independent decision-making. And there is more. Not only the mediating role of technologies can be anticipated, but also the STS-idea that technology only 'works' in a socio-technical network. Developing social media as a centralised platform serves as a good example of creating 'platform dependency' where users are only connected to each other on condition that they submit to certain data-mining activities of that same platform. These examples of manipulative 'nudging' user experience through technology thus show in a negative way the power of design.

Beyond the question of whether anticipating the social dimension of a technology is done for ethically defensible intentions, it is important to realise that such anticipation always carries the risk that things will turn out differently and that, in other words, there will be 'unintended consequences'. History shows that this is the rule rather than the exception. After all, as we argued in the previous chapter (see 1.2), a technology is only what it becomes in the hands of its users and society. In this sense, it is interesting to follow Rommetveit (2021) in **thinking about innovation in terms of handling 'ignorance'**. Contrary to what self-assured R&D discourse would have us believe, the social fate of an innovation involves radical ignorance. In this sense, the effective power of designers immediately raises the question of their responsibility in mobilising and exploiting this ignorance in anticipating new technological futures. In this regard, a troubling paradox can be observed.

On the one hand, this ignorance is fully exploited on the side of R&D and this in the name of cutting-edge developments, whereby the unknown is invariably seen as a terrain to be conquered and mined. Within the context of big data, Thatcher et al. (2016) call this the utopian imaginary of 'digital frontierism', which they challenge by contrasting it with the metaphor of 'digital colonialism'. On the other hand, the very same ignorance is often employed as a licence to innovate and to limit society's grip (policy, ethics,...) on innovation: precisely because

regulators ‘do not yet know for sure’, they cannot and should not yet intervene, the argument goes. The burden of proof falls on the side of those who are critical about the social fate of the technology, whereby issues of regulation and social acceptance of technologies are delegated to a later time, after the ‘real’ work of innovation has been done and the products have been introduced in society. This paradox can also be understood as yet another attempt by innovators, when it suits them, to reproduce the linear innovation model and its separation between innovation and society, while in practice they have long since abandoned that same model.

…but this power is socially and culturally constrained

In tandem with the insight that designers have more power than often thought, the conception of innovation as a socially embedded process also shows that designers have relatively less power, and this on two levels. First, on a meso-level, their power is constrained by the fact that a host of other actors, each with their interests and strategies, play a role in the design process. As we have argued, STS shows that the ultimate function and meaning of a technology is not only decided by the designers. SCOT in particular points out that this is rather the result of a negotiation process between involved actors and interested groups from the immediate design environment, e.g. designers, shareholders, funders, users and other stakeholders. ANT in turn emphasises the socio-technical network within which a technology ‘functions' and in which a technology must therefore be inscribed. Again, the 'success' of a technology depends on the strategic alignment of a host of other actors, which in ANT's view can include both people and things: suppliers, devices, supply networks, network principles, design standards, etc. In summary, STS paints a picture of designers as influential actors engaged in negotiation and conflict with other interested actors.

Second, on a macro-level, CTT points out that a technological innovation practice is at all times culturally embedded, whereby the whole notion of the designer's intentionality may be called into question. The power of the designer must here be thought of as relative, in the sense that in the exercise of their power (as described above) they invoke courses of action that are in turn culturally predetermined in a non-negligible way. It is important to recognise that this cultural bias, as we have called it, makes its appearance in designers’ actions in two ways. On the one hand, they appeal to their imagination to project certain functions into the future, whereby the 'seeing' of affordances is in part shaped by the cultural glasses through which designers look at the future and through which certain futures seem preferable to others. On the other hand, in doing so, designers also appeal to what can be called a discipline’s 'technical heritage', i.e. the 'culturally biased knowledge sedimented in technical disciplines shaped by a history of technical choices.’ (Feng and Feenberg 2009)

Feng and Feenberg (2009) give the example of the bicycle and the roadway and how they are embedded in the social fabric in the Netherlands and North America respectively. 'It does matter,' Feenberg argues, 'that a person living in Amsterdam is inclined to think of cyclists as natural users of roadways, while a person living in America does not. It matters, because this taken-for-granted understanding - what in essence is culture - becomes a background condition to the design of technology.' And he further points to the dominant meaning attached to a

roadway in both cases: in Holland, it is accepted that bicycles and bicyclists are 'legitimate' users of the road (indeed, cyclists commonly have the right-of-way); in North America, these same devices and people are oddities, either grudgingly accepted or met with hostility by the road's primary users, motorists. In North America, the word "road" brings to mind cars; in Holland, the same word brings to mind both cars and bicycles. The Dutch road not only incorporates bike lanes, but just as importantly, social expectations about the proper use of bicycles. It is in this sense that the ultimate social function and meaning of a technology must also be understood 'as a function of the way in which things appear to be "natural" to the designer'. A comparison between Dutch and American roadways tells us that designers must accommodate themselves to existing social worlds, which implies submitting to existing power relations and hierarchies. The stifling effect of such passive coercion is a significant obstacle to the realisation of alternative designs.

2.2. Technology design as culturally embedded abstraction: the ethical charge of concepts

In this section, we focus on the role that concepts play in design practice as we are convinced that, as carriers of cultural charge, they can provide concrete entry points for critical ethical deliberation and proactive alignment with social and cultural values.

The technological design process understood as a process of abstraction

To explore this in more detail, it pays to think of the design process, in line with CTT's take on primary instrumentalisation, as a process of abstraction. In the design phase, an engineer deliberately breaks the cohesiveness of the everyday lifeworld to study, within the contours of the laboratory, a make-shift artefact in detail, thereby decontextualising objects and simplifying them to highlight those qualities by which they are assigned a function. It is characteristic of a technological design process that a social need or challenge is delineated for which the artefact to be designed should provide a solution. One sees this very clearly reflected in vision texts and project descriptions of R&D labs where, typically, a societal need or challenge is posited, how this can be addressed technologically and the kind of R&D that will be needed to produce the necessary innovation. Cancer biomarker research, to take one example, aims to contribute to 'precision oncology', including 'early detection' and 'personalised treatment', by developing 'predictive biomarkers' that will be developed on the basis of 'innovative sequencing techniques'. In a similar logic, carbon capture R&D projects are framed as addressing the societal need of 'reducing greenhouse gas emissions' and 'combating climate change' through the development of 'innovative chemical technologies' that promise the re-utilization or safe storage of emitted carbon emissions. And in a similar manner, location-based services are seen as "smart solutions" to address a plethora of identified societal needs, such as addressing traffic congestion, cultivating and enabling more efficient and environmentally friendly mobility patterns, encouraging healthier lifestyles, etcetera. In essence, this is a matter of abstraction each time, which happens by delineating a societal need and envisioning a technological solution for it. It is important to recognise that this abstraction is an unavoidable step that makes technological design possible at all. Indeed, there is nothing extraneous about this. As

already hinted at, the ability to abstract in order to act upon it in the world is something that characterises us as human beings and makes us the technological being we are in essence and origin. Anthropologists indeed conjecture that the ability to think of objects as means, that is the capacity to abstract, together with the upright stance and opposable thumb together form a constellation that predisposes human beings to engage technically with the environment. In this, humans achieve an exorbitant development of potentials exhibited in small ways by other higher mammals.

At the same time, it is crucial to acknowledge that this abstraction implies a decisive, non-neutral act, whereby the socio-technical complexity of the world is deliberately delineated, solidified, interpreted and consequently staged in the design lab in terms of a well-defined problem for which a technological solution is designed. In this sense, the linear innovation model contains a grain of truth: the designer does withdraw from society to study a well-defined aspect in detail. Only, in 'withdrawing from society' through abstraction, something happens that is anything but socially and culturally neutral. It is not because the designer withdraws into the laboratory with his quest that s/he would therefore act 'in isolation' from social and cultural interests and concerns. This is what CTT claims with 1st and 2nd instrumentalisation only being distinguishable on an analytical level. As we have illustrated with the example of the tree leaf used to put a bicycle chain back on, there is no 1st instrumentalisation without a minimum of 2nd instrumentalisation. One cannot design a function without assigning minimal meaning to it by delineating what this function should serve. And it is through this minimal meaning that cultural bias, as discussed in 1.3., inevitably makes its appearance in technological design and this in two ways: on the one hand, through the imagination of possible futures, on the other, through the mobilisation of the technical heritage of the design discipline. We elaborate this in more detail in the next section.

Design as abstraction and the ethical charge of concepts

If, to return to the same example, an R&D group sets out to pursue 'precision oncology' by developing 'biomarkers' through the technique of 'direct episequencing', this involves an abstraction from a socially embedded issue to the contours of the epigenetics lab. As argued, this abstraction is anything but neutral and knows two dimensions. On the one hand, a prospective one, where the future is brought into the lab through the culturally embedded vision of 'precision oncology'. So the term 'precision oncology' carries the ethically charged promise of providing the right treatment to the right patient at the right time, itself inspired by a distinctly Western view on life and death, of cancer as an invasive and unfair disease to be defeated at all costs, and of technology as holding the ultimate weapon to win the war on cancer (Engen 2022). In this sense, precision oncology can be understood as what Jasanoff (2015) calls a **'socio-technical imaginary'**, which she defines as ‘a collectively held, institutionally stabilised and publicly implemented vision of desirable futures, animated by shared views on forms of social life and social order, attainable through and supportive of advances in science and technology'. Defined as collectively held, socio-technical imaginaries can be national or transnationally borne, as is the case with 'circular economy', 'zero carbon society', or 'digital sovereignty', but equally at the level of a collective such as that of a design-lab boasting to produce ‘smart solutions’ based on location based services. Besides the prospective dimension,

design as abstraction entails a historical dimension in which history is brought into the laboratory through the culturally embedded technical heritage of the design discipline. Typically, this technical heritage is encrypted in the **metaphors** technical disciplines mobilise. In the example given above, 'biomarker' and 'sequencing' are two such metaphors that are central to biomarker research, carrying with them a culturally charged technical, material and social history.

By focusing on the imaginaries and metaphors that a given design practice mobilises, we explicitly draw attention to the crucial role concepts play in the abstraction process inherent to technological design. More specifically, we draw attention to the 'cultural charge' that imaginaries and metaphors, as conceptual abstractions, carry. In this context, it is interesting to dwell for a moment on the very different way in which two of the project's core concepts have been dealt with in SIMPORT, notably sovereign and intuitive. Although not explicitly cast as a 'socio-technical imaginary', the project proposal deals extensively with 'digital sovereignty' as a vision of a desirable future that unfolds around a continuing digitalisation within Europe. In doing so, the proposal points out that this imaginary, although serving as an overarching goal, cannot simply be taken for granted. One of the five research questions explicitly states: "what does digital sovereignty mean in dealing with personal location information?" In response, one of the six main work packages is entirely dedicated to "ethics and digital sovereignty in the management of personal location information." Throughout the project, attention was thereby repeatedly brought to the collective aspect of digital sovereignty that is at risk of being snowed under in a straightforward understanding of sovereignty as individually deciding what location data one is willing to share with third parties. In a context where AI is becoming ubiquitous, many business models do not focus on the personal data of individuals but on high-resolution yet anonymous mass data. Conceived as a collective concern, digital sovereignty should thus take into account that, in the age of predictive analytics, an individual or a group can be treated unequally on the basis of anonymous data that others disclose about themselves, possibly in best faith (Mühlhoff 2021).

In comparison with this reflexive approach to the concept of 'digital sovereignty', it is striking that the concept of 'intuitive', although equally central to the project, remains untouched in the project proposal - as if it were a neutral, purely technical term, the meaning of which is considered obvious. Besides being in the title, the term figures centrally in research question 4: "Which forms of interaction are suitable for intuitively enabling users' decisions and wishes regarding the handling of their location information?" Other than that, the project proposal text leaves this term undiscussed which indicates to what extent the term 'intuitive' is part of the technical heritage of the Human Computer Interaction domain. This was also confirmed in the interviews with the project collaborators (see Part 2) where the term 'intuitive' was regularly used by developers to casually indicate what they thought the so-called 'fine grained control options' that the project intended to develop should look like. In the course of the project, critical questions with regard to 'intuition' were raised by the ethicists, but it was found that, unlike the critical aspects around the collective character of digital sovereignty, these hardly took hold. The following quote by one of the ethicists (further discussed in 3.1) nicely captures the ethical charge that is deeply embedded in established design concepts like 'intuitive':

> “While we had this discussion around collective aspects of privacy and this was something that stuck with people, like the individual level being not enough, I would say that around intuitivity, I didn't get to people, like I'm still having meetings and people go like, oh, we should do it intuitive and that's so good. And I'm always like, no people, I told you it's not good. You can't use it that way … What is intuitive? What should it even mean? It usually means something like easy but easy for whom? … I think it's one of those myths that people tell … that they just believe and having this idea of there's one intuitive way to do it, and then it's easy for everybody. It's a really tough idea to break.”

2.3. Responsibility in the face of ambivalence

In the previous chapters, we have immersed ourselves in the social nature of technology, asking what this means for the power position of the designer. What emerged from this was an ambivalent picture. Yes, the designer has power, a great deal of power in fact. Through our interweaving with digital devices, for example, the design of software algorithms give shape to what we think, feel and do. But at the same time, the power of a designer is not that of the modernist agent from the linear innovation model who develops a 'perfect’ technological solution from a kind of omniscient point of view. Against that simplistic model, we sketched innovation as a complex and messy process that is thoroughly interwoven with society and in which the designer is one actor amidst a host of others, each with their interests and concerns. In this picture, the power position assigned to the designer is thoroughly ambivalent, mainly for two reasons.

