Data Trusts

Ethics, Architecture and Governance for Trustworthy Data Stewardship

WSI White Paper #1

February 2019

Kieron O'Hara

University of Southampton

Web Science Institute

Copyright © Kieron O'Hara 2019

The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the University of Southampton, the Web Science Institute or its Board of Directors.

This work is licensed under a Creative Commons Attribution — Non-commercial — No Derivatives Licence. To view this licence, visit (www.creativecommons.org/licenses/by-nc-nd/3.0/). For reuse or distribution, please include this copyright notice.

Web Science Institute
Building 32, Highfield Campus, University of Southampton, SO17 1BJ
wsi@soton.ac.uk

About the Author

Kieron O'Hara is an associate professor in electronics and computer science at the University of Southampton, UK. His interests are in the philosophy and politics of digital modernity, particularly the World Wide Web; key themes are trust, privacy and ethics. He is the author of several books on technology and politics, the latest of which is *The Theory and Practice of Social Machines* (Springer 2019, with Nigel Shadbolt, David De Roure and Wendy Hall). He has also written extensively on political philosophy and British politics. He is one of the leads on the UKAN Network, which disseminates best practices in data anonymisation.

About the WSI

The Web Science Institute (WSI) co-ordinates the University of Southampton's (UoS) world-leading, interdisciplinary expertise in Web Science, to tackle the most pressing global challenges facing the World Wide Web and wider society today. Research lies at its heart, positioning it as a leader in Web Science knowledge and innovation and fuelling its extensive education, training, enterprise and impact activities. The WSI is also UoS's main point of contact with The Alan Turing Institute, the UK's national institute for Data Science and AI, of which UoS is a partner university.

https://www.southampton.ac.uk/wsi/index.page

https://www.southampton.ac.uk/wsi/enterprise-and-impact/policy.page

Executive Summary

In their report on the development of the UK AI industry, Wendy Hall and Jérôme Pesenti recommend the establishment of ***data trusts***, "proven and trusted frameworks and agreements" that will "ensure exchanges [of data] are secure and mutually beneficial" by promoting trust in the use of data for AI. Hall and Pesenti leave the structure of data trusts open, and the purpose of this paper is to explore the questions of (a) what existing structures can data trusts exploit, and (b) what relationship do data trusts have to trusts as they are understood in law?

The paper defends the following thesis:

A data trust works within the law to provide ethical, architectural and governance support for trustworthy data processing

Data trusts are therefore both constraining and liberating. They **constrain**: they respect current law, so they cannot render currently illegal actions legal. They are intended to increase trust, and so they will typically act as further constraints on data processors, adding the constraints of trustworthiness to those of law. Yet they also **liberate**: if data processors are perceived as trustworthy, they will get improved access to data.

Most work on data trusts has up to now focused on gaining and supporting the trust of data subjects in data processing. However, **all actors involved in AI** – data consumers, data providers and data subjects – have trust issues which data trusts need to address. Furthermore, **it is not only personal data that creates trust issues**; the same may be true of any dataset whose release might involve an organisation risking competitive advantage.

The paper addresses four areas.

1. Trust and trustworthiness.

With regard to trust, the aims of data trusts are twofold. First, data trusts are intended to define a certain level of **trustworthy** behaviour for data science. Second, they are intended to help **align trust and trustworthiness**, so we trust all and only trustworthy actors. The appropriate form of trust is based not on rules, but on **social licence** to operate.

2. Ethics

An appropriate ethical regime will help create and support a social licence. Hence a data trust must generate a meaningful **ethical code** for its members. This will vary, depending on whose trust the data trust is intended to solicit. However, the code should constrain all who operate within it. Hence a data trust is expected to have a **membership model**, and all the members of the trust would respect the ethical code when acting within the model.

One possible example for the foundation of an ethical code is proposed in the paper: the **Anonymisation Decision-Making Framework** (ADF), proposed by UKAN.

3. Architecture

The data trust might not actually have an architecture as such – it might be merely a code of governance. However, this paper discusses one possible architecture, based on the Web Observatory developed at Southampton University, to create a **Data Trust Portal**. The architecture allows data to be **discovered** and **used**, promoting **accountability** and **transparency**, without the data leaving the hands of data controllers. A data trust is not a data store.

4. Legal status

The paper sets out the manifold reasons why a data trust cannot be a trust in a legal sense. However, it takes inspiration from the notion of a legal trust, and several instances of this are also set out. The key issue is defining the set of **beneficiaries**, and defining what their **rights** within the trust will be. Again, the appropriate set of beneficiaries will depend upon the set of agents whose trust is to be solicited by the data trust.

To conclude, data trusts could help align trust and trustworthiness via a concentration on ethics, architecture and governance, allowing data controllers to be transparent about their processing and sharing, to be held accountable for their actions, and to engage with the community whose trust is to be earned.

Introduction

In their report on the development of the AI industry for the UK government, Hall and Pesenti introduce the idea of a data trust as a means of facilitating data sharing, in order to support industry's, government's and academe's access to the data that is the raw material of AI development (Hall & Pesenti 2017). They specify that data trusts should be "proven and trusted frameworks and agreements" that supply the trust that will "ensure exchanges [of data] are secure and mutually beneficial". In the background is the unspoken assumption that the US and China have the advantage of being larger markets than the UK (Hall and Pesenti's focus), and less fragmented markets than the EU (Lee 2018). Another assumption is that data sharing is inherently risky for a number of reasons, including that sharing personal data might put the interests of data subjects at risk, exposing an organisation to a fine or to reputational damage, and that companies might lose trade secrets or competitive advantage by sharing. Hence data sharing needs a 'shove' to establish the practice, and data trusts might help to absorb at least some of the perceived risk of data sharing.

Hall and Pesenti leave open the exact nature of data trusts, and define them only functionally. Hardinges (2018), in a survey of this nascent field for the UK Open Data Institute (ODI), whose mission is to increase safe data sharing and to open up as many data stores to as much processing as is consistent with safety, found five particular interpretations:

1. A repeatable framework of terms and mechanisms.
2. A mutual organisation.
3. A legal structure.
4. A store of data.
5. Public oversight of data access.

The ODI researchers eventually narrowed down their quest to a single definition (Hardinges & Wells 2018), which they based on the notion of a literal legal trust: "a legal structure that provides independent third-party stewardship of data". A trust is a legal relationship in which an asset is run by a trustee for the benefit of a beneficiary. Even though the trustee owns the asset in law, she is not allowed to run it for her own benefit, but has a fiduciary duty to ensure that the benefits fall to the beneficiary. The idea of a data trust, then, leans on this concept from common law jurisdictions such as the UK and the US: whoever have the rights over the data must commit to administering the data for the benefit of beneficiaries, rather than for themselves. Delacroix and Lawrence (2018) argue that data trusts as Hall and Pesenti cannot be literal legal trusts.

In this paper, I will broadly endorse the ODI conception, while also agreeing with Delacroix and Lawrence, and look in detail at how we might implement something like this concept, while also in passing considering the reasons for rejecting some of the other interpretations. I will also consider what technologies and standards might already be in place to support this implementation. The key thesis of this paper is:

A data trust works within the law to provide ethical, architectural and governance support for trustworthy data processing

In particular, a data trust needs to fulfil two functions. First, it needs to be an arena in which data processing and data science can take place transparently, allowing data controllers to be held accountable. On top of this, it should also allow data scientists to interact and debate what constitutes trustworthy behaviour in their profession.

Second, the data trust also needs to be an interface between data scientists, data subjects and other stakeholders. This should allow stakeholders both to hold data scientists to account themselves, and also to inject their own views about what constitutes trustworthy behaviour by data scientists (i.e.

what they trust data scientists to do). Delacroix and Lawrence argue that "it is unclear what, if anything, such frameworks have in common with the Trust structures" that we find in English law (2018), but I will argue in the course of this paper that data trusts can take quite a lot of inspiration from, even if they cannot actually be, legal trusts.

We also should note the long list of agents who have a need for trust. Data controllers need to trust that their data will not be misused by data users. Data users need to trust that the data they get access to is of high quality and good provenance. Data subjects need to trust that data about them will not be used to harm (or even to irritate) them. And all data scientists need to trust that untrustworthy practices will be stamped out – trust in data science as a whole suffers with each Cambridge Analytica story. The data trust is not just about the trust of data subjects, but of many more. It follows that there is no 'one size fits all' data trust, but a range of models should be available, as argued, for different reasons, in (Delacroix & Lawrence 2018). The structures described in this paper are intended to be extremely flexible, in order to foster the trust of different communities, not just the data subject, unlike most previous research (Edwards 2004, Delacroix & Lawrence 2018).

One final preparatory caveat: I have already used the term 'data controller', which is a term of art from data protection law referring to the person who determines the purposes for which and the manner in which personal data is processed, i.e. exercises overall control. The trust issues that arise in data sharing are not restricted to the sharing of personal data; non-personal data can be sensitive too, if for different reasons. In this paper, I will use the term 'data controller' loosely to mean whoever exercises control over any dataset in a data trust, whether or not it is personal data, and consequently, whenever I refer to data or datasets, I make no assumption that the data is personal data unless stated explicitly. However, if I refer to the data subjects of a dataset, naturally that implies that the dataset contains personal data.

The structure of the paper is as follows. The next section looks at the notion of trust, how trust in the use of data is currently promoted, and how it could be. The following section considers some of the ethical issues, on the understanding that the regulatory background, which in the UK and EU is based around the General Data Protection Regulation, is not sufficient for maintaining trust. Next, I speculate about what kind of architecture might implement a data trust. The penultimate section examines in some detail the parallels and divergences between a trust in law and a data trust on Hall and Pesenti's and the ODI's pragmatic, practical view, and argues that a data trust can take inspiration for its structure from the legal concept of a trust, but it should and could not actually be a legal trust. Finally, a concluding section will revisit the topic of trust.

Trust and trustworthiness

Data processing is highly regulated. There are different jurisdictions across the globe, but the EU's GDPR has set high standards, and combined them with powerful punishments (fines of tens of millions of euros are possible), with the aims of making data controllers more accountable, and of helping data subjects to ensure that their preferences are respected, and that personal data held about them is accurate, proportionate and not excessive. The GDPR regime has been criticised for being too powerful, although it sets a useful international benchmark. The US regime is patchier, covering some sectors more than others, resulting in a focus on sensitivity and the potential for harm; health data, financial data, and data about children are regulated more than less problematic data.

Yet there is still something of a trust deficit around data processing, despite these regulatory regimes. While this may be surprising at first blush (and indeed at the time of writing, GDPR is relatively new and so could reassure more people once the lines of its practical operation become clearer), some reflection on the data protection regime will make it clearer why it is not well set up to support trust in this area.

To begin with, trust is a relative term – X trusts Y to do something in a particular context (O'Hara 2012). The data protection regime is set up to support one particular type of X and one particular type of action; the X in question is a data subject, and the action is the processing of personal data from which X is identifiable. This already limits the regime in two important ways. First, regulation is often, and inevitably, behind the curve of innovation. The Data Protection Directive of 1995 was painstakingly developed for a standalone database world, just as the World Wide Web came along to make linking data easier. Similarly, the GDPR of 2018 protects us against many of the excesses of the Web, just as big data came along, allowing decisions to be made about us and profiles attached to us without any input from personal data, which is anonymised or aggregated out of scope. The focus on personal data is already too weak to protect us from all the inappropriate interventions that data processing can afford.

Second, many of the trust problems that concern Hall and Pesenti (2017), and also the ODI researchers, go beyond the problems of the data subject, covering the doubts of data providers, data consumers and other stakeholders. Data protection does little for the concerns of these stakeholders.

There are also deeper reasons why even an overhauled data protection regime is not well-placed to support trust, which I will consider in the next subsection.

Rights and neoliberalism

The data protection regime combines two complementary ideological positions. In the first place, data protection is part of a rights-based approach. The individual is perceived to be in possession of certain rights, which she can use to defend herself against harm. The European Convention on Human Rights of 1953, developed in the aftermath of the horrors of Nazi Germany, included an article enshrining her rights to a private life. Data protection regimes add more detailed rights to this basic idea; the GDPR grants a right of access to data subjects to see their own personal data, as well as some rights to erase personal data held by others, rights to explanations of decisions made about them on the basis of algorithmic processes, and so on. In many cases, data processing can be consented to via a contract between subject and processor. The Charter of Fundamental Rights of the European Union of 2000 includes rights both to privacy and data protection.

Yet the original Data Protection Directive was conceived in the context of the European Single Market, and so has a dual aspect – it gave data subjects some rights to protect their privacy, and gave data controllers rights

to gain value from the data. Following it, the GDPR also protects some data sharing practices, and aim to provide a framework for data controllers to process personal data accountably in a stable and predictable environment. From this angle, the data subject is seen as the defender of her own interests in a complex marketplace. This neoliberal view of the data protection regime sits alongside other mechanisms where the onus is on the individual to understand and express her own preferences, and to ensure they are met, where possible, through her own efforts. Such mechanisms include consent regimes, which envisage data subject and data controller entering into a contract when the consent button is pressed, and personal data stores, where the data subject undertakes some administration of her own personal data. Tim Berners-Lee's recent promotion of 'personal online data stores' (pods) falls into this category.

These twin approaches of rights and neoliberalism each have several merits which I will not review here. However, neither of them is very conducive to the development of trust. There are two reasons for this, one major and one minor. The minor reason is that they focus on particular projects for processing data, and rely on the individual pushing back where she believes that she may be harmed, or at least may not benefit from, such projects. This is small scale; the individual is supposedly trying to ensure that various detailed rules are followed. Yet trust is a big picture view of the world, not a detailed vision of how people should behave. A trustor expects a trustee to look out for her interests in various, possibly unspecified, ways. The patient (at least, one without medical training) does not trust the doctor to carry out specific, detailed procedures; she trusts the doctor to make her well. The saver does not trust his accountant to put so much of his money here and so much of his money there, but rather trusts her to maximise his income or security according to his general appetite for risk, and trusts her not to benefit herself over and above the fees he pays her. Trust is not legalistic; a technical breach of the rules will be overlooked in a trusting relationship, as long as the intentions behind the breach were benign and the consequences not too terrible. Indeed, in many technical areas, the individual may not even know what her own interests are, and will trust professionals not only to defend her interests, but also to define them. Data protection, on the other hand, is a legalistic regime, giving the data subject too narrow a focus to generate trust in the way her data is handled overall.

Secondly, both the rights-based approach and neoliberalism place too much onus on the individual. The individual is to defend her rights. This is, as is frequently argued, quite a burden. Most have better things to do, and few have the expertise to do it well (Delacroix & Lawrence 2018). Even if the individual engages, she will find herself with quite a burden as she tries to deal with giant corporations under conditions of asymmetric knowledge. For example, when the data subject signs a consent form or clicks a privacy policy, she rarely understands what this actually means, and so the contract between the two parties is one-sided to say the least.

But most importantly, both the rights-based approach and neoliberalism are products of a lack of trust, assume trust is in short supply, and make trust difficult to build. The relationship between the individual and the other is deliberately set up antagonistically. In the rights arena, the individual is warned that the world is full of potential threats to her well-being, and by bad actors who will not treat her with the dignity proper to a human being, and that she therefore needs conventions and courts to protect her. Under neoliberalism, which aims to expand freedom by shrinking public space and growing the powers of private actors under market conditions, the individual is told that she must pursue her own interests, because no-one else will do it for her. Under neither of these

conditions is the individual (or the other, for that matter) incentivised to seek out the compromise or to initiate the dialogue that will enable them to bootstrap trust where it is not pre-existing.