First, once technologies are taken up and used, they mediate our perception and our ability to act according to an ambivalent pattern of amplification and reduction. New possibilities thus always come at the expense of new impossibilities; certain groups will manage to bend a technology to their will, others will suffer the consequences. Data-fuelled AI for instance carries the hope that it can help uncover patterns in human practice in order to make social processes ‘more sustainable’, ‘more efficient’, and ‘more reasonable’. At the same time, we increasingly see the problems thrown up by AI applications: it is used for surveillance, on its basis new forms of capitalism become reality which are much more closely interwoven with everyday practices, and it reproduces and reinforces inequalities and everyday racisms, to name just a few of the most important points of discussion (Benjamin 2019, Eubanks 2018; Noble 2018; Zuboff 2018; Buolamwini & Gebru 2018, O'Neil 2016). Even before anything can be said about how to deal with it, this irreducible ambivalence calls on designers to observe a certain degree of humility in speaking and thinking about their own power (Jasanoff 2003).

Second, it became clear how the power of the designer is always socially and culturally constituted. Every technological affordance, from the moment it is even conceived, already carries within it a cultural bias. For example, we drew to mind how a central design term like 'intuitive' is not innocuous and carries within it the legacy of several decades of interface design, which was done primarily from a profit perspective. In this respect too, the power of the designer is irreducibly ambivalent, by which we mean that there is no such thing as a 'neutral' understanding of what constitutes an intuitive interface design. And the same goes for other

principles that ethics guidelines typically put forward. What fairness, privacy, social benefit, etc. should mean in a concrete application entails an irreducible ambivalence. Clinging to the belief that there is a 'neutral' interpretation amounts in practice to submitting to the ruling order: it seems 'neutral' because it has achieved the status of the dominant ideology, which always strikes us as self-evident. In this sense, such principles should much rather be understood as starting points for a reflexive deliberative process. Instead of giving neutrally formulated definitions, they would therefore be better accompanied by a warning such as "beware, fairness may not be what you think it is," or "in the end, your design decisions will give shape to what human autonomy will be". That is the inescapable responsibility of technological design: it takes sides anyway.

The irreducible ambivalence that technology design faces, both in terms of the outcome and intention of its practice, is aptly summarised in the following quote, to be taken as a persistent provocation, rather than a conclusion[5]:

> *"If you think technology can solve your problems, you don't understand technology - and you don't understand your problems."*

And so the question arises of how to actively shape a responsible attitude in the face of irreducible ambivalence. We already cited humility and reflexivity as two important pointers. Beyond that, it is of utmost importance to recognise that an attitude of responsibility is a context-dependent matter. Innovation is not a process that can be steered from one omniscient point. If you want to take responsibility as a designer, it has to be done bottom-up, from within the design practice that is thoroughly intertwined with the social context. In any given instance, designers make and remake numerous, overlapping decisions, each of which is situated within a broad range of contextual dimensions that imply a diverse array of potential social, cultural, ethical, environmental and economic decision input or influences (Fisher 2022).
In order to explore value-sensitive responses to the ambivalence it faces, an ethical design approach will thus have to orient itself in the social fabric of its own practice. In part 2 of this report, we explore such a self-reflexive approach in more detail by focusing on the perspective of innovation practitioners. On the basis of interviews with SIMPORT collaborators, we try to gain insight into how they map out their responsibility, how they make sense of this complex and messy process of innovation and what (im)possibilities they see to make ethics matter in their daily practice. This seems to us a more fruitful approach to arrive at meaningful recommendations than simply laying down ethical principles.

[5] Credited to a variety of people, this quote figured in a 2023 art installation by Laurie Anderson at Moderna Museet (Stockholm).

Part 2: A practitioners' perspective on ethics by design

Part 2 presents an insider perspective on the concrete practice of integrating social and ethical considerations in software development. It does so by drawing on in-depth interviews with SIOMPORT-collaborators about their experience with the Ethics By Design (EBD) approach pursued throughout the project. The underlying premise is that EBD's actual potential crucially depends on whether and how practitioners can make sense of this buzzword and feel empowered to translate it into practice. Chapter 3 presents a narrative analysis of the interviews. We show how SIMPORT collaborators, developers and ethicists alike, in talking about their experience with EBD, draw on three narrative repertoires, respectively on ethics, software development and integration. In chapter 4, we turn this logic inside out and argue that these repertoires, and the core narratives they contain, shape the 'geography of responsibility' by which practitioners assume and ascribe responsibilities vis-à-vis the ethical dimension of their work. As we will argue in Part 3, we consider this 'geography of responsibility' key to unlocking the potential of EBD.

3. Making sense of ethics by design: a narrative approach

No matter how often and how loud projects proclaim to follow an EBD approach, there is no clear-cut methodology, no blue-print of how to put this in practice. Bringing this into focus, we find it instructive to consider 'ethics by design' as a buzzword[6]. A buzzword gains power precisely because it represents a positive value while remaining somewhat vaguely defined. Who could be opposed to design that guarantees an ethical outcome? While there are good reasons for scepticism, Part 1 also showed that there are good reasons to take EBD's promise of integrating ethical and social considerations seriously. Therefore, apart from the promises that such buzzwords advance and the ideals and principles they claim to pursue, we propose taking a closer look at the practice and seeing what concretely happens under the EBD label. It is for this reason that this chapter will give SIMPORT practitioners ample voice.

This way, by focusing on the practitioner's perspective and discussing EBD with them, we are convinced to learn more about the possibilities and constraints to make ethics a core value in software development. However, the practitioners' perspective is not sacrosanct. Indeed, there are several reasons why we should not accept what they say at face value. To begin with, they may not see the relevance of EBD, even though it is there. In that case, the key is to bring it to their awareness by, for instance, expanding their view of their own practice and showing them that an ethical perspective is indeed relevant. It may also be that they do see the relevance, but do not feel responsible for it and prefer to outsource ethics to others or to a later stage in technology design. Here the question arises about the 'ethos' of both the practitioner and the organisational context in which they are embedded. And finally, practitioners may see the relevance, but feel unable to act on it, due to contextual or institutional factors. 'I don't get paid for that' is an oft-heard quote here, pointing, for instance, to valuation and reward schemes in place within a particular culture (Sigl 2020).

In all this, we will follow a narrative approach, which in short assumes that agency, the ability to act, appears in and through narrative. The possibilities one sees for action are always enabled and constrained through a narrative about how one understands oneself and one's relationship to the social and material context. It is from this perspective that we want to hear from SIMPORT practitioners whether and how they give meaning to EBD, and which narratives they thereby mobilise about themselves and their relationship with others and the world. To this end we conducted semi-structured interviews towards the end of the project. Besides more general questions related to the theme of the project, i.e. digital sovereignty in relation to location based services, the interviews focused on how practitioners had experienced the EBD trajectory pursued within the project, how they assessed its impact, which aspects they liked or disliked, how they related to it in their own practice and what they thought could or should have been done differently. As a result, two types of information emerged. On the one hand, evaluatively, interviewees make sense of 'what EBD has been'. They talk about the activities that took place under the EBD label, express their appreciation, criticise or amend them. On the other hand,

[6] The Merriam-Webster dictionary defines it as "important-sounding usually technical words or phrases often of little meaning used chiefly to impress laymen."

imaginatively, interviewees make sense of 'what EBD should be'. This tells us something about what place EBD occupies in the interviewee's imagination. In practice, the two types of information are often intertwined. For instance, an actual EBD practice is often talked about 'evaluatively' by expressing where it fell short against the 'imaginary' idea the interviewee has of EBD.

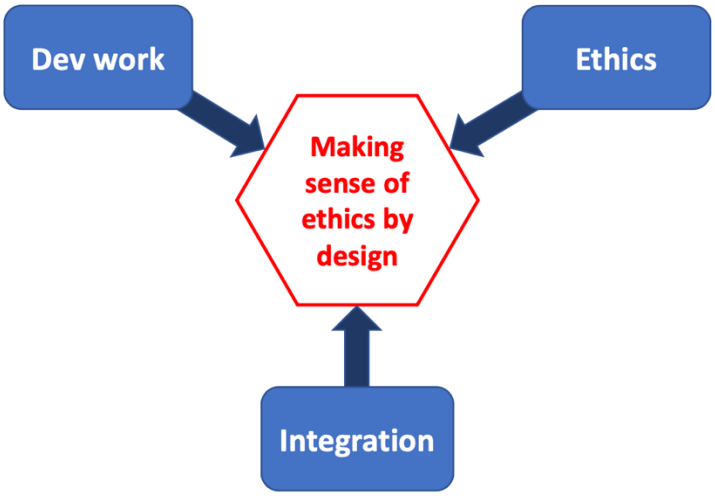

Figure 1 Narrative repertoires drawn on by SIMPORT practitioners to make sense of ethics by design.

What is in a way striking about these interviews is that they do not contain unequivocal answers as to what EBD has been, or what it should be. At the same time, it is self-evident: just because EBD is a buzzword, everyone, developers and ethicists alike within the team, is busy making 'sense' of it. Anecdotes are told, accounts of what people thought of the 'ethics worksheets', and what they did with them as well as how they perceived the presence of the 'embedded ethicist' in the weekly development meetings. Ideas are expressed about how things should have gone, about how things could have gone as well as expectations people had that may or may not have been met. The interesting thing about the interviews is that one can hear the 'sensemaking' so to speak live at work. Interviewees hesitate, doubt, recant their words, they wrestle with what they at times refer to as 'the ethics thing'. In analysing the interviews from a narrative approach, it became clear how this mass of information could be clustered around three 'narrative repertoires', respectively on 'software development work', 'ethics', and 'integration of ethics and software development'. As schematically depicted in Figure 1, SIMPORT practitioners, in discussing EBD over interview, are understood to be drawing on the discursive reference points contained in these respective repertoires.

A first narrative repertoire interviewees employ includes ideas and statements around 'software development work': how does an interviewee understand oneself or another as a 'developer', and what does one see as a developer's task and domain? Then there is the repertoire around 'ethics': how does an actor understand oneself or another as an 'ethicist', what does one see as the ethicist's task and domain? And finally there is the repertoire around 'integration': in what terms is the integration of ethics and development work thought and spoken of? It may be clear

that the repertoires around 'dev work' and 'ethics' already have a long history, and will therefore contain some culturally embedded ideas. In contrast, the repertoire around interdisciplinarity and socio-technical integration is relatively young, although some jargon has crept in within the development world via the notion of 'privacy by design'. In what follows, we describe these repertoires in more detail through the core narratives they contain, which we identified in the interviews. Such a 'core narrative' contains a cluster of 'narrative elements' or 'discursive reference points' that all fit within an overarching idea, for example, as we will see immediately, that developers are narrowly focused on functionality. In the identifiers following each quote, ethics team members are represented as (ETH) and software developers as (DEV). We have further numbered each participant in order they appear.

3.1. Making sense of development work

A first narrative repertoire used revolves around development work. A general aspect that often recurs is that developers are a specific kind of people who are very focused on their core business, or as one project leader put it: "they just want to develop" (DEV-1).

Narrow-mindedness

An often-cited characteristic of developers is that they are very strongly and rather narrowly focused on functionality, or as one project leader put it, they are "people who are supposed to solve a problem, create a solution, on a technical level" (DEV-2). In that context, the term "narrow-mindedness" was used to refer to a natural quality that developers are supposed to possess, and which does not need to be excused:

> "Developers, or software architects, and also business developers with a digital focus, they have a natural focus on bringing out some successful application, and for days they are working on it and their everyday focus is on issues like usability, user experience, validity of architectures and so on (…) It needs no excuses, it is obvious that people who are so much specialised … get some kind of narrow-mindedness, naturally somehow" (DEV-2).

This narrow-mindedness can be clearly related to what we said above about 'design as abstraction', where a designer deliberately breaks the cohesiveness of the everyday lifeworld to study a make-shift artefact in detail to highlight those qualities by which they are assigned a function (see 2.2.). As we have argued, concepts often play an important role in that abstraction process and this was also confirmed by one of the ethicists who talked about the concept of 'intuitive' acting as a kind of 'guiding myth' in the development process:

> "And you come into this kind of environment of developers and you have people who believe in the one way in the one-size-fits-all … I think it's one of those myths that people tell, that they just believe. Like, having this idea of there's one intuitive way to do it, and then it's easy for everybody. It's a really tough idea to break" (ETH-1).

A compartmentalised, agile workflow

A second aspect that often recurred revolved around the idea that software development follows a typical workflow in which the work is compartmentalised into subtasks that are carried out by a team where everyone has a well-defined role. Coordinating and streamlining these

subtasks is done through regular development meetings ('dailies', 'weeklies', etc.). In this context, regular reference was made to terms from agile software development such as 'scrum' meetings led by a 'scrum master', as is the case in the following quote:

> "consider a development team … that, I don't know, creates a software in an agile process for a company, or a larger project, where you do sprints, and everybody has their role as developers, designers, the scrum master, project managers, product owners" (DEV-2).

Breaking up the work into subtasks is thereby seen as something very typical of software development, as expressed by one of the developers:

> "It's just like, nice to break down into goals and sort of do one thing after another … software development processes are sort of made for this sort of thing (DEV-3)"

That built-in way of compartmentalised working was also cited several times as a reason why Covid had relatively little impact on this project. Although it was launched in full pandemic, with everyone working remotely from home, there is a strong feeling that the project got off to a good start. In this context, great importance is attached to the weekly 'scrums' or 'dev meetings', as well as the monthly 'stand-ups' where the progress of different components is briefly discussed. This relatively standardised workflow seems to give developers a foothold, regardless of whether this happens live or remotely.