Social licence

Ensuring that data processing is trusted needs a different approach. The operation of a technology or technocratic policy requires some kind of big picture approach to act as the locus of trust. One way of viewing this is to see data science as analogous to other kinds of technological intervention that need to be accepted by a community and other relevant stakeholders before they can operate successfully or profitably. Doctors need to be trusted by their patients (Carter et al 2015), and those drilling or mining for natural resources need to be trusted by stakeholders, particularly the local community (Gallois et al 2017), if coercion is not to be used. These technological interventions are often justified using the resources of a *profession*, such as professional codes of conduct. The profession and its resources provide the big picture crucial for trust. At the moment, data science is only beginning to develop its professional standing. There are plenty of rules – GDPR provides plenty – but they haven't solved the trust problem, and more rules will not help.

The sociologist Everett Hughes provided the valuable notions of *licence* and *mandate* (1958). Licence is 'granted' informally by society for some occupational groups to carry out activities that are part of the job, and members of those groups claim a 'mandate' to define what proper conduct looks like. This produces what Hughes called a "moral division of labour", where society and profession collaborate in "the setting of the boundaries of realms of social behaviour and the allocation and responsibility of power over them". This is a negotiation. The delicate and informal nature of the licence provides no guarantee that trust will be preserved if the professional goes too far – Carter et al describe how the highly trusted medical profession in the UK presided over the disastrous roll-out of the care.data scheme to use primary care data for medical research and other purposes (2015).

Key to the negotiation of a social licence is communication. As (O'Hara 2012) argues, trust involves aligning the trustors' and the trustees' understanding of what the trustee is committed to, which involves communicating clearly and precisely what the trustees' intentions are. If the trustors fail to understand precisely what the trustees intend to do, then their trust may be based on false assumptions, and their trust could be misplaced, despite the trustees' behaving in a perfectly trustworthy manner by their own lights. Communication requires engagement and response, and trust will be more forthcoming if the would-be trustees have a good track record for responsive practice in the past (Gallois et al 2017). Furthermore, communication needs to be a genuine dialogue, not merely the broadcasting of what from the scientific point of view are truisms expressed in jargon; engagement is required to seek a vocabulary that is meaningful to both sides of the conversation. Furthermore the trustors' attitudes towards evidence and their risk assessments also need to be understood and accommodated (O'Hara 2012). Gallois and colleagues argue that communication accommodation theory is a good frame for the necessary engagement (Giles 2016, Gallois et al 2017).

Data trusts as explorations of trustworthiness

A data trust, then, could serve the data science profession as a focus for a social licence, and a locus in which the social mandate could be negotiated. The data trust would specify a set of boundaries and responsibilities for data controllers, and give the controllers a space in which they could negotiate the social mandate for their profession. The data trust would then have a clear set of aims.

Firstly, unlike the rights approach or the neoliberal approach inherent in data protection, its starting point would be the compromise between trustor and trustee that is essential for creating trust in the first place. This involves genuine mutual communication and consultation. Trust may be hard to build – trust of data processing is all of a piece with trust of companies (or government), of global capitalism (or state power), of security and infrastructure, and so on.

Secondly, again unlike the other two approaches, the expertise of the data scientist is a central part of the picture. For example, sending the data subject a notification of where his data has been sent, and which third parties now have it in their control, whether anonymised or fully personal data, is well-meant transparency, but hardly useful to the data subject (O'Neill 2009), who not only has better things to do but who also may struggle to understand a highly complex document containing several names of companies of which he has probably not heard, performing actions, such as auctioning adverts, whose significance is unclear to him, and which may not do him any tangible harm. In the rights-based and neoliberal approaches, the data subject is on his own. With a data trust, data scientists can (and should) engage with data subjects and other stakeholders to determine what kind of treatment of data is acceptable, and the scientists themselves may well, if they present themselves sympathetically, be able to inject a good deal of their expertise into this discussion. They might then be able, if they can take their stakeholders with them in the conversation, to determine to a large extent which data processing is probably OK, and which not. Individual data subjects may not care, or be interested in engaging, but in a big data repository, enough subjects, or representative groups, may be able to feed in opinions. The data scientists should absolutely not assume, *ab initio*, that they have a monopoly of rationality, and that merely stating their case should be enough to win everyone round. Trust of expert systems is a complex matter. The data scientist needs to earn the mandate to impose and defend the standards of the profession.

Thirdly, the data trust would be a centre for data processing that could be used to hold data scientists accountable, auditing how they treat the data and who is allowed access.

Fourthly, and relatedly, the data trust would aid transparency by being inspectable and scrutable. This would allow individual data subjects to complain and intervene, as with the data protection approach. More to the point, however, this would also allow representative groups (e.g. patients' groups, or taxpayers' representatives) to monitor data use. But the real advantage of a data trust is that it would allow data scientists to be transparent and accountable to their peers. Data scientists all suffer by untrustworthy behaviour in the profession. For example, Facebook claims innocence in the case of Cambridge Analytica, but even if this is justified, it has suffered reputational damage because of its association. So have some of the political campaigns which employed Cambridge Analytica. A data trust, importantly, would provide an arena in which data scientists could clean up their own act.

Finally, a data trust might even help with determining which processing is legal. GDPR provides for a number of grounds for data processing, of which one of the most important is consent. If a data trust were well-enough known and trusted, then it might become the focus of consent. Data subjects would be asked at collection time whether they consented to the use of their data within a (specified?) data trust, for purposes consistent with the principles underlying the trust. This has the advantage of being clear and flexible, resisting the GDPR's tendency to close down big data opportunities, without succumbing to a hopeless determinism about the rise of big data. The data trust itself could also be a convenient point of contact for a

data subject who wished to withdraw consent at a later date.

The data trust would have to obey the law, naturally. However, this would not be its *raison d'être*. As we have seen, merely being legal is not sufficient to support trust. It follows from this that the data trust should be a voluntary arrangement, rather than mandated by law. If the latter, the trust could easily descend into a box-ticking exercise, as data protection often does. The point of the data trust is to signal and to demonstrate the trustworthiness of the data processing. Voluntary participation is an important part of the signal.

Put another way, legislation and regulation constrain data processing, but not sufficiently to promote widespread trust. If it would promote trust beyond that promoted by centralised regulation, the data trust should act as a further constraint on data processing, beyond what is ruled out by law. Such voluntary constraint, when credible, is a means of promoting trust. This shouldn't necessarily be seen as a cost to the data processor, however, as the result of trust may well be the creation of more opportunities for processing as a result (more collaborations, more data subjects willing to give consent, especially open-ended consent, greater supply of data under fewer formal conditions). Hence the voluntary constraints imposed by a data trust may liberate the processor to achieve more.

I have so far written mainly of trust. In fact, the key issue is the *trustworthiness* of the processing. Trust and trustworthiness are two sides of the same coin: trustworthiness is the virtue of reliably meeting one's commitments, while trust is the belief of another that the trustee is trustworthy (O'Hara 2012). Trust without trustworthiness is a severe vulnerability. Hence what is needed is a means for (a) establishing the parameters of trustworthy data science, and (b) demonstrating to would-be trustors that the data science is indeed trustworthy, so that they could be confident that their trust is warranted.

A data trust should be means to both of these ends. As an arena for data scientists to share and process data, it should enable the debates and discussions about what counts as trustworthy behaviour to take place. As an interface between data scientists and data subjects (and other stakeholders), it should enable the engagement to take place that will signal trustworthiness, and also allow the other stakeholders to help determine what constitutes trustworthiness.

Ethics

As noted earlier, there is a trust deficit around data processing despite the increasingly powerful legal regime in the EU based around the GDPR. Regulation will not, of itself, create trust, although it may be one of the means for stamping out untrustworthy behaviour; similarly for consent and contracts. As argued earlier, they simply operate at the wrong level, and in this case do not support an already existing social licence.

As well as regulation, an *ethical* regime is needed to help create that licence, so that the data scientist's actions can be judged not only legal or illegal, but also right or wrong, and ultimately that the data scientist can be judged to be virtuous or vicious. Data trusts could catalyse the development of such an ethical regime, in which the data scientist is seen as someone acting not only in her own interests, but also as someone acting in (or against) the interests of her stakeholders. The data trust would be the means of ensuring that stakeholders' interests were considered in any decisions made about processing. Of course, no data scientist should process data illegally, but the data trust could be the means for deciding whether legal data processing was in stakeholders' interests, against them, or neutral. If the processing was against their interests, then the governance structures of the trust should be sufficient to hold the data scientist to account.

Rules will not cut it; they can always be bent. Even when the letter of the law is adhered to, its spirit may not be. Rules cannot do justice to the sheer complexity of ethical life, which varies so much by context. They struggle therefore to distinguish trustworthy and untrustworthy behaviour. Trustworthiness is a virtue, and the neo-Aristotelian language of virtue ethics is helpful here.

A key notion in virtue ethics is that of *human flourishing*. The virtuous person promotes human flourishing. Happily, this phrase was used in the British Academy and the Royal Society's report on data management, an important starting point for working out the appropriate stance for ethical data science: "The promotion of human flourishing is the overarching principle that should guide the development of systems of data governance" (British Academy & Royal Society 2017).

Promoting flourishing is not something for which rules can be written; rather, this is something that must be reasoned case-by-case, using what is called *practical wisdom* which is sensitive to context (Lovibond 2002). A data scientist with such practical wisdom will look after data virtuously, not only making the right decision in any particular case, but able to plan ahead and consider other variables in her deliberations. She will be able to express her wisdom to others, and in particular to engage with stakeholders, stating her case in a way that is meaningful to them, and responding to their replies by adjusting and revising her plans if necessary. These abilities are central to practical wisdom, and also central to the creation and maintenance of trust.

There is no exact characterisation of the right ethical framework to help data scientists develop practical wisdom to promote human flourishing – 'human flourishing' itself is a (deliberately) vague term in this respect. In the rest of this section, I will consider a recent framework for data stewardship which might help provide some guidance.

Example of an ethical framework: the ADF

The Anonymisation Decision-Making Framework (ADF – Elliot et al 2016) was developed to support the complex task of anonymising data, under the legal regime of the Data Protection Directive in the EU. It was developed by the UKAN organisation, a joint venture of the Universities of Manchester and Southampton, the ODI, and the Office for National Statistics. It was adapted for the Australian data protection regime as the De-

Identification Decision-Making Framework (DDF – O'Keeffe et al 2017), and is currently under further development to bring it into line with GDPR.

It is therefore a work in progress, but the aim here is simply to show how the framework might help inform the ethical principles underlying virtuous data stewardship in a data trust. Other principles could be followed; much would depend on the context, the domain, the potential for harm, and the nature of the stakeholders whose trust was being sought. The point about the ADF is that it is a framework, not an algorithm or a set of rules or a set of boxes to tick to anonymise data; anonymisation is an art as much as a science, and the ADF is designed to reflect that. It requires, not the ability to follow rules, but rather to exercise practical wisdom in responsible data stewardship.

Let me also emphasise that the use of this example, of an anonymisation methodology, does *not* mean that all data in a data trust should be anonymised (although some of it may be). It is rather that the ADF contains principles for responsible data stewardship that may be applicable *outside* its intended sphere.

The ADF consists of three main activities, divided into subcomponents (Elliot et al 2016). Because we are not concerned with anonymisation *per se*, we do not concern ourselves here with the second activity, which contains the technical processes of disclosure risk assessment and control. We are concerned with the first activity, which is an audit of the data situation, and the third, impact management.

Data situation audit

Ethical data stewardship must involve understanding the flow of data and its ramifications. In the ADF, this involves various aspects, including understanding what use cases there are for the data, and mapping how data would flow in these cases. It also involves understanding the legal issues surrounding the data, not least the basis for processing (and if this is consent, consent for what?).

There are two particularly crucial aspects of the data situation audit. The first is understanding stakeholders' trust in the system. This is not simply whether this is high or low, but also what the stakeholders understand the data controller to be committed to, and for whom. Note that the stakeholders' understanding of the data controller's commitments may be different from the data controller's understanding. It might also take into account the warrants or reasons for stakeholders' trust.

The second concerns the idea of a data environment. The insight of the ADF is that whether data is anonymous is not a function of the data alone. Much depends on the context in which data is held. Anonymity is also not a binary; the point of anonymisation is to reduce the risk of reidentification via the data to a negligible level, not to transform the data permanently. As the context changes, so will the risk. Much therefore hangs on the context.

To express this, the ADF introduces the notion of a *data environment* as a technical term (Mackey & Elliot 2013). The data environment is characterised by four things: the *agents* who have access to the data; any *other data* to which the data can easily be linked; the *governance* of the data; and the *infrastructure* used to store it, including hardware, representation languages and cybersecurity measures. Data will typically be held, or planned to be held, in a range of data environments, all of which need to be mapped and understood by data controllers (the aggregation of the data environments is referred to in the ADF as the *data situation*).

The data environments are important within the ADF because they will help determine whether data is, or will be, anonymous in the sense that no-one could reasonably be likely to identify individuals from the dataset.

Outside of the anonymisation methodology, understanding of the data environments in which the data is held will help data controllers estimate risks to privacy or other types of well-being of the data subjects.

Note that the methodology could also easily be applied to non-personal data as well. Part of the problem of privacy in the big data era is that non-personal data can be influential in individuals' lives, for example via profiling. Or non-personal data can be combined by an intruder with other data that she holds to find out more about a target. The boundary between personal and non-personal data (or personally-identifying data from non-identifying data, in US terms) is no longer the same as the boundary between risky and safe data, even if the boundary is clear (which is doubtful).

The output of the data situation audit, then, will be a greater understanding of the context in which data is held, including the attitudes of the stakeholders, and the evidence needed to estimate the risk of an attempt to use the data for illicit purposes. The data trust can help fix much of the context of any shared or potentially sharable data, and so enable increasing precision in reasoning about the risks involved with sharing data.

Impact management

The second important aspect of the ADF which could be imported into a data trust is the plan for managing the impact of a data breach. This area of data management is often overlooked, so responses to emergencies are often *ad hoc*, opaque and improvised. The immediate instinct of an organisation is to minimise liability, which can result in slow responses and even dissembling, while messaging is cleared with lawyers. The result is an apparent shiftiness, which is easily taken as a signal of untrustworthiness. Even if the organisation has done everything it could and is not to blame for the breach, an ill-thought-out communication strategy gives an impression of a cover up, that it has something to hide. At best, it means that the organisation is focused on its own problems of liability, and not on the harms to its stakeholders.

The data trust therefore does need to have plans in place to deal with the worst. The exact details of course cannot be predicted, but it is important that a response is lined up, and the people expected to deal with it, and to communicate with stakeholders, as well as to initiate any procedures within the trust itself, should be trained and ready for their tasks.

Impact management in the ADF has three components. First, there needs to be a plan about how data sharing will be managed. Within a data trust, much of this will be standardised within the trust's governance and architecture. It will also include monitoring the new environments in which the data is held. For example, if dataset A is shared with organisation O, does O hold other datasets that will enable the inference of sensitive data? If dataset A is a database of children, does O hold a dataset B of mothers of babies, which might be combined with A to discover underage mothers in a region, far more sensitive information? If so, then the new environment for dataset A needs to be specified so that there is a strong firewall between A and B, and it would be O's responsibility to ensure that it is in place. O's new arrangements should also be transparent within the trust, so that it can be held accountable if its arrangements are inadequate.