A guidelines culture

What also recurred frequently was the idea that software development commonly uses guidelines, cheat sheets or some other 'instruction manual-like' form of guidance. One developer called it "a full written documentation on how to do some things and maybe also with some kind of examples" and added "yeah, that always helps a lot. (DEV-4)" A project leader put it as follows:

> "I think developers in particular are very keen on this because they know the interface guidelines and the programming guidelines and they read it and then they implement it and it's sort of fairly straightforward" (DEV-2).

Interestingly, one of the ethicists linked the omnipresence of guidelines to an individualist learning culture where self-education is in high esteem, calling it "a really big approach of like … you can just learn that yourself (ETH-1)." It was immediately added that this was not the case in the SIMPORT team:

> "People work together and they are in the GitHub together … and I think all the dev meetings show that people really help each other out and are very open about that, like answering questions and posing questions. And I think that's a very, very nice culture that we have in the team (ETH-1)."

The good collaborative culture should be somewhat nuanced by the often echoed observation that, as a primarily academic research project, SIMPORT was a kind of sandbox project where, unlike commercial environments, there was space and time for collaboration and exploring possibilities. Consider the following quote, which also refers back to the core narrative of narrow-mindedness::

> "I think we had it easy by having relatively few sort of other goals, right? I mean - that's maybe too easy - but I could imagine in real life, I mean in other contexts, there's sort of competing factors that, don't necessarily go against the ethics thing, but just make it harder to do it right. Like I think the most simple one is just time, right? It costs time to do all these things. And if you just for whatever reason need to do a thing quick, yeah, I think it comes natural to people to drop these thoughts.(DEV-3)"

Ethics as a context switch

The last quote contains another narrative that often recurs when interviewees try to convey what development work entails, namely that it is something utterly different from ethics. One of the developers put it as follows:

> "I think it's also that just in the day-to-day work, it's sort of hard to keep track of these things and just easier to focus on technical things, you know, it just comes natural and then you need to sort of continuously check: hey, these technical decisions I'm doing right now, do they also have an impact that is not only technical? Which, when you're constantly doing purely technical work, is difficult because in a way it's also a context switch. (DEV-3)"

At this point, the open spirit of collaboration within the team was somewhat nuanced by one of the ethicists in the sense that:

> "as an ethicist, you have this kind of runway on which you can go, which is larger with developers who are more open about these kind of issues. But once you hit like certain core problems, you might see, ok, so this is the point where they start to block. (ETH-2)"

3.2. Making sense of ethics

Although the project proposal repeatedly refers to philosophy and critical social theory on the one hand and Responsible Research and Innovation and 'midstream modulation' on the other, it became clear from the interviews that 'ethics' and 'ethics by design' were the guiding terms used to interpret the specific SSH input to the project. Thus, the developers refer to 'the ethics people', 'ethics expert' and conversely, the SSH participants refer to themselves as the 'ethicists' and framed their contributions in terms of 'ethics', i.e. 'ethics workshops' and 'ethics worksheets'. It is a point we will return to, as we are convinced that this specific 'ethics' framing has had a decisive impact on the expectations associated with it. In line with this, it was noted in the interviews how 'ethics' was used as an umbrella term for both the developed EBD process and its content. Thereby, reference was regularly made to 'the ethics thing', which tellingly highlights an underlying tension : on the one hand, ethics is often perceived as vague, as that indefinite 'thing' that especially eludes developers; on the other hand, the designation 'thing' suggests that people would like it to be more concrete, more tangible.

Ethics as knowledge: epistemic authority versus non-committal chatter

A first cluster of narrative elements revolves around the assumption that ethics involves a reservoir of knowledge that possesses epistemic authority. For instance, one of the developers described the role of the ethicist as "somebody who knows about ethics and can put the finger in the wound and say, okay, this might not be sound," only to lament moments later that "ethics

are hard to measure (DEV-5)." The assumption that ethics is a scientific field that can make unequivocal statements about right or wrong is primarily brought into play by developers in a negative way to indicate their disenchantment with the presumed non-committal nature of ethics. This is for instance echoed in a developer's allusion to the idea that a philosopher is "often considered as a chatter or as a blabber, somebody who is talking a lot, but without any relevance (DEV-2)". Also telling is the following quote about an 'ethics paper' written by the ethicists and circulated within the project team:

> "It [the ethics paper] contained general concepts and maybe some questions of a general nature, but nothing to really go by. What is the very point? What does it tell us? And what could be the impact on our doing here? (DEV-2)"

This 'non-committal' character was particularly raised when talking about the ethics workshops held mainly in a first phase of the project, a phase described by one of the developers as "an ethics intense time. (DEV-4)" While the workshops were generally labelled "interesting" (DEV3-4-5-6) and of importance "to discuss the basics" (DEV-5) and "to have everyone on board" (DEV4-5), this was almost always nuanced from a developers' angle as non-binding and hard to apply. Consider the following quote:

> "We had a number of workshops in the beginning, where we were discussing the basics, and kind of tried to get onto a similar level, at least when it comes to terms and understanding of concepts and everything. But most of us are not philosophers, they have a technical background. So that was, I would say, good to know. But then it was a bit hard to turn this into a working knowledge that we could apply. (DEV-5)"

This criticism of the 'non-committal' nature of ethics stuck to the ethicists who, in turn, tried to make sense of it. At one point, they decided to create 'ethics worksheets', a kind of one-pager in which an ethical issue is unpacked in a concise manner. Commenting on a specific worksheet s/he had been working on, an ethicist alludes to a sense of despair in trying to position ethics against this ever-returning criticism of being 'non-committal'.

> "What to say, you know, what problems should be discussed, what societal issues, what technological, theoretical, ethical issues should be discussed? For me it was very difficult to pinpoint what we as ethicists want to say in that regard. … At least in my experience that wasn't clear, like having very defined points on what we as ethicists are trying to say in a sense of, you know, maybe call them academic points (ETH-2)".

It would be all too easy to interpret this quote in terms of incompetence on the part of the ethicist. In our reading, it mainly says something about the gap that exists between the two domains and which, seen from the developers' side, should be bridged by making ethics 'applicable' as an epistemic authority. (see 3.3 for an elaborate discussion).
The characterisation of ethics as 'non-committal' persisted right up to the end of the project and shaped the relationship between developers and ethicists to a considerable extent. For instance, in preparation for one of the last project meetings, it was suggested within the ethics team to already present part of the draft version of the 'guidelines document', to which one of the ethicists remarked:

> "It makes us so naked. And it will invite yet another round of complaints that all we do is vague and abstract. (ETH-3)"

Ethics as guidance: facilitating a deliberative process versus settling ethical decisions

This brings us to a second cluster of narrative elements which, rather than the status of ethics as a source of knowledge, concerns the practical role ethics is supposed to play. A commonly used term in that context is 'guidance': ethics is supposed to provide guidance in making development decisions. However, what that guidance is supposed to consist of varies between two extremes. At one end, ethics' role is seen as taking a restrained position as a facilitator of a deliberative process and at the other end as the body that settles ethical decisions on the basis of clearcut principles. In this context, it is interesting to see how the main deliverable of the ethics work package, which you are hereby looking at, was spoken about within the SIMPORT project. Where the project proposal description talks about the design and evaluation of an EBD approach that was to culminate in "guidelines and process description of Ethics by Design", over the course of the project this evolved into the term 'ethics guidelines'. One of the project leaders explains how the emergence of this term relates to different conceptualisations and expectations:

> "My impression is that there were different conceptualizations of what it would be. So I think at some point the term 'ethics guidelines' was more frequently used. So because I think the project was already running almost a year and there was no guidance in the ethical sense. Just general concepts and maybe some questions of a general nature. But nothing really to go by. (DEV-7)"

'Ethics guidelines' in this way functioned as a sort of 'boundary object' between ethicists and developers (Star and Griesemer 1989). While it enabled communication and collaboration between both worlds, the notion of 'ethics guidelines' also contained virtually differing conceptualisations. From the developers' side, it was viewed from their experience with programming guidelines (see above) as a concrete tool for settling development decisions, while from the ethicists' side, the term left enough room to manoeuvre towards a more deliberative approach. The latter is illustrated in the following quote:

> "We wanted it to be a reflectionary process in which the engineers themselves have to reflect on their ideas and issues. We did not see ourselves as the ethical authority that brings the ethics from the outside into the project, but that kind of facilitates ethical thinking within the group. (ETH-1)"

Developers' appreciation of this deliberative approach, which was emphatically explored in the first phase of the project with the ethics workshops and worksheets, is ambivalent. On a positive note, one developer spoke in this context of ethics as "an awareness, a general way of thinking about how to do things ... that you carry with you in your daily work … and if there were decisions that affected them more strongly, you would be aware of that (DEV-3)." Others were more sceptical, such as one developer talking about one of the ethics workshops as "a meeting where we consider philosopher A and philosopher B and we have a 2-hour discussion with arguments back and forth and that's it" and concluding that "that's not realistic. (DEV-5)" Consider also the following quote, which illustrates the ambivalent stance:

> "I mean, broad conversations about ethics are nice and can be entertaining and interesting. But when it comes to designing and developing things, I guess that you would need to look at the time you're investing in this area. (DEV-5)"

From the ethicists' corner, in turn, it was repeatedly pointed out that ethical guidance can truly take effect only when developers adopt a little bit of the humanities' thought style:

> "the 'thinking style of the humanities seems to be completely alien to some of the developers. … The thought style of the humanities is inherently personal in the sense that it cannot be mechanised and proceduralised in the same way as development work. Many people today lack an understanding of this; being alien to the humanities is a societal problem at large scale, it is the signature of our time, if you will. (ETH-3)"

3.3. Making sense of integration

A third narrative repertoire relates to the question of how to integrate ethics and development work. As we elaborate in more detail in 5.1, thinking about interdisciplinarity and in particular about integrating ethical considerations on software development is relatively new. Unlike for software development and ethics, there are hardly any 'entrenched' narratives and it should therefore come as no surprise that, in talking about integration, people are more likely to hesitate, they will try out new formulations or they will borrow from domains considered similar or related, such as 'privacy by design'. Words are experimented with, metaphors are sought, new imaginaries take shape, all in an attempt to make sense of what largely remains to be made meaningful.

Bridging the gap between the two cultures

A first cluster of narrative elements revolves around the idea that there is a gap between software development and SSH that needs to be bridged. As we have seen, both sides repeatedly indicate the extent to which 'their' work differs from that of 'the other', and how 'different' their mindset is. One developer sees "a clear gap," adding that "how to bridge over that gap, that is an interesting question. (DEV-2)" This sense of a gap to be bridged is nicely expressed in the following quote from one of the ethicists::

> "being an ethicist in such a project is defending in two directions: translate to the humanities that developers actually think about their work and explain to developers that not all philosophy is to be done by books, and that it is more practical. (ETH-1)"

The gap discussed here obviously resonates with what CP Snow said in his influential 1959 'Rede Lecture' at Cambridge's Senate House about Western society being characterised by a divide between the two cultures of humanities and sciences (Snow 1959). Snow believed that this divide was harmful to society because it led to a lack of understanding and communication between these two groups. Scientists and humanists, according to Snow, spoke different languages and had different ways of thinking, which made it difficult for them to work together and address society's complex problems.

Talking the same language

In this respect, it is striking that interviewees repeatedly mention the importance of speaking the same language to bridge that supposed gap. Consider the following quote from an ethicist:

> "I think that's a very nice culture we have in the team. So I always felt like I could ask all my stupid questions. Yeah, the gap is not too wide and I think that's something that works. People do not have similar backgrounds, but from the experience and stuff they have a common language to talk about things. So it works." (ETH-1)

It is suggested here that this common language cannot just be installed, but rather emerges from the experience of working together where concrete stuff needs to be talked about. Thereby, it is noteworthy that several interviewees indicated that due to COVID circumstances, where it took more than a year for people to meet in person, "it took a long time for people to speak the same language (DEV-7)." Whereas speaking the same language is clearly considered a necessary condition for interdisciplinary integration, it is therefore not a sufficient condition. This was especially echoed when talking about the 'ethics workshops' at the beginning of the project. Whereas they were deemed "interesting" (DEV3-4-5-6) and relevant "to get everybody on board" (DEV-4-5), "to get on a similar level" (DEV-5) or "to discuss the basics" (DEV-5), the resulting 'common language' was found "a bit hard to turn this [common language] into a working knowledge that we could apply (DEV-5)". All this resonates strongly with what has been written in the STS-literature on interdisciplinary collaboration and the emergence of a so-called 'trading zone' through the mutual sensitisation of involved disciplines. The concept of trading zones was introduced by the historian of science Peter Galison (1997) to describe how two communities with vastly different practices and discourses can interact and negotiate a joint enterprise through the formation of inter-language as a language between languages (Collins et al. 2007).

Seamless integration

As a kind of ideal image of what that bridged gap might look like, the term 'seamless' has been repeatedly invoked to indicate how developers and ethicists work synergistically together towards a common goal. One suggestion made repeatedly in this regard is that the ethical component should be built into the typical compartmentalised workflow of developers (see above). Consider the following quote:

> "If such things, such guidelines, however they might look, could seamlessly slip into a process that is already known to developers, that would be best case I think. So, just to have an extra step in agile development for example – just to name something – would in my opinion have a much bigger impact than simply some guidelines that can be found somewhere on the Web." (DEV-6)

The idea of a renewed 'ethically aware' development process came up repeatedly, often referring to agile and scrum methodologies. While bringing this up, interviewees often implicitly – and sometimes explicitly – indicated the extent to which they felt ethics may 'disrupt' the development process. The use of the term 'seamlessly slip into' in the last quote is telling in this respect. Some are quite happy with "an ethical expert looking over your shoulder and making you stop and think about certain things (DEV-7)", others feel that ethicists should "catch them

[developers] in the process, with empathy and be as constructive as possible" because as developers, we "don't want anyone to stop us (DEV-1)."