The second component is to plan how to communicate with stakeholders, particularly in the event that something goes wrong. This involves each organisation in the data trust maintaining a line of communication with stakeholders in the data it holds. It may not need direct communication with every stakeholder, e.g. every data subject in a set of personal data. However, if the stakeholders' trust is to be maintained, each organisation

within a data trust will need to be able to keep them informed.

Finally, a plan is needed for when things go wrong. If there is a data breach, how can it be closed down quickly? Who needs to be informed, by whom, and with what messaging? If an organisation within the data trust is held accountable, how will it be disciplined? Will it be expelled? If so, how will this be managed, for instance if it has shared valuable data with other organisations in the trust.

The ethical anchor of a data trust

The shape of the data trust is becoming clear when we consider the ethical requirements. Organisations will bring data to the trust to share with each other under specified ethical conditions. Each organisation, therefore, must commit to a common set of ethical standards which will be determined by the trust itself. The commitment must be voluntary, but there must be measures which can be taken against organisations that do not live up to their commitments.

I argued above that, given that detailed rules are not very effective at engendering trust, and given that trustworthiness is a virtue, a virtue-based ethic looks appropriate. This also fits in with the idea floated by the British Academy and the Royal Society that ethical data stewardship should support human flourishing, which has been the goal of virtue ethics since Aristotle's *Nicomachean Ethics*, where it is called *eudaimonia*. We also see that rule-following or box-ticking needs to be supplemented by context-sensitive practical wisdom, or what Aristotle called *Phronesis*.

A data trust therefore needs to develop methods to support data controllers' practical wisdom, or pragmatic practices, for understanding and acting in the interests of the relevant stakeholders, in the sense of enabling them to flourish. This requirement does not determine any specific ethical code, although it seems clear that trustworthy, virtuous data stewardship should involve the virtues of *caring*, for the interests of the stakeholders, and *prudence*, the ability to discipline oneself and to manage the risks one undertakes, both in one's own interests and in the interests of those with whom we have dealings.

I have also argued that certain aspects of the ADF could usefully be repurposed to fulfil some of the caring and prudential aspects of data management. Indeed, I would claim that the ADF constitutes an approach to virtuous data stewardship in itself. Hence the ADF could be taken off the shelf as an important part of the ethical basis for a data trust.

Architecture

A data trust could simply be an arrangement of governance or a legal agreement. However, it is possible to imagine that many of the institutions or practices that would support trustworthiness within the trust could be programmed into an architecture, and reasonable to believe that this would be desirable. In this section, I will consider what some of these *desiderata* might be, and then sketch an architecture, based on an existing model, that might underlie a data trust.

The basic idea of a data trust is a virtual place where data is made available to share. Different organisations would bring data to the trust. The trust would not need to store data. We can think in terms of a membership model: different organisations would be members of the trust, which would mean that they would (i) be *either* data controllers bringing data to the trust for sharing, *or* data users wishing to share data via the trust, or both, and (ii) agree to abide by the ethical principles underlying the data trust.

Desirable properties of a data trust architecture

Many of the properties of a data trust architecture will fall out of this specification of how the trust should operate. In this section, I will set out 8 properties that would seem to be important in many if not all contexts where trustworthy data sharing needed to take place. Different conceptions of data trusts may require a different set.

1. **Discovery.** Potential users need to be able to discover the existence, properties and quality of the data in the first place.
2. **Provenance.** Potential users need to be able to assess the quality of data, by getting access to metadata about its provenance and other properties. The system within which they gain access should also be able to generate an account of the provenance of the new operations on the data.
3. **Access controls.** Data controllers need to be able to retain control over who gets access. Users need to engage with data controllers to discuss the terms and conditions for sharing. The liability for data protection breaches, therefore, remains with the data controllers where the data is personal.
4. **Access.** If appropriate, users need a mechanism to get access to the data. Access need not be unconditional, and could be mediated, or be to a limited quantity of the data, or to a redacted, anonymised or pseudonymised version.
5. **Identity management.** Data controllers need to be able to identify those attempting to get access through time.
6. **Auditing of use.** A record of uses of data needs to be generated and stored. This needs to be transparent, so that it can be checked for compliance with the law, and compliance with the ethical principles agreed by trust members.
7. **Accountability.** Ultimately, data controllers are accountable for the use of the data under their control, and the audit must enable them to be held accountable for misuse. Similarly, those receiving data and misusing it must also be held accountable.
8. **Impact.** The value, use and misuse of data also ought to be assessed via the records kept in the data trust.

A Data Trust Portal

In this section, I will sketch out an architecture which I will call a Data Trust Portal (DTP). This is not the only architecture that would fit the 8 desiderata given above, but it does fit the bill. I take inspiration here from the idea of a Web Observatory used in Web Science as a means of sharing data on and about the Web

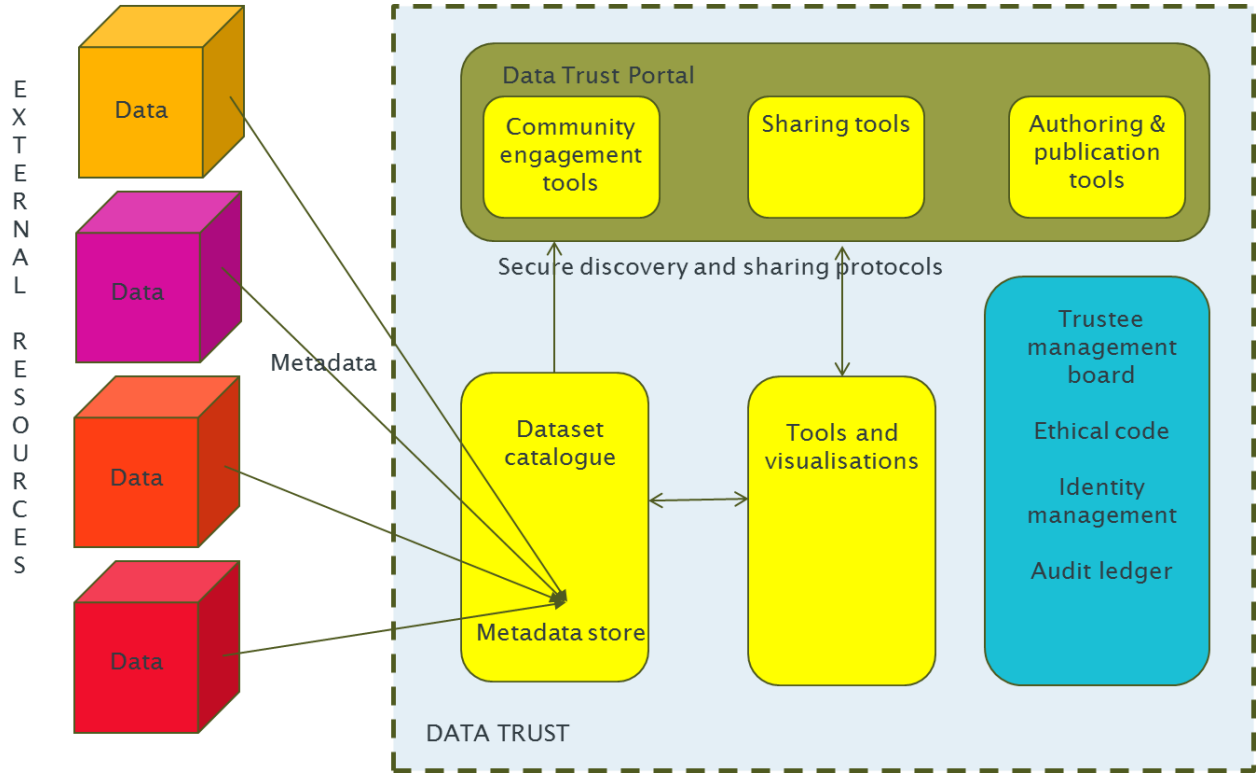

Figure 1: Architecture for a Data Trust portal

safely and ethically (Tiropanis et al 2013, Tiropanis et al 2014, Tinati et al 2015). Many of the ideas are extended or adapted for the specific needs of a data trust. The suggested DTP architecture is shown in Fig.1.

Note that the data does not get into the DTP at all; the DTP is *not* a data store, nor a distributed database. The data is held by the original data controllers, in their own controlled environments, and they retain their data protection responsibilities if the data is personal data. They do not transfer the data (unless they wish to), and remain in ultimate control of access. Different datasets can be treated differently. If, for example, they would only allow data users to access the data on specific premises, e.g. a safe haven with no Internet access, then that is their decision. If they are happy for a copy of the data to be transferred to a user, then they can design the arrangements for this, including creating their own terms and conditions, and can determine any rights for the data users to transfer the data to a third party. Data sharing arrangements can be automated, and the automation can apply to all, or only some, of the datasets. Access to the data need not be free; nothing in this arrangement precludes charging for access. Sharing data on a data trust should not entail surrendering control. In this way, data controllers' trust of the sharing process should be maintained, because they only relinquish control on their own terms (this meets property 3 above). Note also that individuals (who might also be beneficiaries) could bring their own data (e.g. from wearable wellbeing devices) to the data trust as well, if they were willing and able to abide by its ethical terms. They could share their own data with other data controllers, or even, if they had the expertise, ask for access to other datasets to make their own data more meaningful.

They post *metadata* into the DTP, into a metadata store; this could be any metadata felt useful, but should include provenance, or provenance summaries (meeting property 2), and also basic information about size, content, representational schema, etc. The metadata are used to build a searchable

dataset catalogue, of all the datasets available in the DTP (this meets property 1). The data trust need not only deal in raw data, but could also share useful analysis tools and visualisations of the data, either created by the data controllers themselves, or by data users.

A DTP will need a relatively centralised management to ensure accountability, although it may adopt a peer-to-peer structure if peers were trusted to hold each other accountable. They would each have incentives to do this, since one untrustworthy member of a data trust could taint all the others. The management component would include managing the identities of those supplying and those consuming data (property 5), creating and maintaining the ethical code, and providing an audit trail of all data use via the trust (properties 6, 7 and 8). The portal itself would be a platform, where data controllers and users are enabled to meet to work out their arrangements; the data consumer will find the data he is interested in in the catalogue, and then approach the data controller via sharing protocols to negotiate the terms upon which he will be allowed to share the data (property 4). He may, of course, be refused access at any time, perhaps because the data are so sensitive that only certain data users would be allowed access, or perhaps because the conditions placed on the data sharing are so stringent that the costs outweigh the benefits of access.

The Web Observatory which inspires this architecture was conceived as a potential network of observatories (Tiropanis et al 2014). This would *not* hold with a DTP; in order to maintain the ethical standards set out by the data trust, linking with other data trusts would of necessity involve ensuring that standards were and remained compatible and equally high. Much would depend on the specific architecture, and of the make-up of the trust. For instance, a public service DTP run by a city partnership to share data about that city might link to a similar DTP run by another city, allowing the sharing of data, under controlled conditions, between service providers in the two cities.

In general, the trust problems of data sharing could be addressed gradually by this structure; a data controller could advertise data, but only share it under stringent conditions (or not at all) until he was satisfied that the data trust was promoting trustworthy behaviour. As he became more convinced, he could gradually increase his commitment to sharing within the trust, if he was comfortable doing that.

Legal status

For reasons to be discussed in this section, it is probably too complex a project to make a data trust a literal trust, in the sense of the 3-party fiduciary arrangement that developed in English common law. In general terms, this is partly because the proposed arrangement in the data trust differs from the property arrangements typical of a trust, and partly because a trust is a development of common law, and is not always found in civil law jurisdictions (Penner 2016, 52ff.). However, the notion of a trust, in which property is owned and managed by a trustee for the benefit of a beneficiary, could still inspire the ideas inherent in a data trust.

Appropriately, trusts emerged from the medieval Court of Chancery, which existed alongside courts of law to ensure equity, that is, to provide remedies when the strict operation of law produced injustice. Equity is therefore, in its origins at least, reflective of ethical considerations rather than legal ones; it did not rest on how the law stood, but on how people should act 'in good conscience'. We can see a data trust as playing a similar sort of role – expressing how data controllers should behave in good conscience, rather than merely working out what is legal for them to do.

It is worth pointing out that trusts can be voluntary, or established by law (TABOLs – Trusts Arising By Operation of Law). I have argued above that participation in a trust should be voluntary, and so the law should not determine that a trust has to be set up. The data trusts I describe here are analogous to express trusts, that is, they are intentionally set up for a purpose (Penner 2016, 16, Delacroix & Lawrence 2018). There is also no central register of trusts; (Hall & Pesenti 2017) argue that a Data Trusts Support Organisation should be set up. This might provide a register of data trusts, even if an incomplete one, which would enable their discovery, and the dissemination of experience and best practice (i.e. the development of professional standards).

We might describe a TABOL as a top-down type of trust, where law mandates the creation of a certain type of structure. Others have described a bottom-up style, where data subjects would compel their data to be managed by trustees, and would set the terms of its management (Delacroix & Lawrence 2018). The proposal of (Hall & Pesenti 2017), explored in this paper, is rather a middle-out style, where the data controllers are prime movers, wanting to maintain warranted trust without losing control. I would argue that the top-down approach would require some legislation in a world where the full effects of GDPR are not yet known, which would be not only unlikely but positively unwise. The bottom-up approach, as with many others such as personal data stores and indeed the data protection regime as a whole (see above), requires a somewhat proactive attitude from data subjects; it is not impossible to imagine, but would undoubtedly place a burden on data subjects however willing a cohort of trustees can be mustered (it is noted as a 'challenge' by Delacroix and Lawrence). The proposal of (Edwards 2004) that a data trust is created whenever data subjects share personal data with data collectors is the extreme example of a bottom-up data trust, and of course in such case the trusts must be 'implied' rather than express (Delacroix & Lawrence 2018). Apart from the administrative difficulties this complexity would produce, it also misses the point that, in our age of aggregation, anonymisation and profiling, it is not only personal data which could cause problems for individuals. The middle-out approach has not been explored in detail, and has many pragmatic points to commend it as a 'good enough' solution to a social problem that does not concern everyone.

A trust has three specific roles – the *settlor*, who creates the trust, writes its terms, and disposes of the property (Penner 2016, 25);

the *trustee*, who owns and manages the property; and the *beneficiary*, who receives the benefits of the property. In the case of a data trust, the settlor is the person or group who sets up the trust and defines its remit. The trustees are the data controllers who remain in charge of the data, as can be seen in Fig.1. That leaves the beneficiaries.

Who are the beneficiaries?

There are many candidates to be the potential beneficiaries of the data trust. Much will depend on the purpose of the trust, as defined by the settlor, and on whose trust is being solicited. Each different potential set of beneficiaries will demand different principles and different structures. Potential beneficiaries include:

- Data subjects.
- The general public.
- A particular population. For example, a data trust run by service providers for a city or a region might specify the population of that region as the beneficiaries of the trust.
- Data consumers. For example, social scientists may wish to gain access to sensitive data in such a way as to retain their ethical credentials and to not alienate the population they wish to study. Or it may be that the apparatus of the data trust would give data consumers confidence in the provenance of the data in the trust, as expressed by the metadata.
- Data providers. Those making data available through the trust may want to ensure that the data they share isn't misused, or doesn't give a competitive advantage to those with whom they share it.
- Customers or clients. An organisation may wish to join a data trust as part of its reputation management. If that organisation has a poor reputation for misusing data, then it might signal to its customers that its practices have changed by joining a data trust.