That seamlessness was being called for is also related to the empirical way the EBD track took shape within SIMPORT. In an attempt to bridge the gap towards developers, and in response to the criticism of ethics being non-commital (see above: ethics as non-committal), the ethicists pursued what they have called a "trial and error approach" (ETH-1), at one point introducing, for example, ethics worksheets (see above). This empirical approach in turn seems to have led some developers to the perception that there was no streamlined baseline approach and that the ethics input came to them disparately and in scattered batches. Consider the following quote from a developer:

> "I think it's a process that's still under construction, but at the moment, I can't yet see how those things could look like. And I also do not yet see like that there is like a pattern or like this thread going through the ethics things that we have done. That would create for me, from a developer's perspective, a real benefit that I could incorporate (DEV-5)."

Whereas that trial and error approach was interpreted by developers mainly in terms of a lack of applicable input, one ethicist made the interesting observation that it was not so much the content that should have been more coherent, but mainly the form. Here, s/he insisted on the frequency and continuity with which ethics should have been given a structural place in the process from the start. "It needs to be a regular thing," s/he said, "like we lost a lot of time finding time slots that worked for people. (ETH-1)"

The ethics toolbox and the ethics hammer

A very interesting metaphor that is repeatedly advanced as a kind of ideal in the context of 'seamless integration' is that of the 'ethics toolbox', one of the developers even talks about an 'ethics hammer':

> "But it's not like I would have been handed out like five ethics tools that I could use. And this is the ethics hammer, and I use it like this. And then my software part is going to be ethically more sound. So this, from my perspective has not yet happened. And this was actually what I was expecting in the beginning of the project. (DEV-5)"

The image of the hammer is telling, especially if we tie in with what we discussed in Part 1 about the typical 'embodiment' relationship in which a tool like a hammer is incorporated. A person who is hammering is not consciously thinking about the hammer, all his attention being focused on the carpentry project at hand. It seems that something similar is hoped from a 'ready-to-hand' ethics hammer that a developer should have in his toolbox to fix any ethical problem he might encounter while developing. As was further confirmed through participant feedback, the idea of an 'ethics toolbox' largely finds its origin in the fact that software developers, especially in the domain of location based services, are familiar with 'privacy-by-design', where privacy and security principles are typically brought into the development process via programming guidelines and checklists. Consider for instance the following quote:

> "So my frame of reference was also always a little bit like the privacy by design framework, where you have certain things that you consider at various stages and there might be some sheets that you can use to make sure you don't forget anything and you keep certain records. (DEV-7)"

Division of roles

Although it sometimes fades into the background as being taken for granted, the division of roles is key in the narrative repertoire around integration. Especially on the developers' side, it is quasi-evident that integration work should primarily be done by ethicists. This obviousness can be seen in the way it is often mentioned in passing, such as when a developer sees a clear gap that needs to be bridged (see above), wondering out loud: "So how can outcomes from ethical people or from philosophers be something that is sufficiently specific for the engineering people? (DEV-2)" Developers, so it seems, are happy to implement ethics input, and some are more eager than others to do so, but it remains to ethicists to facilitate and smoothen this implementation.

To a large extent, this 'division of roles' was inscribed in the project proposal. As we indicated earlier, 'ethics' and 'ethics by design' have been important discursive reference points in framing the integration of SSH-perspectives in development work, already within the project proposal. This made it quasi-evident to look towards the 'ethicists' when it comes to the actual work of integration. Assigning the EBD work package to the ethics partner in the project also facilitated this. This also raises the question of the importance of ownership of the integration process: how can developers be expected to take ethical considerations into account as long as they are allowed to see them as extraneous to their own practice? While we will elaborate on this point in 5.2, it is interesting to zoom in on the presence of an ethicist in the development meetings. On the developers' side, this is repeatedly cited as the most valuable and effective aspect of the EBD trajectory, as in this quote:

> "The highest impact on a day to day basis, I would say, would have been ETH-X working on those ethics topics and sitting in the dev meeting and asking questions and raising issues potentially (DEV-5)"

Here it is interesting to note that the decision to sit in the dev meetings happened on one of the ethicists' explicit initiative because s/he felt that what s/he picked up in the monthly team meetings was not enough to do her/his job properly. "I went into the dev team because the monthly wasn't enough information," s/he says and s/he immediately adds: "But I think it's the other way round as well". Interestingly implied here is that 'bridging the gap' (see above) can happen from either side and that it is therefore also up to developers to bridge the gap to ethicists, delve into ethics and make development practice accessible to ethicists. Where a developer argues that "ethics people should also be super computer scientists (DEV-1)" we see no reason of principle why this sentence should not be mutually applicable as 'dev people should also be super ethicists.' (Some of this report's recommendations elaborate on this point. See 5.2)

4. Geographies of responsibility at work in SIMPORT

For a practitioner to make ethics matter in software development work, three conditions must be met. First, one should be aware of the ethical and social dimension of development work; second, one must feel responsible for it; and finally, one must be empowered to take up this responsibility. To capture these three conditions, we advance the metaphor of 'geography of responsibility'. According to Sigl (2020), the metaphor of a geography helps to imagine how R&D practitioners position themselves within an imagined map of responsibilities, where some responsibilities appear close enough to be considered in their practices, while other responsibilities are imagined as being too far away to matter in situated R&D practices. The geography metaphor thus suggests responsibility as a context-dependent concept that derives from those relations through which an identity is constructed. In this sense, the narratives we employ to make sense of who we are and what we do in relation to others and the world, are the discursive matter that 'geographies of responsibility' are made of.

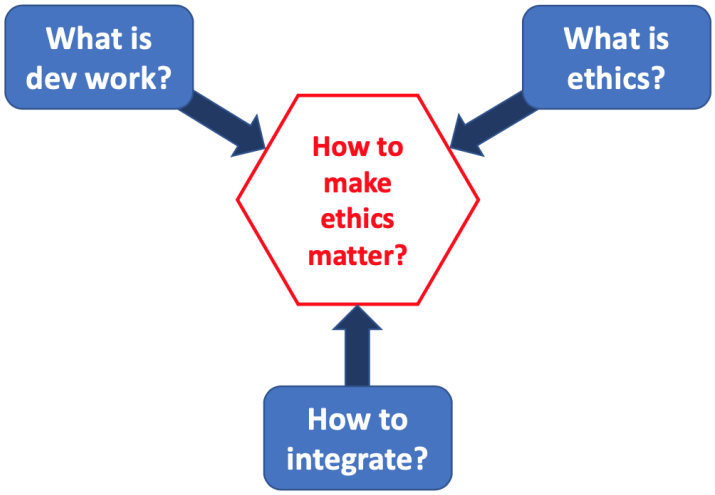

Figure 2 Narratives about development work, ethics and their interrelation enable and constrain practitioners in sketching out a geography of responsibility that guides them in responding to the question how to make ethics matter.

In the previous chapter, we showed how SIMPORT practitioners make sense of EBD using three narrative repertoires, on development work, ethics and integration, respectively. In this chapter, we go on to show how practitioners carve out their particular geography of responsibility by positioning themselves at the intersection of these three repertoires. Indeed, it is the interaction between these different repertoires, the arrangements they may form or the paradoxes and contradictions they create that matter when assuming or assigning responsibility. Thus, as illustrated in Figure 2, we want to show how narratives enable and constrain practitioners in creating their own 'geography of responsibility', how it guides them in talking about assuming and assigning responsibility (Felt, 2017) and, in the case of the SIMPORT project, in answering the question: 'how to make ethics matter in software development?'

Indeed, this is what we see happening in the interviews. Particular narratives are combined and brought into play to indicate what people consider themselves responsible for and whether or not they feel capable of assuming this responsibility. A range of possible 'geographies of responsibility' thus emerges, in which one clear pole of attraction stands out, namely the one that gathers around the pervasive call for principled guidelines and which we have captured here under the term 'ethics toolbox'. In contrast to such bureaucratic geography of responsibility that sees ethics mainly as something that can be formalised, compartmentalised and delegated, we also observe glimpses of an alternative geography by which practitioners in concrete practices feel empowered to reflect and respond to the complex ethical questions that arise at the interface of software development and society. We will term this a geography of response-ability and indicate where this was seen at work in the SIMPORT project. In chapter 5 we will argue how this geography can be further cultivated and fleshed out by focusing on narratives, practices and attitudes of response-ability.

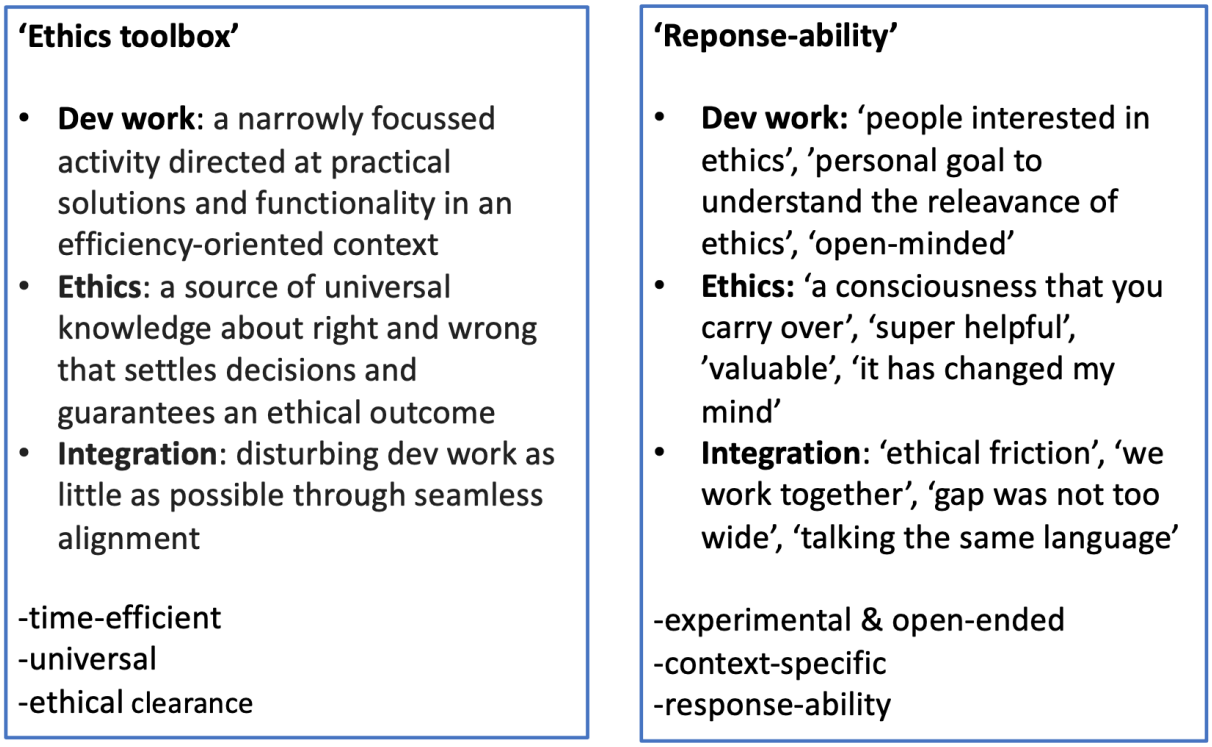

Figure 3 A schematic overview of the two distinct geographies of responsibility operating within SIMPORT.

4.1. The 'ethics toolbox' and the dawn of new bureaucracies of virtue

By using the label 'ethics toolbox', we want to highlight three aspects of this particular geography of responsibility. Just as a good toolbox holds the promise of broad applicability, efficiency and a quick fixing of the task at hand, the 'ethics toolbox' promises a time-efficient, universal approach that guarantees an ethically sound outcome of any development project. We by no means claim that SIMPORT has subscribed to this geography of responsibility. However, as regularly indicated throughout 3.3, we do see in the SIMPORT interviews that the

'ethics toolbox' exerts a great appeal for practitioners who try to make sense of the EBD trajectory and especially when they criticise what it has been or try to articulate what it should be. It is a kind of imaginary point to which they refer when they themselves are at a loss, an imaginary point that seems acceptable and attractive to many.

A particular combination of narratives…

Interestingly, our narrative analysis reveals how this imaginary point is based on a particular combination of particular understandings of development work, ethics, and integration. Where development work is understood as a narrowly focused activity directed at practical solutions and functionality in a strictly time- and efficiency-oriented context, ethics, in turn, is seen as a decision-making authority that guarantees an ethical outcome. Finally, integration is understood as an activity that disturbs the course of action (read: development work) as little as possible by seamlessly aligning itself. The concrete tools and guidelines that should put this geography of responsibility into practice and which are always referred to, but which no one has ever seen, now form the 'moral glue' with which these divergent expectations and promises are patched together: they not only promise to be tangible, time-efficient and universally applicable but also to guarantee ethical clearance: who would not want that?

…embedded in stabilising meta-narratives

In understanding the appeal of the 'ethics toolbox', it is equally important to see what gives this geography of responsibility its stability or, in other words, what keeps it so firmly in place. In doing so, it is instructive to identify the common factors shared by the specific narratives around dev work, ethics and integration: on the one hand, the prevalence of the linear innovation model, and on the other, a profound belief in technological solutionism. As discussed at length in Part 1, the **linear innovation model** consists in seeing the world of R&D and society as separate entities whereby society is a downstream field of application for technological solutions that can only be pursued in the seclusion of the R&D lab (Sigl et al. 2020; Godin 2006). While social impact and relevance of ongoing research is self-confidently communicated, undesirable implications are typically dismissed as 'unintended consequences' for which it is only a matter of time before they too are adequately resolved (Wynne 2011; Jasanoff 2003). Although the linear model of innovation is outdated from a social science point of view, it remains a very powerful imaginary that is held on to in every possible way. And so it happens that, under pressure of increasing concerns about the social and ethical aspects of innovation, the linear model is not opened-up but adjusted by equipping the R&D lab with an 'ethics toolbox' to pre-emptively fix such unintended consequences. **Technological solutionism** then refers to the techno-deterministic idea that technology is the answer to any challenge we face (Morozov 2013). Deeper even, it goes back to the ineradicable idea of solutionism, which sees the world as imperfect and analyses it in terms of problems that can, at least in principle, be solved. As indicated, it is an essential feature of what it means to be human. At the same time, as human beings, we should dare to be self-critical wherever this characteristic is elevated to the level of ideology. The question of whether every problem should be solved technologically perhaps deserves as much attention as the question of what socio-technical shape that solution should take. Technological solutionism stabilises the "ethics toolbox" geography of responsibility by positing the prospect of the 'right' technological solution. Here, social and

ethical considerations are not refuted or outsourced, but merely seen as additional tools to be deployed upstream in the innovation process, to eventually produce the right technological solution. What is crucial here is that this extra input does not substantially challenge the R&D trajectory, in the sense that it might question the underlying problem framing or the research directions pursued. Rather, it optimises the R&D trajectory, by 'designing out' any 'unintended consequences'. In its most bureaucratic form, technological solutionism sees 'ethical and social considerations' as additional tools for a linear innovation process, where ethical deliberation is reduced to a box-ticking exercise and application of the 'right' principles is believed to lead self-evidently to the 'right' technologies.

New bureaucracies of virtue?

The seductive force of the 'ethics toolbox' geography however, we argue, threatens to pull practitioners into 'new bureaucracies of virtue' that serve to normalise behaviour but is not authentic ethics (Felt 2017). In employing the term 'new bureaucracies of virtue' Felt (2017) laments the increasing subjugation of reflexive work to standardisation, division of labour and control from a capitalist logic of efficiency. Within SIMPORT, we indeed observed the seductive force at work of such bureaucratic ideals as formalisation and compartmentalisation.

By **formalisation**, we make reference to translating ethical work into the activity of filling out forms, as in the case of specific ethical review processes. In academia, we see this clearly at work for instance in ethics forms related to project applications, which foster tick-boxing rather than wider reflection throughout the research process. A form seems attractive because it creates 'tangible evidence', in this case, proving that ethical concerns have been considered. At the same time it is a tool of standardisation of what counts as the formal fulfilment of ethical work and allows for control and accountability (Felt 2017). As repeatedly indicated, this 'discrete charm of form' (Becker 2007) exerted a clear influence in SIMPORT practitioners' account of ethics by design, with the experience of privacy-by-design and the ubiquity of guidelines and cheat-sheets in the development world playing an important role. By **compartmentalisation**, we point towards the outsourcing of ethical work within a clear division of labour between ethicists and developers. Typically here, ethicists are supposed to take care of social acceptance and ethical assurance, which boils down to not allowing discussions on societal values to become embedded in R&D and be admitted to the core of R&D practice. In this context, Frauenberger et al. (2017) also speak of the risk of 'work-packaging' whereby the task of ethical reflection and deliberation within a research or innovation project becomes confined to a certain sub-task, or work-package, which is then left to the ethics experts on the team to be dealt with, without any input from the other participants of the project. This aspect also rears its head in SIMPORT. As already indicated, compartmentalisation often starts as early as naming integration work as a form of 'ethics', for which responsibility is also formally assigned to the 'ethics' partner. In the interviews, we saw this division of labour and the insistence on outsourcing the ethics work to the ethicists very clearly present (see 4.3).

Reflecting on her three year experience as an embedded SSH-scholar in a nanotechnology R&D group, Ana Viseu nicely depicts the 'ethics toolbox' geography of responsibility she was facing:

> "[T]he other scientists seemed to view my role as one of managing a narrow list of possible risks and consequences, so that if a researcher followed my instructions and ticked boxes, then I would bless them as 'social and ethical' and they would be free to do their work with no concerns. I was routinely (wrongly) introduced as an ethicist and was expected to find minimal, non-disruptive ways of dealing with social and ethical issues." (Viseu 2015)

4.2. Glimpses of 'response-ability' and the deconstruction of entrenched narratives

While the seductive force of the 'ethics toolbox' idea was clearly present in SIMPORT, and helped shape the responsibility landscape, our analysis also reveals **glimpses of an alternative geography of 'response-ability'** by which practitioners feel empowered to explore value-sensitive responses to societal and ethical issues and considerations. During the interviews, we repeatedly noted how the 'ethics toolbox' and the underlying narratives around development work were repeatedly refuted and deconstructed. As soon as development work or ethics are discussed on the basis of concrete collaboration, the entrenched core-narratives crumble and developers turn out not to be so narrow-minded after all, ethics is no longer a barrel full of non-committal knowledge that is of no practical use, and integration is given a concrete content according to the concrete ways in which people actually worked together. It is also notable here that those who actually did the work, rather than the supervisors, tend to speak about this in a concrete, nuanced and appreciative way. In contrast to the illusory 'ethics toolbox' promise of a time-efficient, universal approach that gives ethical clearance, the contours of what we put forward as a geography of 'response-ability' emerge. Characterised by an experimental and open-ended nature, such response-able practices cultivate an attitude where responsibility is not shrugged off but taken up in never easy but always unfolding, context-specific ways.

The idea of the narrow-minded developer tightly focused on functionality is at most a guiding myth. Although the interviews show that this myth is often employed, we also observe how this narrative is broken up at regular intervals. For instance, a number of developers indicate that they were excited to enter an EBD project and eager to learn about ethical issues. As one developer says: "I didn't personally have a deeper knowledge about those ethics topics. So, a personal goal for me was to understand the relevance of that (DEV-6)." Yet another puts it as follows:

> "I do feel like it [ethics of emerging software applications] is sort of an important topic that I would like more people to think about and sort of realise how things are sort of changing. What changes could be possible due to sort of emerging use of technology in that field. And yeah, I think this is the broader motivation for me behind this project. (DEV-3)"

As mentioned, the ethicists in the project also repeatedly indicate that developers are indeed open and willing to think about more than just the technical aspects. An interesting clue in this regard came from some developers who, when listing the members of the 'development team' in response to the question about who they worked most closely with, spontaneously mentioned the ethicist participating in the dev meetings. This indicates that, contrary to the

‘ethics toolbox’ narratives of dev work and ethics, an SSHer can indeed be considered a full development member. Evidently, further opening up this idea of development work requires developers to adopt the narrative of the social life of technology and broaden their ethos accordingly with a **self-reflexive attitude** (see Part 3).

Mirroring this, we also see how the core ethics narrative of the ‘ethics toolbox’ is regularly challenged and opened up, revealing the relevance of SSH perspectives for innovation and dev work, without reducing ethics to a yes/no decision authority. The idea that ethics is vague and non-committal for instance completely crumbles when practitioners talk about the input from the user and persona workshops. In addition, the presence of an ethicist in the dev meetings was perceived as “really cool" (DEV-6) and "super helpful" (DEV-4). Here, the idea of an ethicist looking over one's shoulder and making one pause to think about things was perceived as something productive rather than disturbing. Consider for instance the following quote:

> “So, like I mentioned in the weekly dev meetings, there's ETH-X joining and, even though s/he can't participate so much in the … like the deep technical stuff, s/he always had like ethics vision goggles on and yes, I think s/he always hinted us to some directions where we might want to go or which step we shouldn't go and which step we should have avoid. And I think that's the most valuable part of ethics by design. (DEV-4)”

As the last quote suggests, we are convinced that ethics, and SSH more broadly, has an important role to play in the R&D process itself, where it can bring critical perspectives on the social and cultural embedding of science and technology to fruition from a so-called 'embedded position'. As also evidenced by the SIMPORT-project however, (see ‘ethics as non-committal chatter’ in 3.1), the search for a fruitful position for SSHers within an R&D process is anything but evident. On the one hand, SSHers are worried that they cannot contribute to R&D practices because they are perceived as adverse critical observers whose ‘armchair’ critique is non-committal and detached from what matters for R&D practitioners. On the other hand, they fear losing critical distance and turning into uncritical research assistants. Smolka (2020) launched the term '**generative critique**' to further thematise this challenge on which we further elaborate in the context of Responsible Innovation (see 5.1.)

Finally, The SIMPORT experience also shows that the integration-narrative can be fleshed out differently than along the ‘ethics toolbox’ line of non-disturbing integration. We must and can indeed move away from the idea that integration is easy and that it hardly requires any time or effort. Integration is not a given that just needs to be implemented, not a universal procedure, but an outcome that has to be reinvented time and again by confronting the complex questions that arise at the interface of software development and society. As SIMPORT demonstrates, this requires a willingness to experiment with new, sometimes clumsy practices while embracing their open-ended character and the conviction that **new possibilities do unfold in working together**. This is what practitioners articulate when they talk about developing a common language, seeking coherence in a trial-and-error approach, or when they say they engage in "ethical friction” (ETH-1). Above all, this is echoed in the convincing tone in which several practitioners say: "we work together". It is from collaboration and the mutual confrontation and sensitisation this implies that empowering narratives of response-ability emerge (See Part 3).

PART 3: Towards response-ability in software development

In Part 3, we take stock of the theoretical and empirical findings of this report. We sketch out our vision of being an ethical developer and a developing ethicist and present our main recommendations for response-able software development, underpinned by the framework of Responsible Innovation. In the online version of this report (see https://simport.net/ethics-by-design/), we present some concrete methodologies (card-based discussion formats, prompts for integrating ethical reflection into agile software development processes, and ethics worksheets) closely drawing on the material presented here.

5. Making ethics matter in software development

How to make ethics matter in software development? It is fair to say that this is the central question on which this report is hung. In this final chapter, we take stock of the insights we have developed, drawing on both the SSH perspectives into the social nature of technology and innovation of Part 1 and the narrative analysis of the SIMPORT project of Part 2.

If we are serious about making ethics into a core dimension of software development, we must resist the bureaucratization of virtue as it is pushing through the landscape of digital innovation via the tidal wave of ethics guidelines. As we argue, ethics in the full sense comes with the cultivation of response-ability, which relies on the cultivation of specific sensitivities and modes of reflection. Response-ability cannot be mechanised and 'proceduralised' because it requires two things that are irreducibly personal: The response-able person (1) must be sensitive to what is at stake in relation to their own and other's actions, and (2) must be capable of navigating available options to channel that sensitivity into thoughtful and reflectionary action. Thus, the core of response-ability is being personally responsive, which refers to a capacity of sensing and judging. Accountability (understood as reporting to external observers who tick checkboxes) is only secondary. The idea that there would be 'operationalisable' quick fixes for the ethical and social challenges posed by the digitalization and datafication of the private and public spheres is an illusion. In the face of such far-reaching issues, response-ability and unguaranteed decision are the only signposts, making a persistent appeal for ethical deliberation and reflection throughout the whole technology development cycle.

In what follows, we pursue this thread in two steps. First, we propose Responsible Innovation as a framework within which a more robust approach to Ethics by Design becomes conceivable. As an intellectual movement closely aligned with the SSH perspectives presented in Part 1, Responsible Innovation has gained prominence within both policy and research in recent decades. Understood as translating SSH's critical view of the social nature of technology into technology development practice, Responsible Innovation serves as a stepping stone to shape response-ability within software development projects.

In a second step, we sketch out what we envision as the ideal of ethical software development in the form of six key recommendations. These recommendations proceed along the three central dimensions of narratives, practices and attitude (ethos). We connect our recommendations to the narrative analysis presented in Part 2 and the concept of 'geography of response-ability' we put forward. Finally, we provide a number of concrete pointers on how to set up a work culture that fosters Ethics by Design (EBD) by prioritising an ethos of response-ability.

5.1. Responsible Innovation as a broader framework

As an endeavour to integrate ethical concerns into the software development process with a view to developing more socially robust innovations, EBD does not stand alone. Although the boom in ethics guidelines and the way they are framed often give that impression, EBD need not start from scratch and can build on pre-existing repertoires of interdisciplinary collaboration aimed at the integration of ethics and technology development. In the light of our repeated emphasis on responsibility, we argue that EBD should be framed within a much broader debate on the relationship between research, innovation and society that has unfolded over the past few decades and which, under the term 'Responsible Innovation', is increasingly emerging as what Shanley (2022) calls an intellectual movement. Below, we briefly outline the historical emergence of Responsible Innovation and then elaborate on socio-technical collaboration as one of its key approaches.

The emergence of Responsible Innovation

In the aftermath of World War II, the devastating impact of the atomic bomb led to a profound awareness of the destructive power of technology and its ethical implications. Beyond that, the emerging environmental consciousness of the 1970s and 1980s drew attention to the potential damage of unchecked technological progress. Towards the end of the twentieth century, the persistence of complex sustainability issues, such as climate change, biodiversity loss and social inequality, led to the recognition of the inadequacy of classical discipline-based research in addressing them (e.g. Gibbons 1999; Kates et al. 2001). Mode2 (Gibbons et al. 1994) and Post-Normal Science (Funtowicz and Ravetz 1993) were among the main conceptual innovations in these discussions, arguing that the relationship between science, innovation and society was changing, putting forward inter- and transdisciplinarity to go beyond the socio-technical divide (Snow 1959) in search of socially robust knowledge and technology.