Not all these beneficiaries can be pleased all at once. The purpose of the data trust should realistically be to benefit one or two of these classes of beneficiary. The rules and ethical principles of the trust should be tailored to create the optimal signals of trustworthiness to those classes. Hence a data trust designed to create trust among data providers may look very different from a data trust designed to promote trust among data subjects. And it may be that some individuals might contest a definition of 'beneficiary', for example if a 'local' scheme is seen to benefit companies or outsiders not thought of as local by the community itself (Gallois et al 2017, 51). The concept of a data trust to promote trust should not be oversold (cf. Gallois et al 2017, 51). However, conversely, just because a data trust is aimed at a particular class of beneficiary, that does not mean that it cannot also gain the trust of other communities. In general, one would hope that trustworthy data stewardship would raise the level of trust all round.

Note once more that, depending on whose trust is being solicited, the data trust may not always deal in personal data. If – for example – the data trust was intended to share data in the commercial sector to enable firms to keep their data confidential, yet still to collaborate on large-scale problems, such as developing low-carbon systems across sectors, then the data may be non-personal. We still refer to 'data controllers', but only by extension – this is not the legal definition which is only relative to personal data.

How could a beneficiary benefit?

A trust is run for the benefit of the beneficiaries (Penner 2016, 21-23). However, this should not directly be the case of a data trust. The data will be shared or processed for the direct benefit of the sharers or processors. In a standard property trust, the trustee cannot run the property for her own benefit,

even if she and the beneficiary share the benefits. In contrast, a data trust is supposed to benefit those donating data to it (otherwise why would they take part at all?) even while the beneficiaries also benefit – see above – if only indirectly. In this section, I will speculate on some of the potential benefits, suggesting issues that data trust settlors should consider when drawing up terms and conditions.

It might be thought that a potential benefit for the beneficiaries is to get access to their data, as many advocates have argued in recent years. However, this is unlikely to be the case. In a traditional trust structure, the beneficiary has no rights to the property, only to the benefits from the property. So, for instance, if a trustee holds a house in trust for a beneficiary, the income, from rents for example, goes to the beneficiary. However, the beneficiary has no rights to use the house, so the trustee can sue the beneficiary for trespass if the latter enters the house without the trustee's permission (Penner 2016, 18, 53).

A data trust might be set up deliberately to provide data subjects with access to 'their' data, but it need not be. The data could remain confidential and only shared within the trust; nothing about a data trust structure implies that the rights to access to the data should be extended beyond the current rights holders. On the other hand, a data subject could put her own data (e.g. from her own wearable devices) into the trust and she could enter as a trustee as well as a beneficiary, as noted earlier.

Furthermore, unlike the benefits of at least some trusts, the beneficiaries cannot sell or transfer the benefits onto third parties, unless there is express provision for this in the data trust. If the beneficiary has that status because of a special relationship with the data or the data controllers (e.g. that she is a data subject, or that she is a resident in a particular city or region), then that is the qualifying factor and she cannot sell on the benefits, which are anyway likely to be indirect.

That leaves open the question as to whether the data controllers could sell the data, or access to the data, to third parties outside the trust, and whether, if so, some or all of the income received should go to the beneficiaries. That again will depend on the terms of the data trust, but if at least some of these tangible benefits do not go to the beneficiaries, one would wonder what the data trust was meant to achieve and exactly how it was supposed to engender trust.

The settlor of a trust does not enforce its terms; in law that is the job of the beneficiaries themselves (Penner 2016, 25). The main powers with respect to beneficiaries' rights are to be able to complain about the behaviour of data controllers in the trust, and to seek remedies. In a legal trust, beneficiaries can sue a trustee for breach of trust if they feel the latter is not acting in their interests. How could this principle transfer to the context of a data trust? The powers could take one of two forms. It may be that beneficiaries could demand that the data from which they benefit should be used in a particular way. Or alternatively, they could be given rights to challenge any actual use of the data, without any extra ability to be proactive. Since a data trust would normally preserve the arms-length relationship between the beneficiaries and the data, the latter would presumably be more common. I have already argued that engagement with beneficiaries is an important potential function for data trusts; this, if implemented, would be a formalisation of that engagement.

This is inversely connected with the powers that the trust gives to the trustees. Trusts can usually do one or more of three things. They can impose a fixed duty on the trustee to do something specific benefiting the beneficiaries, or they can impose a duty to achieve some outcome that benefits the beneficiaries, while leaving it up to the trustee

to decide how to implement it, or they can give the trustee a right to do something that she is under no obligation to do (Penner 2016, 67ff). A data trust is likely to do one or both of the last two of these things, demanding that certain benefits go to the beneficiaries, or that certain costs do not, while leaving data controllers still in control of the data processing. The extent of those rights and duties will be related to the extent of the rights and privileges of the beneficiaries.

Can a data trust be a trust?

A legal trust is the inspiration for a data trust. However, data trusts are not trusts, without some clever crafting of its terms anyway (Delacroix & Lawrence 2018 would agree, I think, with this assertion about data trusts as I have described them, although they argue that the bottom-up trusts they advocate *could* be genuine trusts). As noted, the settlor (who need not be an individual, but may be a committee of all the relevant data controllers) must create the terms for membership of the trust, deciding questions such as what the ethical principles should be, who the beneficiaries are, what rights they have, what rights the data controllers have, what happens if a data controller goes bankrupt or the organisation fails, how controllers withdraw from the trust, whether controllers can process or share their data outside the trust, and so on. There are many templates from trust law about how to set these things up, but there are various reasons why data trusts would not behave as most ordinary trusts do.

First of all, we should note the reason given in the previous subsection, that data trusts are intended to benefit trustees (i.e. data controllers) directly, and may benefit beneficiaries only indirectly. Indeed, the trustees/data controllers in a data trust would hope to benefit twice over – once through the processing of the data, and again through the maintenance of trust of the beneficiaries. That may mean restrictions on what can be done with the data (e.g. perhaps it can't be sold to third parties), depending on the principles of the data trust, which may mean that the benefits of the data to the data controller cannot be maximised as they could be outside the trust. However, this kind of self-denial is exactly what is supposed to foster trust of the beneficiaries in the data controller, and is therefore the whole point of being in the data trust.

Delacroix and Lawrence (2018) argue that "a fiduciary obligation towards data subjects is incompatible with the data controllers' responsibility towards shareholders", and indeed that this is "the only logical conclusion" about the potential for conflict of interest here. We should begin by noting that this, if true, is only true of private sector for-profit data controllers, and even then only if we assume that the data controllers' fiduciary duty to shareholders totally outranks their fiduciary duty to data subjects and other stakeholders. However, even if we concentrate on the private sector case under that strict ordering of fiduciary duties, the point of being in a data trust is to increase trust in the handling of data. This could be argued to be in the interests of even the most rapacious shareholder in three ways. Firstly, trust in a company is an aspect of goodwill, one of its intangible assets. Reputation damage can cause serious financial problems for a company; Cambridge Analytica went out of business within two months of its scandalous data handling practices being reported in the media. Secondly, building trusting relations can help long-term profitability, even at the cost of short-term gain (this is the sort of puzzle often explored in game theory, for example with the prisoner's dilemma). The data trust sketched here could be the focus of a good deal of reciprocal behaviour with long-terms benefits over and above any short-term opportunity costs. Thirdly, recall that in the proposal sketched here, it is not necessarily the data subjects whose trust is being sought (this argument will not therefore concern Delacroix

& Lawrence 2018, who do focus on the data subject). The data trust sketched here is flexible enough to enable companies to develop robust relationships with all kinds of individuals and organisations, from data subjects through to those sharing data through even to regulators. Clearly this must be compatible with long-term profitability.