This evolution was fuelled and reinforced from within academic research in the intertwined fields of philosophy and sociology of science and technology, arguing that research and innovation cannot adequately be understood in isolation from society. Primarily prompted by SSH insights on the social nature of technology (see PART 1), a body of work concerned with ideas such as 'technology assessment' (Guston and Sarewitz 2002; Rip and Kulve 2008) and 'anticipatory governance' (Barben et al. 2007, Guston 2014, Nordmann 2014) emerged, centering on proactive decision-making and evaluation of emerging technologies to address potential risks, ethical concerns, and societal implications before their widespread adoption. The Human Genome Project has played a pioneering role in this regard, launching, more than 30 years ago now, a call for proposals from social scientists, ethicists, lawyers and others to explore the ethical, legal and social implications (ELSI) of mapping and sequencing the human genome. The inclusion of such an 'ELSI programme' within a high-profile international research effort has subsequently spurred a flow of similar programmes, mainly in 'hot' domains such as bio- and nanotechnology. Hereby the European ELSA (with Aspects instead of Impacts) variant emphasised not only to focus on the negative downstream 'side-effects', but to take into account the broader set of upstream questions of the priorities and directions for research-driven societal change.

While the establishment of these concepts and related practices are considered to be important milestones in the governance of research and innovation, there was often a feeling of unease when these reflections were relegated to separate work packages in projects or toward the end of innovation processes (Volker et al. 2023). Around 2010, in response to the growing plea to make ethical and social issues and considerations an inherent part of research and innovation practices, the closely related concepts of Responsible Innovation and Responsible Research and Innovation gained traction in framing responsibility-related issues for academics and policy makers alike. Although the terms Responsible Innovation and Responsible Research and Innovation emerged in parallel, they are nonetheless considered to be different things. While Responsible Innovation (RI) typically refers to an academic discourse that seeks to "open up technovisionary science and innovation, creating spaces for discussion, analysis, debate... aiming to provide a measure of social agency in technological choices (Owen and Pansera 2019, 39)", Responsible Research and Innovation (RRI) represents a specific policy artefact rooted in the European Commission's Science in Society program[7].

When RI took shape as a central pillar of the EU's Horizon 2020 funding program, it seemed to represent "an emerging zeitgeist for responsible innovation" (Owen et al. 2013). As a buzzword, it was heavily loaded with expectations: most importantly that R&D should become 'responsive' to societal issues and concerns, i.e. that research processes would be reshaped to trigger "innovation that looks different" (Owen et al. 2013). Crucially, this implies a second expectation, namely that researchers reimagine and broaden what they see as their responsibilities, from traditional core values such as research integrity to broader responsibilities of considering societal issues and concerns and responding to them in their practices. For this reason, the term 'responsibility' was put centre stage to express the expectation that these interactions are meant to change both research processes and their outcomes. A crucial point here is that 'responsibility' contrasts logics of choice that are still dominant in research and innovation with a logic of care that focuses on engaging in long-term collaborations addressing matters of concern (Volker et al. 2023).

In the slipstream of EU's Horizon 2020 program, a distinct field of RI scholarship emerged with the establishment of a high-impact journal (Journal of Responsible Innovation) and globally distributed RI programs and projects. On the one hand, the focus here is on conceptual-theoretical research; on the other hand, concrete RI projects are being set up, which in turn are subjected to theoretical analysis and reflection. In this sense, RI's motivations clearly intersect with a number of key concerns within STS, such as understanding the complex trajectories and societal dimensions of technologies, exploring and anticipating the potential impacts of emerging technologies, promoting the importance of interdisciplinary collaboration, and prioritising the inclusion of and engagement with diverse stakeholders throughout the process of technology development (Volker et al 2023). In its scholarly approach, RI regularly distances

[7] Unless we are talking about the specific EU policy artefact, in what follows we will use the acronym RI to refer to both RI and RRI as part of one intellectual movement. To the extent that any boundaries exist between the two in terms of the ideas, people, institutions and resources which have formed the basis of an international network, they are porous (Shanley 2022).

itself from a specific economic focus often implied within the EU interpretation of RRI, that innovation should align itself with social expectations centred on economic growth and job creation (Van Oudheusden 2014).

Socio-technical collaboration at the 'midstream' of technology development

An interesting perspective on RI is obtained by thinking in terms of the stage in which technology governance intervenes with respect to an R&D process. 'Upstream' governance deals with the pre-R&D stage and attempts to steer technology development through research policy and technology assessment. 'Downstream' governance focuses on the post-R&D stage and seeks to manage the uptake of technology through regulation and market mechanisms. With a central focus on bringing ethical and social considerations to bear on decisions taken right in the R&D laboratory itself, RI can thus be seen as a form of 'midstream' governance (Fisher et al. 2006; Schuurbiers and Fisher, 2009). A central approach to achieving what Fisher calls midstream modulation of R&D decisions is that of socio-technical collaboration, which can be characterised by three central features (Fisher 2015). First and most evident, socio-technical collaboration connects scholars and researchers across socio-technical divides. Second, collaborators work in close proximity, usually by involving an SSH scholar in a techno-scientific space. Third, it combines knowledge production about techno-scientific practices with contribution to change in how techno-scientific practitioners identify and engage with socio-ethical dimensions of their work. In our view, socio-technical collaboration is aptly described as an attempt to translate the conception of RI as a critique of science and technology development into practice through interdisciplinary collaborations between scientists and engineers on the one hand, and SSH scholars on the other. Within STS, socio-technical collaboration is also captured under what is called the 'engaged program.' For engaged STS researchers, understanding the social nature of science and technology is seen to be continuous with promoting socially responsible science and technology. Continuing the tradition of laboratory ethnography studies that gave birth to the STS discipline, STS researchers hereby take an embedded position in an R&D laboratory, using participant observation as a way to make their SSH insights fruitful in the midstream of technology development.

The way in which EBD took shape within the SIMPORT project, including by embedding an ethicist in the software development team, manifestly belongs to what counts as socio-technical collaboration within RI. In this sense, it is apt to note that in the recent 'ethics boom' in the context of digital technology, the link to RI and more specifically to socio-technical collaboration is rarely, if ever, made. We are convinced that the ambition to integrate ethics into software development has much to gain from a closer engagement with the intellectual movement of RI, and in particular its rich, documented and researched experience in rendering critical SSH perspectives fertile for technology development. In the next section, we therefore briefly outline the STIR approach, which is a proven, generic socio-technical collaboration methodology deployable across R&D discipline, originally developed by Fisher (2007). Based on our findings from Part 2, it is our conviction that the core of the STIR methodology, and in particular its structured, iterative and self-restrained approach, lends itself to being embedded in 'agile' software approaches such as scrum (See also Zuber et al. 2022). In the online version of this report (see https://simport.net/ethics-by-design/), we elaborate on such a scenario.

Socio-Technical Integration Research (STIR)

In what became the first STIR pilot study, Fisher immersed himself as an 'embedded' scholar in the Thermal and Nanotechnology Laboratory in Boulder, Colorado (Fisher 2007)[8]. Over the course of three years, he interacted with numerous researchers, conducted interviews, participant observation, and archival research. Through his fieldwork, Fisher noticed that researchers were often already integrating social and ethical concerns into their work without being aware of their own role in such a de facto integration. He discussed his insights with the researchers and assisted them in articulating further latent ethical reflections. To facilitate this process, he developed what he called the STIR 'decision protocol' (Fisher 2022). Based on four components – opportunity, considerations, alternatives, outcomes – and attendant questions, the STIR-protocol, as illustrated in Figure 2, is used to unpack concrete, day-to-day research decisions as they unfold.

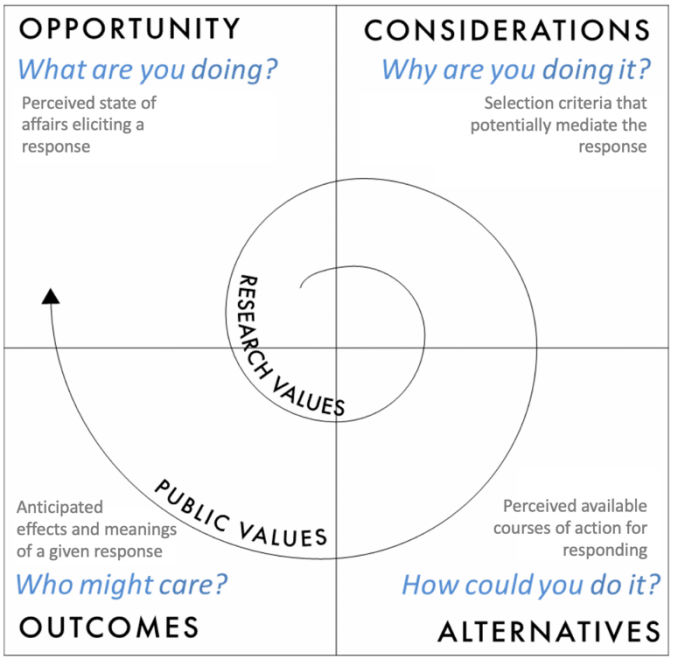

Figure 2 The STIR Decision Protocol (adapted from Fisher 2022): Research decisions are unpacked according to four dimensions that can be captured in four short questions: what are you doing? (opportunity); why are you doing it? (considerations); how could you do it? (alternatives); who might care? (outcomes). An empty four-quadrant diagram serves as a visual underlay (digital or paper sheet) to collaboratively map responses in real time.

On the basis of an early version of this protocol, Fisher collaborated with his technical counterparts in a structured way to map out their decision-making processes on a regular basis. They explored together how ethical and societal considerations were already or could be aligned with technical alternatives, and how researchers could further cultivate or bring about this alignment in practice. As a result, researchers enhanced their capacities for integrating socio-ethical considerations with technical alternatives and started to perceive socio-technical

[8] In the subsequent historical description of the STIR approach, we closely follow Smolka's account in the context of her development of the concept of generative critique (Smolka 2020).

integration as “an integral part of th[eir] work” (Fisher et al. 2014). In addition, they voluntarily altered their decisions and practices concerning experimental setup, material choice, and safety strategies. Hence, the STIR pilot study realised a twofold agenda. On the one hand, it rendered research and innovation processes more akin to what would eventually be articulated as RI by integrating societal and ethical considerations into these very processes and products. On the other hand, it resulted in knowledge production about whether such an integration is possible in the first place, about how it unfolds over time, and on how it co-produces ethics, science, and technology on the laboratory floor. Since then, over 70 known STIR studies have been conducted, demonstrating the effectiveness of the approach not only at assessing capacities but also at enhancing learning, deliberation, and practical change in a variety of R&D domains (Fisher 2022).

In view of our narrative analysis elaborated in Part 2, it is instructive to think of STIR as a method of ‘ethical cartography’ which serves to map out the geography of responsibility that is at work in a particular R&D laboratory. Pursuing the cartography metaphor, the SSH scholar may be seen as possessing the cartographic expertise to map the social and ethical aspects of day-to-day research decisions, whereas the R&D experts are to be considered the local guides who know the technical terrain inside out. Focusing the STIR-dialogue on any research decision that presents itself (e.g. which image to use in a presentation, which database management system to use, which colleague to involve, etcetera), a repeated appeal is made to the discursive reference points R&D experts draw on to make sense of and deal with the decisions at hand. If sustained for a sufficiently long time[9], mutual sensitisation of the disciplines involved will take place, enabling ever more in-depth exchanges (see the discussion on ‘talking the same language’ in *3.3. Making sense of integration*). Thus, step by step, deeper interdisciplinary and reflexive insight is gained into the ‘geography of responsibility’ (and the narratives, imaginaries and metaphors therein contained – see also *2.2. Technology design as culturally embedded abstraction*) that R&D actors mobilise to motivate research decisions, which in turn opens it up for responsible reconfiguration.

Understood this way, STIR serves as a practice for **generative critique** because it facilitates explorations of alternative ways of doing research and innovation in a collaborative learning process (see the discussion on generative critique and the ethics narrative in *4.2. Response-able practices and the deconstruction of entrenched narratives*). By using participant observation as a resource for collaboration on issues of direct relevance to techno-scientific practitioners, an embedded scholar avoids armchair critique. The collaborative nature of interactions in turn helps circumvent a service-subordination mode of interdisciplinarity. Ethics cannot be ‘outsourced’ to an embedded scholar because what counts as ethical or not, unfolds together with the geography of responsibility and may be reconfigured over time in discussions about routine techno-scientific practices.

[9] While an entry-level STIR study typically spans 12 weeks, this approach can be deployed flexibly, for instance in the form of more sustained or repeated campaigns.

5.2. Recommendations for response-able software development

In chapter 4, we argued that the key to unlocking the actual potential of Ethics by Design (EBD) lies in cultivating a 'geography of response-ability' in which practitioners feel capable and empowered to navigate the complex socio-technical landscape of digital innovation in relation to their own daily work. We argued and illustrated that whether practitioners feel empowered to do ethical work in software development depends on how they map out themselves and their practice in relation to others and society. In this respect, the idea of a geography suggests that certain responsibilities are 'on their map' and thus considered relevant while others are not, as well as whether those responsibilities that are 'on the map' are considered within reach and thus assumed rather than delegated. Part 2 also found that the one-dimensional geography of the 'ethics toolbox' exerts a strong appeal within the realm of digital innovation and how this, accompanied by the promise of a universal, operationalisable, ethical quick-fix, hinders the development of context-specific response-able practices, narratives and attitudes.

Our recommendations therefore focus in the first place on addressing the **illusory idea of the 'ethics toolbox'** and the underlying narrative resources that shape and stabilise this particular geography of responsibility. In Part 2 we observed how the 'ethics toolbox' imaginary collapses in the face of concrete SIMPORT practices that, however imperfectly, attempted to integrate ethical work into software development. As discussed in 4.2, we repeatedly observed how, discussing SIMPORT's concrete collaborative practices during the interviews, the entrenched core narratives crumbled and gave way to something else: developers turned out not to be so narrow-minded after all, ethics ceased to be a barrel of non-committal knowledge that was of no use in practice, and integration took on a tangible meaning of actual ethical development work. Throughout their account of what happened, SIMPORT practitioners showed how, at regular times, they felt effectively empowered to think, anticipate and respond to the often complex questions of digitalisation and datafication that arise at the intersection of software development and society. Hence, our recommendations capitalise on the importance of holistic, non-compartmentalising (counter-)narratives for EBD. We seek to cultivate narratives that make practitioners own the ethics aspects of their work; that is, we take practitioners by their response-ability as ethical agents that is indispensably their *own*. This means reconfiguring the presumed geography of responsibility into a geography of response-ability whereby ethical aspects are not framed as contained sub-regions that can be managed through specific sub-routines of standardised work-procedures, but rather are a persistent dimension underlying all positions on an innovation project's route map.

Centrally, the transition to such a geography of response-ability means *owning* ethics on the part of the developers. True response-ability is rooted in sensitivity (curiosity, wanting to know, wanting to reflect) and the capacity to channel perceptions and observations into thoughtful and reflectionary analyses, narratives, practices and attitudes. As such, response-ability is indispensably personal, it cannot be delegated or mechanised. It takes time and space to cultivate appropriate sensitivity and awareness both personally for each of the practitioners involved and collectively for teams of socio-technical collaboration.

Narratives, practices and attitudes

Overall, our recommendations rely on three entangled aspects or 'angles' involved in an empowering ethics approach: Response-able narratives, practices and attitudes.

(1) Narratives are the ideational, discursive representation of geographies of responsibility: Narratives influence beliefs, values, orientations, motivations and provide representational ground to what 'is at stake', what 'one can do', what is the 'bigger project', what is 'expected of oneself' etc. Narratives constitute the discursive layer of a geography of responsibility by pre-structuring conceivable practices, actions, alternatives, points of responsibility etc.

(2) Practices, in turn, form the procedural, performative layer of geographies of responsibility. Things that are done, and how they are done, shape and confirm narratives by providing first-hand experiences, practical knowledge, inherited habits and an operative status quo with a certain inertia. Narratives and practices can therefore mutually sustain and support each other, so that none of the two forms a causally closed system, but rather rely on each other.

(3) Attitude, or ethos, to use the ancient Greek term (the German translation 'Haltung' fits better than the English 'attitude'), in turn, is the concept that describes the field of mediation between practices and narratives that is to be located in each single subject in the field. Nobody just mechanically follows inherited practices, nobody just buys into established narratives. Ethos refers to the way practitioners 'hold' themselves within the field of narratives and practices by means of specific sensitivities that are linked with specific capacities of (re-)acting. Engaging in some of the practices and narratives with bad feelings, in others with full conviction, doubts, hesitation or, to the contrary, pleasure or maybe cynicism, sensing subtle frictions or unease or believing 'to change the world for better' – these are some emotional markers of one's individual attitude through which one is embedded and entangled in the geography of responsibility that is made up from narratives, practices and all of the practitioner's attitudes.

Importantly, attitudes can be cultivated through (joint as well as personal) reflection. Through attitudes, one can actively relate to practices and narratives, for instance, by retreating from responsibility by resorting to a narrative of compartmentalisation and delegation, or assuming responsibility, by striving to see potential impacts, or by actively selecting counter-narratives or subversive practices to experiment with possible changes. Thus, one's own attitude can (and must) be the locus of active ethical interventions. An attitude of taking responsibility, of striving to see the bigger picture of remote impacts or side effects of one's own business, or, on the contrary, engaging in cynicism or responsibility-averting ideas of compartmentalisation and delegation is ultimately each practitioner's active ethical choice. One is not passively exposed to narratives and practices; even if one cannot easily change many of them, ethics begins with the development of a sensitivity and reflective attitude (or stance) towards those practices and narratives. Narratives and practices of self-reflection (best to be done in groups) are important instruments for cultivating response-able attitudes.

In sum, all three aspects – narratives, practices and attitudes – mutually fertilise each other; doing ethics effectively means working on all three dimensions at the same time. As attitude is the central mediator that translates narratives into practice and, conversely, inherited practices perpetuate narratives through the individuals involved, ethics is inextricably personal: Everyone is involved in ways that cannot be delegated, taking responsibility is something everyone has to do, it is essentially personal ('höchstpersönlich'), it cannot be done for you.

Recommendations

Along the three entangled aspects of narratives, practices and attitude, we have six recommendations in pursuit of a geography of response-able software development.

Response-ability from the angle of practices:

1. **Make time and space for building context-specific response-able practices from within innovation contexts instead of importing generic approaches**. What societal responsibility can mean in concrete innovation settings should remain an ongoing, continuously re-debated empirical quest. This quest should be led in a way that is self-owned by the project members and sensitive to the perspective of external stakeholders (e.g., groups and individuals affected by the innovation). Three pointers flesh out this recommendation:
    - **There is no "ethics module" offered by external ethics service providers** that can be plugged into your project to take care of ethics questions on your behalf. Counter to the habit of dissecting and compartmentalising development processes into small, manageable chunks, there is no delegation and compartmentalisation of ethics work. All practitioners, across hierarchies and seniority levels, must have time and space to become part-time ethicists themselves.
    - In the pursuit of such an open-ended, empirical and context-sensitive approach, **being a part-time ethicist should take a starting point in concrete software development practice.** Regardless of whether software developers are aware of it or not, their practices always already involve a wide range of social and ethical considerations. This implies that any development decision, no matter how mundane, can be the starting point for mapping out their geography of responsibility that is at work. In this respect, embedding iterative ethical deliberation in agile environments with an emphasis on team empowerment (such as scrum) seems a viable, meaningful and necessary approach to pursue. The STIR methodology presented in this report offers a concrete starting point to this end (see 5.1. and 6.3).
    - **Use tools to map out responsibilities, not to shrug them off.** Within the broad field of RI, a plethora of 'instruments' and 'tools' have been developed over the past 30 years, including scenario analysis, mediation analysis, design workshops, focus-groups etc. Throughout this report, we have argued strongly against the idea of 'ethics tools' as promising a quick ethical fix (see in particular 4.1). Yet we are convinced that many such tools, if applied in an open-ended, experimental and

context-sensitive approach, are indeed useful in fleshing out an EBD project[10]. The question is just: useful for what? Rather than instruments to 'ethically' settle decisions, they should be understood as cartographic tools to map out the ethical and social issues on which these decisions impinge in order to inform and sharpen the sense of responsibility of everyone involved (attitude aspect). If, in the end, decisions need to be reached based on (joint) ethical consideration, this will be (and have to be) the decisions of the persons involved, not the output of a mechanised ethical arbitration procedure. Ethics decisions stick with the persons involved, they cannot be the product of a routine.

2. **Ethics by design thrives on socio-technical collaboration**. Ethical and critical engagement with digital innovation promises to be effective and impactful if it relies on sustained collaboration between software developers on the one hand and ethicists on the other. This collaboration should be shaped along the following imperatives:
    - **Ground the collaboration on the basis of equivalence.** This already starts when writing the project proposal by ensuring that the collaboration is framed in terms of two partners contributing to a common outcome. This can be done, for example, by making socio-technical integration a central work package for which both partners are responsible. In contrast to the 'ethics washing' observed in many projects where typically an add-on work package 'social and ethical issues' or, even worse, 'social acceptance' is assigned to the SSH partner, which serves as a free pass to outsource the integration work to the SSH partner. In addition, it is also important to give socio-technical collaboration a structural place in the project workflow from the outset, for example by scheduling integration activities on a fixed frequency and blocking appropriate time of all partners across all seniority levels. If integration has been discussed within the SIMPORT project as 'bridging a gap' between software developers and ethicists, then equivalence implies that building that bridge must be done from both shores (see 3.3)
    - **Provide enough time for the mutual sensitization of all involved disciplines.** Socio-technical collaboration does not happen overnight. Aligning expectations and building trust between collaborators to invest in an open-ended process without clear outcomes takes time. We referred to the notions of 'inter-language' and interdisciplinary 'trading zone' to indicate that collaborators from both disciplines need to familiarise themselves with each other's languages, frameworks and perspectives to enable in-depth exchange (see 3.3). Only in this way can they collaboratively cultivate a 'geography of response-ability' from which they can take development decisions that they are ready to defend from an ethical and social perspective.
    - **Withstand the affective tensions of interdisciplinary collaboration.** For socio-technical collaboration to succeed, and even to strengthen it, it is crucial to be attentive to its affective dimension. Anyone who has ever done interdisciplinary

[10] In search of inspiring RI-related tools, practices and projects, the CTA Toolbox project at Twente University (https://cta-toolbox.nl/) and EU's RRI Tools project (https://rri-tools.eu/) offer good starting points.

work knows that this comes with moments of silence, antagonisms, discomfort and frustration. As we have also seen within SIMPORT, those are moments when one does not understand the other, perceives their approach as irrelevant, counter-productive, naive or a waste of time, or finds their argument nonsensical and hard to translate in one's own practice (see e.g. the discussion on ethics as non-committal chatter in 3.2). Such moments are often swept under the carpet as quickly as possible or simply ignored because they are understood as symptoms of the malfunctioning of interdisciplinarity. If they happen repeatedly, those moments might even be the reason for parts of the team retreating from the interdisciplinary engagement, limiting their scope to their own sub-project and delegating communication in the wider team to specific team members (compartmentalisation). However, following Smolka et al. (2021), such affective tensions may "serve as a resource for engaging collaborative difficulties that stem from solidified disciplinary boundaries, epistemological and cultural differences and asymmetric funding provisions (1079)". Cultivating attention and sensitivity to meaningful differences, Hillersdal et al. (2020) suggest, may in turn lead to "other ways of addressing a research object and ultimately a social problem that do not simply reproduce a focus on boundaries between disciplines (73)".

Response-ability from the angle of narratives:

3. **Culturally entrenched narratives surrounding software development and its societal responsibility should be challenged.** Ethics considerations in development teams have to include open reflection on the role of both software developers and ethicists in society. A particular point of reflection is how the conception of those roles is pre-configured by the culturally entrenched narratives undergirding the idea of the 'ethics toolbox'. Building on the observation that the latter reduces ethical work to bureaucratic practices of formalisation and compartmentalisation that fail to engage all practitioners in a holistic sense, this report particularly challenges the following narratives (see 4.1):
    - **That software developers are narrow-mindedly focused on functionality**, creating the illusion that ethical considerations should be fed to them in the form of an 'ethics toolbox', the simple application of which guarantees ethical clearance.
    - **That ethicists and philosophers** operate from a distance to practice and economic reality, employing vague and non-binding concepts and frameworks (cfr 'non-committal chatter' (see 3.1) and 'armchair critique' (see 4.2)).
    - **That ethics presents a context switch for development work**, and that it should disturb the latter as little as possible (see 3.3).
    - **That innovation happens in seclusion from society,** whereby ethical and social considerations are delegated to 'social' actors such as policy and regulation or SSH scholars, who are then expected to take care of social acceptance of the outcomes (cfr. 'linear innovation model'; see 4.1)
    - **That innovation is the a priori solution**, that it harbours the solution to any problem, even complex ethical and social aspects of innovation; in other words, that nothing

should stop innovation, as any unintended consequences will eventually be solved by … more innovation (cfr. 'technological solutionism'; see 4.1)

4. **Inspiring vocabularies are needed to articulate and discuss the social nature of technology.** That technology is a neutral means to an end (instrumental view of technology) remains a powerful myth, mainly used to keep ethics and societal interference out of the innovation process. In Part 1 of this report, we sought to give relevant words and concepts to the description of technologies as political through and through, loaded as they are from the conception phase on with power, competing interests, and ideology. Following this theoretical discourse, we recommend:
    - **Acknowledge that technological artefacts do much more than fulfil predefined functions.** Always anticipate that the actual functions and use-cases of a technology exceed what developers had in mind. Try to understand and to be attentive to how artefacts create new use-cases, shaping in non-neutral ways how we perceive and act in the world. Anticipating this mediating, always ambivalent role of technology-in-use should be a core task of any development and design process. (cfr. Technological Mediation Theory; see 1.1)
    - **Try to envision technology as a socio-technical network**. That is: A technology can never be explicated without reference to its social and cultural embedding. Technology never stands on its own. It only functions as part of a socio-technical network of devices, rules, perceptions, narratives and practices that need to be aligned. This suggests that the development and design process should accommodate the perspectives of a wide range of interested actors and social groups, with critical attention to mutual power relations. (cfr. Science and Technology Studies; see 1.2)
    - **Acknowledge that technology is culturally biassed and strive to find out how this plays out in concrete cases**. From conceiving an affordance, culturally embedded understandings about the possible, desirable or expected social benefits and purposes play a role. Bringing these cultural entanglements into focus in order to question them critically requires a far-reaching degree of sensitivity and reflexivity, as such biases often sneak into the development process through the involved persons' implicit assumptions, perceptions and values. (cfr. Critical Theory of Technology; see 1.3)

Response-ability from the angle of attitudes:

5. **Cultivate an ethos of self-reflexivity – because thinking about what you are doing changes what you are doing.** There is an essential portion of ethics that is indispensably personal, in that it cannot be outsourced or delegated. Ethics is not a subroutine of the development procedure that can be programmed by a specialised team member or be imported from a library. Rather, ethics is the systematic and cultivated wonder about the social, societal and political status of one's own actions, and as such it accompanies every move you do. This is something everyone has to do, by themselves, very personally. It is, after all, *you,* raising or not raising objections, seeing or not seeing the bigger picture, caring or not caring to

understand what is at stake with respect to societal impact. Much of ethics therefore is in the question, *how,* and *with what kind of awareness,* you are personally doing what you are doing. If you are a software developer, you are in a comparatively powerful societal position so that your actions *do* have an impact on others. Along these lines, we recommend:

- **Regularly make time to reflect on what kind of narratives and 'truths'** are driving your relation to your work and your project. Where do they come from? Why are they compelling to you? Whose interests do they serve? What do they make visible and invisible? Are they powerful? In what way do they empower or disempower you? In what way do they contribute to keeping your team, the apparatus of your company, your brand or business sector, ultimately society etc. together?
- **Regularly make time to reflect on what kind of practices, organisational structures and routines** you are following and why? Where do they come from? Why are they there? How did you inherit them? Whose interests do they serve? Are they powerful? In what way do they empower or disempower you? In what way do they contribute to keeping your team, the apparatus of your company, your brand or business sector, ultimately society etc. together?
- **Regularly reflect, what are your own interests in all of this?** If you prefer certain practices or narratives, why? Psychologically and subjectively, why these and not other narratives or practices? If you lean more on the side of compartmentalisation and delegation with respect to ethics: Why is that a compelling narrative to you? What would you need to assume responsibility yourself?