The second reason why data trusts are not congruent with the model of legal trusts, also noted earlier, is that trusts seem to flourish more in the common law world than in the world of civil law, partly because civil law jurisdictions tend to have a more binary view of property. Some civil law jurisdictions have trusts, including Quebec and Scotland (Penner 2016, 54-58), but not all, so if the trust has international pretensions, then it would need to be able to translate its terms into possibly unsympathetic legal regimes. If we simply take the idea of a trust as an inspiration rather than a strict code, this is less of an issue.

Thirdly, the data trust is a voluntary agreement with a specific purpose of supporting trustworthy behaviour. To that extent, it is not a permanent settlement of property, it is an agreement to conform to specific behavioural and ethical principles. It is time-limited, and it will always be possible for those donating data to withdraw the data if the data trust doesn't meet their purposes.

Fourthly, the point of a trust is to develop and support trustworthy behaviour and therefore create warranted trust. Independent oversight may be useful, but not in all cases. In fact, it is quite plausible that in many cases, especially when data controllers are already trusted and merely wish to maintain existing trust, that the settlors, the data controllers and the trustees are the same people or organisations. Under such an arrangement, for example, it would be possible to audit data use with a permissioned distributed ledger where the peers are the trustees/data controllers.

It follows from all this that a data trust would not be a literal trust, falling under the law of equity. Rather, data trusts take legal trusts as inspiration for a certain type of hands-off arrangement involving fiduciary duties (Penner 2016, 22ff., Delacroix & Lawrence 2018). The key point in any data trust is to define, as part of its ethical principles, the nature of the fiduciary duty of the trustees toward the beneficiaries, and to hold trustees to account against it. The fiduciary duty could be expressed, for example, in the terms of the ADF.

Trust revisited

To conclude, the purpose of a data trust is to define trustworthy and ethical data stewardship, and disseminate best practice. The aim is *not* to increase trust, which many have claimed as an imperative. The aim, rather, is to align trust and trustworthiness, so that we trust trustworthy agents and do not trust untrustworthy ones, and conversely make it so that trustworthy agents are more likely to be trusted, and untrustworthy agents less likely to be trusted. In other words, the aim is to support *warranted* trust.

A data trust is not a mechanism for producing trust. Trust cannot be magicked out of nowhere, the trustor has to be persuaded of the trustworthiness of the trustee (O'Hara 2012). Therefore the trustee needs to handle data in a trustworthy way, to communicate his actions transparently to the trustors, and to be held accountable for those actions. Existing trust in an organisation, for example the UK National Health Service, or a city council, can be leveraged to bootstrap trust, but even in that case trust still has to be painstakingly maintained, as was discovered in the care.data fiasco (Carter et al 2015).

All the would-be trustee can do is to behave in a trustworthy manner, and engage with trustors to understand their views and to communicate his own. The trustee must not make wild promises, or say what the trustors want to hear – rather he needs to manage expectations and only make credible commitments. Although my approach differs from that of Delacroix and Lawrence, I certainly agree with their statement that "a successful data Trust will be one whose constitutional terms better encapsulate the aspirations of a large part of the population" (2018).

To conclude, data trusts could help align trust and trustworthiness via a concentration on ethics, architecture and governance, allowing data controllers to be transparent about their processing and sharing, to be held accountable for their actions, and to engage with the community whose trust is to be earned.[*]

[*] Thanks to Les Carr for comments on the paper, and to audiences at several events and meetings for tough questioning and kicking the tyres.

References

British Academy & Royal Society (2017). *Data Management and Use: Governance in the 21st Century*, London: British Academy & Royal Society, https://royalsociety.org/topics-policy/projects/data-governance/.

Pam Carter, Graeme T. Laurie & Mary Dixon-Woods (2015). 'The social licence for research: why care.data ran into trouble', *Journal of Medical Ethics*, 41(5), 404-409, https://doi.org/10.1136/medethics-2014-102374.

Sylvie Delacroix & Neil D. Lawrence (2018). *Disturbing the 'One Size Fits All', Feudal Approach to Data Governance: Bottom-Up Data Trusts*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3265315.

Lilian Edwards (2004). 'Reconstructing consumer privacy protection on-line: a modest proposal', *International Review of Law, Computers and Technology*, 18(3), 313-344, https://doi.org/10.1080/13600860420002767 62.

Mark Elliot, Elaine Mackey, Kieron O'Hara & Caroline Tudor (2016). *The Anonymisation Decision-Making Framework*, Manchester: UKAN.

Cindy Gallois, Peta Ashworth, Joan Leach & Kieren Moffat (2017). 'The language of science and social licence to operate', *Journal of Language and Social Psychology*, 36(1), 45-60, https://doi.org/10.1177/0261927X16663254.

Howard Giles (ed.) (2016). *Communication Accommodation Theory: Negotiating Personal Relationships and Social Identities Across Contexts*, Cambridge: Cambridge University Press.

Wendy Hall & Jérôme Pesenti (2017). *Growing the Artificial Intelligence Industry in the UK*, London: Department for Digital, Culture, Media & Sport and Department for Business, Energy & Industrial Strategy, https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk.

Jack Hardinges (2018). 'What is a data trust?' *Open Data Institute blog*, https://theodi.org/article/what-is-a-data-trust.

Jack Hardinges & Peter Wells (2018). 'Defining a "data trust"', *Open Data Institute blog*, https://theodi.org/article/defining-a-data-trust/.

Everett Cherrington Hughes (1958). 'License and mandate' in *Men and Their Work*, Glencoe, IL: Free Press, 78-88.

Kai-Fu Lee (2018). *AI Superpowers: China, Silicon Valley and the New World Order*, New York: Houghton Mifflin Harcourt.

Sabina Lovibond (2002). *Ethical Formation: Practical Reason and the Socially Constituted Subject*, Cambridge MA: Harvard University Press.

Elaine Mackey & Mark Elliot (2013). 'Understanding the data environment', *XRDS*, 20(1), 36-39.

Kieron O'Hara (2012). *A General Definition of Trust*, https://eprints.soton.ac.uk/341800/.

Christine M. O'Keeffe, Stephanie Otorepec, Mark Elliot, Elaine Mackey & Kieron O'Hara (2017). *The De-Identification Decision-Making Framework*, Canberra: CSIRO.

Onora O'Neill (2009). 'Ethics for communication?' *European Journal of Philosophy*, 17(2), 167-180, https://doi.org/10.1111/j.1468-0378.2009.00346.x.

J.E. Penner (2016). *The Law of Trusts*, 10th ed., Oxford: Oxford University Press.

Ramine Tinati, Xin Wang, Thanassis Tiropanis & Wendy Hall (2015). 'Building a real-time Web Observatory', *IEEE Internet Computing*, 19(6), 36-45, https://doi.org/10.1109/MIC.2015.94.

Thanassis Tiropanis, Wendy Hall, James Hendler & Christian de Larrinaga (2014). 'The Web Observatory: a middle layer for broad data', *Big Data*, 2(3), https://doi.org/10.1089/big.2014.0035.

Thanassis Tiropanis, Wendy Hall, Nigel Shadbolt, David De Roure, Noshir Contractor & James A. Hendler (2013). 'The Web Science observatory', *IEEE Intelligent Systems*, 28(2), 100-104, https://doi.org/10.1109/MIS.2013.50.

Sarah Washington (2006). *Equity*, 2nd ed., Oxford: Oxford University Press.