6. **Cultivate a continued debate, collective sensitivity and joint self-reflection in your team.** Critical sensitivity and self-reflexivity are ultimately only possible if they are accompanied by dialogue and peer discourse. When reflexivity aims at transcending daily routines and the narrative 'obvious', people in the same as well as in radically different situations can help you find new words and narratives, experiment with new practices and probe for little changes. EBD means that there is room for this outside the imperatives of effectiveness and productivity that is measurable in seed and capital. Three pointers flesh out this recommendation:
    - **Allow spaces and time for joint reflection.** An EBD process requires pausing and reflecting on a regular basis. It requires spaces of encounter and debate outside planned meeting choreographies or project organisation schemes. Time and space needed for this are not outside working hours, but must be part of healthy project management.
    - **Make use of methodological approaches for group reflection.** Space and time for questioning ongoing activities can be created, for instance, in one-on-one interviews or focus-group discussions. This helps cultivate sensitivity to one's own and other's perceptions and activates the narrative frameworks from which these activities acquire meaning. As we have seen in the analysis of the SIMPORT interviews, such a reflective approach often reveals new interpretations and wordings, which in turn open up new possibilities for fleshing out concrete activities. However, while reflexivity is crucial, it is too often too easily advocated, as if it were a foregone conclusion. One cannot *simply* 'see' one's own blind spots and biases, especially

those deeply ingrained in one's professional culture. Engaging with other disciplines and perspectives does help in becoming aware of one's prejudices, but that does not mean that all aspects of one's thinking and motivations are fully known and accessible. Sensitivity and reflexivity are possible, to a certain extent, but there is no meta-perspective from which to approach things neutrally. In other words, sensitivity and reflexivity cannot possibly be moulded into a decision-making 'tool'. This brings things full circle by ultimately pointing to a radical responsibility from which innovation, however reflexively pursued, cannot escape (see 6.2).

- **Stay on it. This is not an item on your to-do list that will be DONE at some point.** Ethics is a specific personal sensitivity (pursued with certain theoretical tools) that needs to be cultivated. Ethics is not something you can get done. It is not a finite set of factual questions (what should we do in order to do it in a 'good' way) that will be answered once, but rather the question of how to hold yourself as a response-able subject within the web of practices, narratives, goals, visions, and other constraints. Much of this requires curiosity to know about the broader picture, and self-consciousness to be able to understand and make sense of the complex societal entanglements of one's own actions.

Abbreviations and acronyms

AI - Artificial Intelligence
ANT – Actor Network Theory
CTT – Critical Theory of Technology
EBD – Ethics by Design
R&D – Research and Development
RI – Responsible Innovation
RRI – Responsible Research and Innovation
SCOT – Social Construction of Technology
SSH – Social Sciences and Humanities
STIR – Socio-Technical Integration Research
STS – Science and Technology Studies
TMT – Technological Mediation Theory
HCI - Human-Computer Interaction
UX - User eXperience
UI - User Interface

Bibliography

Barben, D. et al. (2008) Anticipatory governance of nanotechnology: Foresight, engagement, and integration. In The Handbook of science and technology studies, ed. E. J. Hackett, O. Amsterdamska, M. Lynch, and J. Wajcman, 979–1000. Cambridge, MA: The MIT Press.

Becker, P. (2007). Le charme discrete du formulaire. In M. Werner (Ed.), Politiques et usages de la langue en Europe (pp. 217-241). Paris: Editions de la maison des sciences de l'homme.

Benjamin, R. (2019) Race After Technology: Abolitionist Tools for the New Jim Code. Medford, MA: Polity

Buolamwini, J. and Gebru, T. (2018) Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency: 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html.

Eubanks, V. (2018) Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press

Noble, S. U. (2018) Algorithms of Oppression: How Search Engines Reinforce Racism, Illustrated edition. New York: New York University Press

Bijker, W.E. et al. (Eds.) (1987) The Social Construction of Technological Systems. MIT Press

Collins, H. et al. (2007) Trading zones and interactional expertise. Stud. Hist. Phil. Sci. 38: 657–666

Daly, A. et al. (2021) AI Ethics Needs Good Data. In: Verdegem, P. (ed.) AI for Everyone? Critical Perspectives. London University of Westminster Press. pp. 103-121.

Engen, C. (2022) Introduction to the imaginary of precision oncology. In: Bremer, A. and Strand, R. (Eds.) Precision Oncology and Cancer Biomarkers. Issues at Stake and Matters of Concern. Springer.

Feenberg (2002) Transforming Technology: A Critical Theory Revisited. New York: Oxford University Press.

Felt, U. (2017). "Response-able practices" or "new bureaucracies of virtue": The challenges of making RRI work in academic environments. In L. Asveld, van Dam-Mieras, R., Swierstra, T., Lavrijssen, S., Linse, K., & van den Hoven, J. (Eds.), Responsible innovation 3 (pp. 49–68). Cham: Springer.

Feng, P., and Feenberg, A. (2009) Thinking about design: Critical theory of technology and the design process. In Philosophy and Design: From Engineering to Architecture, eds P. E. Vermaas, P. Kroes, A. Light, and S. A. Moore (Dordrecht: Springer) pp. 105–118

Fisher E (2007) Ethnographic invention: probing the capacity of laboratory decisions. NanoEthics 1(2):155–16

Fisher E. (2022) Socio-technical integration research (STIR) manual. Unpublished manuscript

Fisher E. and Maricle, G. (2014) Higher-level responsive- ness? Socio-technical integration within US and UK nanotechnology research priority setting. Sci Public Policy 42(1):72–85

Fisher, E. et al. (2006). Midstream modulation of technology: Governance from within. Bulletin of Science, Technology & Society, 26(6), 485–496

Frauenberger, C. et al. (2017) In-Action Ethics, *Interacting with Computers* 29(2): 220–236, https://doi.org/10.1093/iwc/iww024

Galison, P. (1997). Image and logic: A material culture of microphysics. University of Chicago Press

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., Trow, M. (1994). The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies. London: SAGE.

Gibbons, M. (1999) Science's new social contract with society. Nature 402: c81-c84.

Godin, B. (2006). The linear model of innovation: The historical construction of an analytical framework. Science, Technology & Human Values,31(6), 639–667.

Gogoll et al. (2021) Ethics in the Software Development Process: from Codes of Conduct to Ethical Deliberation. Philosophy and Technology 34: 1085-1108

Guston, D. H. (2014) Understanding 'anticipatory governance'. Social Studies of Science 44(2): 218–242. doi:10.1177/0306312713508669.

Guston, D. H., & Sarewitz, D. (2002). Real-time technology assessment. Technology in Society, 24, 93–109.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines 30(1): 99-120

Hillersdal, L. et al. (2020) Affect and Effect in Interdisciplinary Research Collaboration. Science &Technology Studies33 (2): 66-82

Jasanoff, S. (2003). Technologies of humility: Citizen participation in governing science. Minerva 41(3): 223-244.

Jasanoff, S. et al. (Eds.) (2015) Dreamscapes of Modernity. The University of Chicago Press

Kates, R.W., Clark, W.C. et al. (2001) Sustainability Science. Science 292: 641-642.

Krijger, J. (2021) Enter the metrics: critical theory and organizational operationalization of AI ethics. AI & Society 37 (4):1427-1437

Law, J. (2007) Actor Network Theory and Material Semiotics. In B. S. Turner (Ed.), The New Blackwell Companion to Social Theory (pp. 141-158). Wiley-Blackwell.

McLennan et al. (2020) An embedded ethics approach for AI development. Nature Machine Intelligence 2: 488–490

McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development? In Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. pp. 729–733

Morozov, E. (2013). To Save Everything, Click Here: The Folly of Technological Solutionism. PublicAffairs.

Mühlhoff, R. (2021) Predictive Privacy: Towards an Applied Ethics of Data Analytics. Ethics and Information Technology. doi:10.1007/s10676-021-09606-x.

Nordmann, A. (2014) Responsible innovation, the art and craft of anticipation. Journal for Responsible Innovation 1(1):87–98

O'Neil, C. (2016) Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Books

Owen, R. and Pansera, M. (2019) Responsible Innovation: Process and Politics, in von Schomberg, R. and Hankins, J. International Handbook on Responsible Innovation. Cheltenham: Elgar.

Owen, R., Bessant, J., & Heintz, M. (Eds.). (2013). Responsible Innovation. Managing the responsible emergence of science and innovation in society. Chichester, UK: Wiley.

Pinch, T., & Bijker, W. E. (2009) The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology (Anniversary Edition). MIT Press.

Rip, A., und H. te Kulve (2008) Constructive Technology Assessment and Socio-Technical Scenarios. In: Erik Fisher, Cynthia Selin, Jameson M. Wetmore (eds.), The Yearbook of Nanotechnology in Society, Volume I: Presenting Futures, Berlin: Springer 49-70

Rommetveit, K. (Ed.) (2021) Post-Truth Imaginations: New starting points for critique of Politics and Technoscience. Routledge.

Schuurbiers, D. and Fisher, E. (2009) Lab-scale intervention : Science & Society Series on Convergence Research. In: EMBO Reports, Vol. 10, No. 5: p. 424-427

Shanley, D. (2022) Making responsibility matter: The emergence of Responsible Innovation as an intellectual movement. Maastricht University: PhD thesis ISBN 978-94-6469-105-4

Sigl, L. et al. (2020) "I am primarily paid for publishing…": The narrative framing of societal responsibilities in academic life science research. Science & Engineering Ethics 26: 1569-1593

Smolka et al. 2021. From Affect to Action: Choices in Attending to Disconcertment in Interdisciplinary Collaborations. Science, Technology and Human Values 46(5): 1076–1103

Smolka, M. 2020. Generative Critique in Interdisciplinary Collaborations: From Critique in and of the Neurosciences to Socio-Technical Integration Research as a Practice of Critique in R(R)I. Nanoethics 14(1): 1-19

Snow, C.P. (1959) The Two Cultures and the Scientific Revolution. Cambridge University Press.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations,' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Social Studies of Science, 19(3), 387-420.

Thatcher J, O'Sullivan D, Mahmoudi D. (2016) Data colonialism through accumulation by dispossession: New metaphors for daily data. Environment and Planning D: Society and Space. 34(6): 990-1006.

Thoreau F (2011) On reflections and reflexivity: unpacking research dispositifs. In: Zülsdorf TB, Coenen C, Fiedeler U, Ferrari A, Milbun C, Wienroth M (eds) Quantum engagements: social reflections on nanoscience and emerging technologies. IOS Press/AKA, Heidelberg, pp 219–235

Van Oudheusden M. (2014) Where are the politics in responsible innovation? European governance, technology assessments, and beyond. Journal for Responsible Innovation 1(1):67–86

Verbeek, P-P. (2011). Moralizing Technology: understanding and designing the morality of things. University of Chicago Press.

Verbeek, P.P. (2005), What Things Do – Philosophical Reflections on Technology, Agency, and Design. Penn State: Penn State University Press

Viseu, A. 2015. Caring for nanotechnology? Being an integrated social scientist. Social Studies of Science 45(5): 642-664

Volker et al. (2023) Translating tools and indicators in territorial RRI. Frontiers in Reseach Metrics and Analytics 7:1038970 DOI: 10.3389/frma.2022.1038970

Winner, L. (1980) Do Artifacts Have Politics? Daedalus 109(1): 121-136

Wynne , B. (2011). Rationality and ritual: Participation and exclusion in nuclear decision-making. London and Washington DC: Earthscan

Zuber, N. et al. (2022) Empowered and embedded: ethics and agile processes. Humanities and Social Sciences Communications 9, 191

Zuboff, S. (2020) The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs