webscience TRUST

Privacy, Privacy-Enhancing Technologies & the Individual

WST White Paper #1

Kieron O'Hara

Privacy, Privacy-Enhancing Technologies & the Individual

WST White Paper #1

Kieron O'Hara

Emeritus Fellow, Electronics and Computer Science, University of Southampton

kmoh@soton.ac.uk

ORCID https://orcid.org/0000-0002-9051-4456

22.1.21, revised 13.2.22

About the Author

Kieron O'Hara is an Emeritus Fellow in Electronics and Computer Science (ECS) 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 *Four Internets* (Oxford University Press 2021, with 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 WST

The Web Science Trust (WST) is a charity promoting the understanding of the Web, through education and research in the discipline of Web Science. It co-ordinates the Web Science Trust Network (WSTNet) of leading Web Science laboratories from around the world.

The Trust engages in both academic and public outreach through the ACM Web Science Conference, the Brave Conversations programme and sponsored White Papers.

http://webscience.org
info@webscience.org

Preface

Law has granted individuals some rights over the use of data about them, but data protection rights have not redressed the balance between the individual and the tech giants. A number of approaches aim to augment personal rights to allow individuals to police their own information space, facilitating informational self-determination. This reports reviews this approach to privacy protection, explaining how controls have generally been conceived either as the use of technology to aid individuals in this policing task, or the creation of further legal instruments to augment their powers. It focuses on two recent attempts to secure or support data protection rights, one using technology and the other the law. The former is called Solid, a decentralised platform for linked data, while the latter is a novel application of trust law to develop data trusts in which individuals' data is managed by a trustee with the individuals as beneficiaries. The report argues that structural impediments make it hard for thriving, diverse ecosystems of Solid apps or data trusts to achieve critical mass – a problem that has traditionally haunted this empowering approach.

Key words: privacy, Solid, data trusts, privacy-enhancing technologies, PETs, data protection

Introduction

Privacy has been a hot topic of debate and technological development for many years, particularly with respect to the processing of data[1] about individuals by digital technology, and the use of data, whether personal[2] or otherwise, to interfere in individuals' private lives (make decisions about them). A number of approaches to protecting privacy have been canvassed during this time, none of which has made the perception of diminishing privacy go away. Conversely, a number of commentators claim that the 'problem' is not a problem at all, a matter of taste at best, in which the harms committed are minimal to non-existent.

I will focus on privacy concerns around the digital economy, whose size is extraordinary (about $1.4 trillion of the value of Alphabet and Facebook alone is down to their exploitation of data about their users – Economist 2020). This size has led to parallel (and not unrelated) concerns about the corruption of culture by an advertising-driven business model, leading in particular to the degradation of public space and democratic politics. Some also fulminate about the size of firms that exploit network effects, which enables them to stifle competition, and thereby suppress innovation (although the tech giants do seem able to innovate effectively nonetheless).

---

[1] In this paper, 'data' will be used as a singular mass noun, rather than as a plural. This usage is defended in a WST blog https://webscience.org/data-are-or-data-is-a-pedant-writes/. I will also use the term 'data' to cover both data, as processed by computers on its syntactic properties, and information, as processed by human systems with semantics. In some contexts, this is an important distinction, but not here (O'Hara & Hall 2021, 12-18).

[2] Personal data is a legal concept from data protection law, meaning data from which an individual may be identified. It is a central concept in the EU's General Data Protection Regulation (2018). The US equivalent is Personally Identifying Information (PII).

As a result, these firms have been brought increasingly under the eye of trustbusters in the United States, China, the EU and elsewhere. These more political issues are not my focus here (see O'Hara & Hall 2021 for more discussion).

Law, particularly in the EU, has focused on granting individuals some rights over the use of data about them. However, data protection rights have not redressed the balance between the individual and the tech giants. A number of approaches have sought to augment these personal rights to allow individuals to police their own information space, facilitating informational self-determination with technological tools to help enforce rights, or even go beyond them.

data protection rights have not redressed the balance between the individual and the tech giants

In this report, I will review this approach to privacy protection, focusing on two recent attempts to secure or support data protection rights, one using technology and the other the law. The former is called Solid, a decentralised platform for linked data, while the latter is a novel application of trust law to develop so-called data trusts to manage individuals' data. The report has four substantive sections. The first will argue the case that privacy breaches are harmful and therefore need addressing, and the second will then review privacy-enhancing technologies. The third section considers Solid, and the fourth data trusts, before a short discussion rounds things off.

The harms of privacy breaches

One particular problem with privacy is that the costs and benefits of maintaining it are rather skewed. The benefits of surrendering it are often clear and tangible, while the costs are less tangible, and are usually in terms of intangible risk over a longer term. So, for instance, someone who wishes to have access to a particular service now might give up some of their personal data for the immediate and evident benefit of using the service, while ill-effects, whatever they are, may or may not emerge, and if they do emerge may only do so after months or even years. Such may be the distance in time that the data subject may not connect the effects with the initial privacy sacrifice. Or the ill-effects may be the result of the cumulative sacrifice of personal data over several interactions, whereas each individual sacrifice had a negligible effect on its own.

To give one recent example, the videoconferencing application Zoom was one of the star performers of the COVID-19 pandemic, keeping families in touch through remarkably difficult and often traumatic times. In the limelight its privacy protections were quickly identified as seriously flawed, although it speedily moved to improve them. However, it provided an important service that people wanted and needed, that was easy to use for a large audience that was not necessarily familiar with videoconferencing. The security problems proved to be neither here nor there to this cohort. It may be that security proves to be a vulnerability going forward with a more experienced user base with more choices, but in the shorter term, it did not stop Zoom becoming a more valuable firm than ExxonMobil.

We should distinguish between authorised/consented breaches of privacy and unauthorised ones (hacks). Someone who fills in an online form with personal data clearly consents at some level, even if that consent is uninformed. We should treat this differently to a criminal hack, where data is taken without even nominal consent. There is an intermediate case where data is processed without consent.

Such processing can be more accountable if the data subject is vigilant. In this paper, I will not look at hacking and issues of cybersecurity, but will focus on people's own management of their data, whether they keep the data themselves, or give custody to others. The purpose of such management should be to ensure that processing of the data is consented to, or alternatively has appropriate grounds. One hypothesis in this area is that one can, or should be able to, realise one's privacy preferences with proactive management of one's personal data. One therefore needs the tools to do this.

one can, or should be able to, realise one's privacy preferences with proactive management of one's personal data

The online economy has for many years relied on a model of putting services online, for which data about the service receiver is required, and this trend has been accelerated by the COVID-19 pandemic. In the Western democracies, businesses have created value both in the accumulation of data, and the construction of large networks of customers, and have been able to reduce (very often to zero) the price of the services they offer in return for access to customers' data, a model called by Shoshana Zuboff 'surveillance capitalism' (Zuboff 2019).

The data tsunami characteristic of recent times has been driven in particular by the widespread adoption of smartphones and social media. Smartphones have the edge in the unmanaged or unanticipated creation of data by users, in that they are unremarked media for a number of interactions. Even when they not in direct use they are generally carried around unnoticed, so are able to provide, for example, location data and data about modes of travel (via signals from the accelerometer).

Although the possibility of misuse of location data is likely to be overstated (particularly as it quickly becomes outdated), nevertheless it is still extremely revealing about a person's religion, eating habits, address, workplace, sex life, and even political orientation, once the data is matched against timelines, news stories, place-of-interest gazetteers, credit card data, and so on (Fritsch 2008). Social media, on the other hand, while arguably more revealing (and requiring less inference), do at least require conscious decisions to post information much of the time.

Privacy is a preference – some people care about it far more than others. Some are happy to place many sensitive details, or even their entire lives, online. Others prefer not to be under scrutiny. However, the shift to surveillance capitalism in the West, and related models elsewhere (particularly China), mean that many more services are migrating online, automatically diminishing privacy. For instance, reading a book or listening to a CD require no intermediary between the reader/listener and the entertainment, whereas the use of e-readers or streaming services generate large amounts of data for the providers – not only what entertainment one is consuming, but also how often, which bits were skipped, which repeated, what was consumed next, and so on. This data enables the providers to tailor recommendations accurately to their customers, improving their services, but at the same time facilitates two kinds of invasion of privacy: the increase in (potentially revealing or sensitive) information about data subjects, and also the reduction in their decisional privacy, as their choices are anticipated or made for them.

How damaging could this be? Well, as the cliché has it, data is the 'new oil', the basis for a whole resource economy. The data about you is of great interest to service providers; Snapchat awards itself access not only to whatever you post, but your entire media library, for instance. Furthermore, if the privacy policy allows it (it normally does), the data could be sold on to a third party, providing them with a ready-made personal profile of you, so you can't necessarily assume that only the people you are transacting with directly have access to it.

And that third party might sell it on, and so on, resulting in enormous complexity (Manjoo 2019). Tracking third party use of personal data is a very hard problem.

data enables the providers to tailor recommendations accurately to their customers .. but at the same time facilitates two kinds of invasion of privacy

Perhaps all that happens is that companies get better at selling you stuff you don't really want, as they learn more about you. OK, but knowledge about you makes it easier to scam you, for example by making phishing emails more convincing (they might be made to look very much as if they come from your own bank/e-commerce supplier/employer). It also may facilitate crime if you inadvertently let crooks know you are not at home (http://pleaserobme.com/ drew a lot of attention to this problem some years ago), and of course minors may make themselves vulnerable to exploitation or grooming by revealing too much.

Even if the personal data held about you is held in good faith, it could still be hacked by the increasingly capable groups operating on the Internet, some of which are sponsored by national governments (according to one recent survey, 84% of companies polled had had some sort of a data breach in 2018 – Lalor-Harbord 2019). Your smart gadgets, which rely on Internet connection, might provide an attack vector for your home (in a famous incident, a smart fish tank was used to gain access to a casino's computer systems – Williams-Grut 2018). One representative survey, for instance, found that 41% of Americans had noticed fraudulent uses of their credit cards, 16% had had their email accounts taken over, 13% had their social media accounts hacked, and 14% had someone attempt to take loans out in their names (Smith 2017). Very few would notice if their computer was recruited into a botnet used in a distributed denial of service attack.

The data that is held and used legitimately could still be used against your interests. Insurance companies might decrease – or increase – your premiums if it knew certain things about you. If your browsing or purchasing habits lead to a particular set of adverts or recommendations being sent to you, maybe that would bias your future in directions set by your past behaviour – and hinder your attempts to grow and develop psychologically. Smart dolls are able to converse with children – perhaps you've checked what they say, but what changes when the software is upgraded? Can you be sure what conversations they are having with your children? Perhaps they are teaching children values you would rather be under your control. Perhaps they are extracting information from your children that you would rather they did not have. Or even just giving your child an accent or vocabulary you didn't like.

Not only that, but individuals may find themselves in a position to use material against you. Could your browsing habits be a matter of concern if they leaked? What about those juvenile posts when you made off-colour jokes that may have been socially acceptable then, but certainly aren't now? Did you post that hilarious picture of you drunk at a party? Or did someone else post it, maybe even without your permission? If a legal case involving you came down to the question of whether you have a drink problem, such a photo could be used – however unreasonably – as evidence that you do. The more pictures of you out there, the more likely that this embarrassing photo could be automatically tagged using matches made by face recognition technology. And potential future employers (or spouses, or in-laws) might then be able to find it and revise their opinion of you (possibly upwards, more likely down). Examples abound of people's inability to shake off past misdemeanours. And these things hang around – might you spoil your child's nascent political career if photos of you lying drunk on the floor emerged during a campaign?

Finally, there is the chilling effect of surveillance. People behave differently, less autonomously, when they are under surveillance. Maybe they behave better, but maybe they forego opportunities. In any case, is that kind of control what we sign up for when we go online? Don't we go online to express ourselves, not to be oppressed by the ambient infrastructure? Ultimately, according to virtually all human rights conventions and treaties, we have a right to a private life, and the contravention of that right is surely, all things being equal, a harm?

These issues are all couched in terms of risk rather than tangible harm. One leaves oneself open to risk, consistent with what has been called a *risk society* (Beck 1992). In risk society, systems are developed that create the possibilities of environmental, economic or social harm, but responsibility for such harms is not accepted by the system developers. Instead, it is outsourced to individuals, in the form of risks for them to manage. Rhetorically, the acceptance of risk is often referred to as *empowerment*.

Controls for the individual

A lack of privacy can and does lead to harm to individuals on a range of scales, from simple and direct harm from exposure of some secret, identity theft or financial loss, through to levels of interference in their lives that go beyond what they would ideally be prepared to tolerate, down to the simple nagging sense that they are not unobserved. Because of this, techniques have been developed for individuals to police and control their private space. Most obviously relevant to the Web and the online world, data protection was developed towards the end of the 20th century.

The data protection paradigm

Data protection is not in itself privacy protection. It is there to provide legal certainty for data controllers to be able to process and share data confident that they are not breaking the law or opening themselves up to liability, while also granting data subjects rights to influence the conditions under which data from which they are identifiable is processed. Data protection law has been particularly influential in Europe, and it underpins the EU's General Data Protection Regulation (GDPR). The rights are complex, but most importantly, before data controllers can get consent for gathering personal data, they must tell data subjects the purpose for processing the data, and once consent is granted, they cannot deviate from that purpose (there are also grounds for gathering data without consent, but these are not relevant to this paper, which is about the powers of data subjects). Data subjects are also able to ask to see the data held about them, and to correct it where it is wrong, alongside a few other rights. Data protection only applies to data from which a person is identifiable (the definition of 'personal data').

Data protection applies to all personal data, although GDPR provides stronger protections for particular categories of sensitive data. Data in general receives less protection in the US, but specific categories of data, such as medical data, financial data and data about children, are regulated very strongly. Privacy policies are key in the US – companies are required to keep to the letter of their policies, with severe penalties when they don't. Those who can prove harm (usually financial or reputational harm) from the misuse of data or a breach of privacy are able to sue for compensation.

The US has pioneered some strong measures against privacy breaches, including eye-watering fines and the necessity of informing the public about data breaches, which have been copied in the EU. However, its third party doctrine, by which individuals who allow the use of their data by a third party thereby lose any rights they possessed, is utterly unhelpful for the purposes of privacy protection.

In each jurisdiction, however, it is the responsibility of data subjects to discover any problems and harms, bring them to the attention of the authorities, and police their own private space. Someone has to be motivated enough to do this assiduously, and to continue to do it. It imposes large costs on individuals, and only provides a kind of after-the-fact protection – in other words, you might hope to be compensated after the harm is done, which might not cut the mustard if you have lost your spouse, your company or your reputation. Is there anything that can be done to police the private space beforehand, to anticipate harms and neutralise them?

The main pillar upon which data protection rests is that of informed consent – data subjects consent to the processing of data from which they can be identified or singled out, for the purposes specified. What is consented to is usually, in an online context, contained in the privacy policy, which the data subject clicks to gain access to a website or service (sometimes the use of the service after a warning is taken to imply consent). It is a cliché to point out that no-one reads privacy policies – a very tired trope in talks and lectures is to ask those in the audience who routinely read the privacy policies before clicking 'OK' to

The cognitive requirements of informed consent are high

put up their hands, and to savour the lack of response, as if it was a new discovery. The cookie notices that now appear on sites in Europe are at best a mildly tedious obstacle to accessing the content, not providers of useful information for the reader's consideration. As (Manjoo 2019) points out, news sites are among the worst trackers, so that reading a news article about tracking exposes you to a tracking ecosystem too complex to comprehend fully. He does point out that, as his news article appeared in the *New York Times*, it became part of the problem itself.

The cognitive requirements of informed consent are high (Barocas & Nissenbaum 2014). They imply that the data subject is able both (i) to understand the immense complexity of what might be happening to the data, and (ii) to negotiate the trade off between the immediate gratification of the service and the uncertain long-term risk of the data being exposed (especially as the risk is largely cumulative and dependent on whatever other relevant data is available, and the specific risk from a particular data release is probably negligible in most cases). Indeed, so quickly does the technological context evolve, in all probability no-one can truly know what risk they are likely to encounter two or three years down the line.

'privacy paradox'.. is less a paradox than a mismatch between one's reflective view and one's preferences in a specific context

It is hard, therefore, to keep control. The so-called 'privacy paradox' notes that people often profess concern about privacy while behaving in apparently reckless ways (Dienlin & Trepte 2014); this is less a paradox than a mismatch between one's reflective view and one's preferences in a specific context (for instance, one may want to get slim, while sometimes eating hamburgers, but no-one talks about the MacDonald's paradox). Indeed, our privacy norms are generally far more complex and context-dependent than the simple technological tools available (Edwards 2013, 319-324). Surveys have often revealed people's lack of trust of those who use their data, even while they fail to use the protections (both legal and technical) that they have (Smith 2017).

However, most privacy regulation has been developed within the liberal paradigm – that is, it is assumed that the point is to allow individuals (as opposed to groups, say, which are not covered in data protection legislation or privacy rights) to secure the level of privacy they prefer (as opposed to the privacy a paternalistic third party believes they ought to have), and that the individual's autonomy is the key value that the system is trying to protect. Within that mindset, no-one has thought of a better idea than consent for supporting both privacy and autonomy simultaneously.

As legal scholar Daniel Solove points out (Solove 2013), if it is assumed (i) that individuals cannot give informed consent, and (ii) that privacy should therefore be regulated in some other way without engaging the cognitively-stretched individual, then the outcome may well be that privacy is better-protected, but at the cost of sacrificing the autonomy of the individual.

Privacy-enhancing technologies

There are obvious ways to restrict the flow of information about you, even if you don't want to waste time reading privacy policies. These include keeping software up to date, using trusted Virtual Private Networks (VPNs) that don't have an advertising-based business model, blocking online tracking (possibly via the VPN), ensuring passwords or other identification methods are complex and unique to particular applications, being careful about posting personal data (for instance, holiday destinations, at least before you set out) and keeping to well-known apps on the major app stores (so that at least the problems are well-understood).

Going beyond these basic (but themselves underused) technologies, there have been many other ideas, often but not exclusively technological, put forward about how data subjects can take more control of the way their data is treated. These ideas are usually collected under the heading of *privacy-enhancing technologies* (PETs). Many PETs, such as privacy-preserving machine learning or differential privacy (Al-Rubaie & Chang 2019, Royal Society 2019), are designed for use by

no-one has thought of a better idea than consent for supporting both privacy and autonomy simultaneously

corporates in order to manage their personal data processing, and companies such as Privitar can provide business-to-business services with integrated PETs for privacy management. However, some PETs have been designed for individuals as means of enforcing their data protection rights. Examples of the ideas in this space include the following.

- **Platform for Privacy Preferences** (Grimm & Rossnagel 2000). The Platform for Privacy Preferences (P3P) was a protocol that allowed the expression of how data would be used, which became a W3C recommendation in 2002. The intended use was that websites could use it to express how it would use the data gleaned from its users, while users could also say what their preferred limits would be. If the privacy preferences of the user failed to match the privacy policies of the website, then a dialogue could be initiated to negotiate whether the difference could be closed. It was seen as difficult to use, and is no longer supported.

- **Do Not Track** (Bott 2012). Do Not Track was designed to be part of an HTTP transaction that allowed potential website users to say whether they consented to be tracked. It was supported by a number of browsers, and the W3C attempted to standardise it. However, its legal status was uncertain, and take-up was low, leading to its eventual disappearance, and the discontinuation of the W3C's efforts before a recommendation was accepted.

- **Personal Information Management Systems** (Bergman et al 2003). Personal Information Management Systems (PIMS) are sets of services that allow individuals to manage and control online data about them, and at the most expansive to determine their online identity. Data is held in online storage systems, and individuals themselves decide how and to whom access to the data will be granted. With the data under their control, data subjects can more easily exercise their data protection rights, such as rights to correct or erase data. It also presents data subjects with responsibilities, such as ensuring interoperability, cybersecurity, and encryption. Under current law, PIMS will allow individuals to manage data about them that is under their control, or to which they are entitled (for example, via GDPR portability rights); one could imagine extensions to the law which would allow individuals to insist on the transfer of personal data about them to their PIMS. However, at present,

there would need to be a tractable (e.g. contractual) arrangement between data gatherer and data subject. If the scope of PIMS was to be widened, then there would need to be a combination of laws and technologies to enforce that, first to give individuals access to the data about them, and secondly to enable them to track its usage (Oswald 2014).

- **Personal Data Stores** (Van Kleek & O'Hara 2014). A Personal Data Store (PDS) is the specific hardware core of a PIMS, a technical architecture to facilitate longitudinal, decentralised and individual-centric collection and curation of data. Such stores face a number of challenges: the requirement to store data for potentially long periods; the usability difficulties of managing data for individuals; the regulatory basis for individuals' access to their own data, and for third-party access to data; the need to comply with accepted data handling standards; and the need to future-proof data gathering against the evolution of social norms. A number of companies put forward PDSs, such as MyDex (https://mydex.org/), a Community Interest Company in the UK that provides PDSs alongside other personal data services, including secure identity services.

- **Camouflage and obfuscation** (Brunton & Nissenbaum 2015). Brunton and Nissenbaum argue that one response to information asymmetry is to camouflage your digital footprint by obfuscating the data that is processed about you, i.e. making it less accurate and harder to obtain and process (this has also been called privacy vigilantism by Oswald (2017)). This might include providing false information (e.g. when registering for a service or site), especially if the information has no effect on the service you receive. For instance, many sites ask for your birthdate; this can be of no conceivable interest, except to establish that you are an adult (which doesn't require the exact date), or to provide a piece of information that can be matched against another database so that data about you are linked.

Other obfuscation options are to use a browser plugin that clicks on every advertisement on a page, or enters random search terms to search engines on your behalf, which will pollute the profile of you that is built up. Tor (https://www.torproject.org/) creates a type of obfuscation, by disguising the routes of Internet traffic; MIX-zoning creates an area where many mobile users can switch identities, making it harder for them to be traced across the identity switch.

- **Hub of All Things** (https://www.hubofallthings.com/). The Hub of All Things (HAT) is a PIMS, based on a server hosted in the cloud which allows data subjects to create and own their own HAT microservers, allowing them to bring their own data from the Internet, exchange data with other applications, and perform their own analytics. It is therefore more of a scheme to allow people to extract value from their own data, rather than preserve privacy, although it may be used for that purpose as well. A microserver is offered through a personal data account from the Dataswift PDA management platform. Individuals own the data in their microserver, but can grant access to other companies via APIs. HAT aims to foster an ecosystem of apps which would use the data on and via the HAT server.

- **The BBC Box** (https://www.bbc.co.uk/rd/blog/2019-06-bbc-box-personal-data-privacy). The BBC Box is a personal data management system being trialled by BBC R&D, a physical device in an individual's home onto which personal data is gathered from a range of sources, collected and processed under their control. Personal data is stored locally on the hardware and once there, it can be processed and added to by apps running on the Box, as with HAT. No third party, including the BBC, can access data without authorisation. Apps have access to the data, but don't take it off the Box.

- **Global Privacy Control** (https://globalprivacycontrol.org/). Global Privacy Control (Melendez 2020) is a standard sponsored by a number of privacy-concerned companies and organisations, from DuckDuckGo to Mozilla, prompted by the untested new legal context of GDPR and the California Consumer Privacy Act (CCPA) of 2018. The aim is not unlike that of P3P, to provide a means for individuals to express their privacy preferences across the Internet, or at least through a single browser, rather than dealing afresh with each website.
  The current (2021) implementation is not legally-binding, but its sponsors believe that it will amass legal force as industry groups and regulators assess it, and as the case law from the new data protection and privacy legislation comes in. However, we should note that one of the problems with P3P and Do Not Track was their uncertain legal status, which helped hinder adoption. The plus for websites is that they would still be allowed to gather data for their analytics programmes, but not for aggregating profiles of individual users.

There are a number of issues with these technologies and protocols with respect to their use by individuals to police their digital footprint. The first is to do with usability. Despite their often user-centric design, they are disproportionately complex to use, relative to the privacy concerns of most users (Alpár et al 2011). In particular most individuals are comparatively unused to expressing their attitudes to privacy in a reflective and formal way (a fact which is likely also to have led to some of the evidence for the privacy paradox).

Secondly, they often provide formalisms for expressing preferences without a clear roadmap as to how to ensure those preferences are respected. Enforcement and tracing are extra steps, often requiring the cooperation of the websites whose behaviour is supposedly being constrained. Furthermore, the interface with the regulatory field – which is currently in flux as we await the settled case law that will follow the implementation of GDPR, CCPA, and so on – is uncertain. Regulators have not as yet accepted many PETs as standards, and have shown little sign of doing so.

these systems are disproportionately complex to use

Third, it has to be said that the take-up of these techniques and technologies has not been great. Yet many of them seem to rely for their potential on an ecosystem of users and network effects. More users for formalisms like Global Privacy Control would help provoke the development of *de facto* if not *de jure* industry standards. Ideas such as HAT and the BBC Box would appear more valuable to users if there was a wide selection of apps prepared to use the data under the specified conditions. And then, of course, if there was movement on the industry side, then the PETs themselves would become more attractive to users, for example included as defaults in many systems. Hence, network effects would be very handy for the PET industry; the problem is how to bootstrap them, especially in a world where most users are apathetic about these issues.

The PET field has been an area of intense research at least since 2000, and in that time no single product has attracted very much attention outside the legal and academic communities.

consumers and citizens have been somewhat more concerned with adopting new technologies which generate even more data

Indeed, it is salutary to consider that, since research in this paradigm began, consumers and citizens have been somewhat more concerned with adopting *new* technologies which generate even more data. The spread of smartphones and social media happened during this period, and in more recent years we have seen the burgeoning of the Internet of Things, promoted by popular devices such as voice-activated virtual assistants. The relative rates of adoption of these technologies compared to privacy enhancing technologies places in question the extent of the demand for the latter.

Certainly no-one – despite the heroic claims of Google/Alphabet and Facebook/Meta – enjoys receiving advertisements, be they ever so targeted. If avoiding tracking could be done easily, by clicking a button, it would be. Indeed, Apple's release of iOS14 attempts to do this, and that could indeed undermine what one advertising CEO calls "a vibrant app ecosystem" (Pitt 2020). Oh dear! The well-known words "we care about your privacy", which most of us see several times a day, are probably the most common lie in the English language (with the exception perhaps of "no-one likes a good joke more than I do, but …"). "We care about potential reputational damage" would be nearer the mark. However, the rub is that not only do the corporates not care about our privacy, but there's not a great deal of evidence that individuals do either, at least to the extent of doing something expensive about it. Since remarkably few people are prepared to pay for online services, it is hard to see a disruptive business model emerging.

Against this background, in the remainder of this paper, I will consider two new ideas for addressing these concerns. The first is technological, a new protocol developed by the inventor of the World Wide Web Tim Berners-Lee. The second is legal, an up-to-date adaptation of venerable trust law in order to empower data subjects and enable them to achieve a critical mass to even out power asymmetries with the consumers of data.

Solid

Solid (Social Linked Data – Mansour et al 2016, O'Hara & Hall 2021, 233-236) is a decentralised platform for social applications on the Web, in which users' data is managed separately from the applications that create it and those that consume it, rather as with a PDS or the HAT. Rather than being ceded to the app's back end storage, the data is stored in a Personal Online Datastore (pod), which is Web-accessible and from which data is portable, enabling simple switching. Users can have a range of pods, which may be provided by independent pod providers. Solid protocols are based on W3C recommendations, which are open standards for interoperability, allowing developers to create applications which range over all the data over which the user has control wherever it is stored on the Web.

The structure of the Web has shifted from a decentralised information space, to one on which most data is controlled by a relatively small number of very large corporations within so-called 'walled gardens'. These corporations derive great power from their monopolisation of those resources once they reach a certain critical mass which is protected by network effects (Zittrain 2008).

The Web is therefore currently experiencing centripetal forces, driving it towards a hub/spoke information flow model, which, on the Solid critique, provokes power asymmetries between users (at the spokes) and the corporations (hubs), creates inefficiencies in data flow and siloes in data storage, and suppresses innovation by making it harder for developers to build on top of existing popular platforms. The effect of the walled gardens is to undermine the Web's principle of uniform IDs (for interoperability) and permissionless development (to support innovation).

Solid's aim, therefore, is to counter this using current Web protocols for peer-to-peer (P2P) networking, setting up centrifugal forces to *redecentralise* the Web (see also https://redecentralize.org/). It is not therefore intended to replace the existing Web (as, for example, has been suggested about a rival redecentralisation scheme, the blockchain-based Elastos, https://www.elastos.org/). Instead, it facilitates competition for social media and large data consumers, competing against their walled gardens with a P2P ecosystem of apps and services.

Privacy is one issue that animates Solid, for (consensual) privacy breaches are encouraged in the current centralised ecosystem as data flows from the spokes to the hubs to the benefit of corporations and against the interests of individuals (Zuboff 2019). It is, however, as Berners-Lee pointed out in his 2018 MozFest talk (https://www.youtube.com/watch?v=elfSzMATcB4), only one of the hooks upon which it hangs.

The Solid vision is that individuals should have complete control of their data

The Solid vision is that individuals should have complete control of their data, and that they should be able to engage in contacts with their social networks without necessarily revealing metadata about who is in contact with whom and when, so that highly targeted advertisements, political messages, or other types of manipulation can only be crafted with their informed consent. The data uses a linked data model (based on the Resource Description Framework RDF – Cyganiak et al 2014), so it can still be shared relatively easily, with user-determined access control. A Solid server therefore goes beyond an ordinary Web server with its two requirements of access control and support for linked data. Individuals can run multiple identities on a range of servers, and even mint a quasi-anonymous ID, using it a single time to prevent it being linked to any other session involving the same individual.

Solid is administered via a project at MIT (https://solid.mit.edu/), and Inrupt (https://inrupt.com/), a company to provide a commercial context for development. Solid's aim is to create a community of users of pods stored in the cloud, attracting providers to engage with the market, providing choice as to where data is stored and how it is managed (and how much is paid for the storage service). It will also be possible to develop or administer one's own pod, although this would require technical expertise, and so cannot be the mechanism that would allow the ecosystem to scale up. The pod is independent of any apps running off it, completing the envisaged separation of data from application. App developers would design their apps' front ends, while the pod ecosystem is in effect a common back end. We might even go so far as to say that the back end of all Solid apps is the Web as a whole. The data produced by the app would be stored in whichever of the user's pods he or she selected. Hence, different or competing apps could be run over the same data, while an app could provide a seamless experience using data from different pods (even the pods of different individuals if they have all granted access, to facilitate a group interaction, for example).

The main effort so far in the Solid project has been developing the Solid servers rather than app development, although it is hard to see an ecosystem emerging without a set of apps to stimulate demand. Progress on the ecosystem, on the account of Lalana Kagal, Solid's project manager, is slow (Heaven 2020), and its growth more likely to be as a result of government mandate rather than driven from the bottom up (McGrath 2021). Inrupt's first commercial service appeared in late 2020, with pilot projects run by the BBC, the NHS, the NatWest Bank, and most interestingly the government of Flanders, whose My Citizen Profile service is intended to give every citizen a pod (Berners-Lee 2020).

The type of control provided by Solid

The aim of Solid is to put individuals in control of their data. The architecture's intended affordance is to put access controls on the data, so that third parties can only process the data when the individual it concerns permits the processing. The level of security is clearly relevant here – it needs to be ensured that a Solid pod does not have an unintended backdoor that allows hackers to gain access to the data.

If the size of the app ecosystem is to be a key enabler of the architecture, then it may be that there is room for only one of these approaches to survive.

The access controls themselves may also include further architectural constraints – for example, the pod may not allow the third party app to have access to any data at all, but may instead return answers to authorised queries, or may be governed by principles of differential privacy. Once data is copied from the Solid server onto the dataspace of the third party app (if that is allowed), then the Solid architecture can no longer constrain what happens to the data. Hence the interaction between Solid and the app must be seamless. If the idea of apps all accessing a common back end is to take off, then the technical requirements may make it hard to roll out an app across other similar architectures, such as HAT, the BBC Box, Elastos or DFinity (https://dfinity.org/), to name but four. The app may require painstaking work on the front end for each environment. If the size of the app ecosystem is to be a key enabler of the architecture, then it may be that there is room for only one of these approaches to survive. Standardisation is likely to be key (Kuebler-Wachendorff et al 2021).

Solid's main purpose is to redecentralise the Web, but as a by-product allows individuals to express their preferences as to who gets access to their data. Here is where the element of control comes in, and we find the basic privacy proposition for Solid: Solid provides the affordances for individuals to get their privacy preferences with respect to that much of their data which is stored on a Solid server. These affordances provide control: they can choose whether or not a piece of data remains private to them, or confidential to a small circle, and they can choose who can have access and who not.

The cost of this control is that individual now have more decisions to take vis-à-vis data. One way of doing this is to take the resource-heavy decision of running their own servers. This is risky in a number of ways, most obviously the server may fail without a commercial backup. Such individuals would also have to be confident that they were able to furnish cutting edge security. More likely, they would delegate these decisions to pod administrators. This requires that they need to choose, e.g. whether to pay for storage and greater control, or to go for some sort of standard free pod on a Solid community server that provided a basic service.

Note, of course, that privacy preferences cannot be guaranteed by the architecture if it has security problems, and doesn't therefore function as per specification. Note also that if a third party has been given access to the data including the ability to copy it, then the individual can no longer guarantee his or her preferences using the Solid architecture – he or she will need other means to enforce preferences and keep control (e.g. with terms and conditions over third party processing of her data).

Note also that using Solid does not necessarily help people get their preferences as to who else's data they get access to (another type of privacy preference – a preference about the privacy of others). Someone may want their apps to range over their partner's and their children's data in order to realise greater value for the household, but they will be unable to do that unless they give permission.

Note finally that preferences about privacy have to compete with other preferences, including those about whether to expend resources on a Solid pod, or on a dinner for two at Maxim's. There are always other calls on money, and even a privacy-concerned individual may prefer to spend it in different ways.

Regulation and rights

It follows from the above discussion that, as well as the architectural affordances given by the Solid architecture, individuals may need contractual arrangements with third parties who access their data, i.e. enforceable terms and conditions which can be invoked either to prevent them from, e.g., sharing the data further, or to punish them if they do. There may still be issues of enforceability and traceability, but at least there would be a legal underpinning to expectations.

It is also worth noting that some of the Solid functionality would help individuals enforce their GDPR rights, and to the extent that the data did not leave their pod, it would be a powerful tool. Typically, legislation in this area is intended to be technology-neutral, and so it is unlikely that Solid itself would be given direct support (or suppression) by any legislation.

In the United States, those holding their own data on their own servers would have extra protection from government intrusion, via the constitutional protections against unreasonable searches and in some circumstances also free speech (these protections hardly apply to the private sector). There have been controversies about companies handing data over to the government upon request (cf. e.g. Strumpf 2016, Scott 2017, Kraft 2017), but if the government wished to enter a citizen's property and take information directly off her server, she would be protected by the Fourth Amendment against unreasonable searches, unless a search warrant was granted by a court.

Part of Solid's pitch is a moral statement that individuals should have control over their data (whether to protect their privacy or otherwise), and that conversely the big tech companies have too much power as a result of the data they are able to amass and hide behind walled gardens and non-interoperable systems. Whether such empowerment of individuals with respect to their data could be supported in some legislatures is a moot point. The EU is generally sympathetic to the empowerment of individuals, while the US government tends to property interests of companies relatively more highly. Meanwhile the Chinese government is keen to maintain its own routes to access to data for social policy, and so may be less supportive of the Solid model (O'Hara & Hall 2021).

A data consumer's view of Solid

This account of Solid has been written from the point of view of individuals with privacy interests. However, the interests of data consumers are also naturally engaged. For app developers, the major issue is that in the Solid ecosystem, they do not get to design the back end architecture, and so only get to design the interface with the user. Hence, they lose their control over the data, which of course may impact their business model. They could refuse to adopt the Solid architecture, or refuse to supply services to users who wished to keep control over their data.

There may be interesting compromises – for instance, an app developer may partially define a back end in terms of a particular pod design. Use of the app might require the user to put all the resulting data into a pod controlled by the user, but designed and hosted by the developer. This may help the developer, for instance by defining the pod's API to make data transfer easier, or alternatively being able to prevent other app developers getting access to the data.

An app provider may also provide its own internal regulation, and could consider using Solid for its information management needs – it may decide, for example, that Solid would be a useful architecture for it to manage data in a GDPR-compliant manner. It could host pods on its own servers, while ceding (some) control to the individuals whose data is stored. At present, Solid appears to be largely aimed at data subjects rather than data consumers, but thought is being given within the Solid project for the app developer's experience. Whether this will be sufficient to tempt developers to enter the Solid ecosystem may depend more on the network effects of doing so, i.e. whether enough data subjects are concerned enough about privacy issues to take more responsibility in the management of their data. It may be that Solid's friendliness to the concept of linked data, through its use of W3C standards such as RDF, FOAF (Friend Of A Friend – Brickley & Miller 2014) and LDP (Linked Data Platform – Mihindukulasooriya & Menday 2015) in its platform, could make it an important delivery platform for linked data, independently (from the data consumer's point of view) of its focus on data subjects' interests.

Finally, we should not forget that solving one political problem (the power asymmetry between individuals and tech companies) and one technical problem (the increasing centralisation of the Web) does not mean that the solution will not be challenged, nor that unintended consequences may not emerge. At present, most political debate takes the form of questioning whether app providers can demand that permissive privacy policies be consented to by users before they get access to services, especially when the providers benefit from large network effects. If and when the Solid ecosystem grows and evolves, it will bring its own network effects, which will doubtless prompt important political questions about inclusion and exclusion. Can people be denied services because they 'stand by their rights'? Alternatively, can a company be obliged to provide services to someone at a loss? These are major questions in many other fields, and there is no reason to think that they won't arise if the Solid ecosystem scales up.

Solid prospects?

To round up, Solid is currently aimed at privacy-aware individuals who (a) wish to have control over who gets access to their data, (b) prefer that their activities are not monetised by platforms, or (c) ideologically prefer an open Web and support the drive to redecentralise it. A problem for the more casual user is the current lack of positive network effects in the Solid ecosystem.

However, as noted above, it is possible to view Solid as an infrastructure that might aid data consumers/app providers manage the data they consume (or, perhaps more accurately, to outsource its management to data subjects). GDPR might well be the catalyst, as handing control to the individual data subject would enable most data consumers to be GDPR-compliant by default. Data consumers may not in that event even count as data controllers in GDPR terms, but may instead be data processors, processing data without determining the purpose or means of the processing.

Data consumers could offer to host one or more pods for users, possibly for a fee, or possibly funded by the value they are thereby enabled to extract. However, the flip side of this approach would be that data consumers could not force users to put (all) their relevant data in the pods on their servers, or to keep it there for any particular period of time.

However these developments go, in the Solid ecosystem the data subject will remain one of the more important humans in the loop. Even in the context of a thriving ecosystem and the consequent network effects for providers and users alike, this must entail a degree of uncertainty for app providers. Data will quite possibly have to be taken from multiple servers, each of which may have very different APIs.

It will be hard to create application-specific queries or requests for data, until a set of standards is in place (hence an attempt to link Solid with the Data Transfer Project – Tung 2020).

By ceding control of the back end, app developers clearly advance such noble aims as GDPR-compliance, but development is likely to be more complex as the data representation is inevitably less application-specific, and standardisation is very much a work in progress.

Bottom up data trusts

Another interesting attempt to empower data subjects developed recently suggests using legal rather than technological tools. *Bottom up data trusts* (Delacroix & Lawrence 2019) trade on the idea of using trust law (Penner 2016) to facilitate their association in order to balance the power and information asymmetries with data consumers such as social networks, so-called 'data leverage' (Vincent et al 2021). The term 'data trusts' has been used for a number of different contexts recently (O'Hara 2020, Ada Lovelace Institute 2021), including by Wendy Hall and Jérôme Pesenti (2017) to refer to a means of reducing risk in data sharing for the purposes of AI, and by Alphabet subsidiary Sidewalk Labs to describe their data-handling institutions during its aborted smart cities project in Toronto's waterfront (Scassa 2020). However, in this paper, I specifically use the term 'data trust' to refer to a bottom up data trust in the meaning given by Delacroix and Lawrence.

A trust is a legal arrangement where a *trustee* governs an asset for the benefit of *beneficiaries*. The trustee might be paid a salary, and is entitled to expenses, but is not allowed to govern the asset in his or her own interests; to do so would make him or her liable to legal action from the beneficiaries. The relation between trustee and beneficiaries is called a *fiduciary* relation. A typical trust arrangement is to run a piece of property for the benefit of owners who for some reason cannot run it themselves – for example, if children inherit a property, it might be run for them by a trustee *in trust* until they come of age. The role of trustee is therefore highly responsible, and open to liabilities.

The data trust model proposes that managing personal data be the object of the trust. Data subjects would join the trust, in effect authorising the trustee to pursue their data protection rights in their names. The trustee would then be able to deal with data consumers, such as social media or advertisers, not as one individual against a powerful corporation, but as the agent of a group with a wider, more representative and therefore more valuable tranche of data. The data of an individual is relatively low in value; not much can be done with it on its own, while if it is withdrawn from a large population (of, say, thousands of data subjects), then its withdrawal won't make much difference to the value of the rest of the group's data. If, on the other hand, data subjects could get to organise in a group using a data trust, then the trustee, as the single point of decision-making, would have the correspondingly greater power that would come from having a correspondingly large quantity of data under his or her control.

Delacroix and Lawrence, like many thinkers in this space, see important value in there being an ecosystem of data trusts. Different trusts could have diverse policies – some privacy-protecting, some value-maximising, some concerned with promoting scientific and social scientific research, and so on – and data subjects could choose which trusts they wished to join (and of course could join more than one, placing different data types appropriately as they considered it sensitive). As with Solid, data subjects could discriminate in their treatment of the data about them.

If such a diverse ecosystem emerged, and grew to significant size, the trustees would be in a stronger negotiating position with respect to data consumers. Data consumers would have to offer terms and conditions that were acceptable to trustees/beneficiaries to get access to sufficient quantities of data. If the market grew strongly enough, then a tipping point might emerge whereby this would become the norm for the consumption of data.

The growth of such a market could possibly be fostered by legislation, or (see below) may even be implicit in GDPR. However, it would depend on individual data subjects being minded and motivated to take the step of placing their data in data trusts – hence they are described as 'bottom up' data trusts, even though the scheme itself actually centralises control over data in some respects.

Exactly how a trust would work is moot. There currently exists no data trust on precisely these lines, and no-one seems close to setting one up (as opposed to making a contractual arrangement for data management and branding it a 'data trust'). To begin with, it is not possible to set up a trust without there being something in trust – in other words, one could not set up a 'shell' data trust and later populate it with beneficiaries. However, it would be possible to start off with a group of willing participants, and collect more beneficiaries going forward.

The rights of the beneficiaries would also need to be determined – do they preserve their data protection rights while they are operated by the trustee? It is not clear that a data subject can renounce his or her data protection rights. Of course, data subjects can fail to pursue the rights, but they still retain them. However, this means that, even if data subjects agree not to pursue their rights while the data is in the trust, they still have the right to do something that the trustee determines should not be done (e.g. they could give consent for the use of the personal data, even though the trustee has withheld it). But this pushes against the idea of a trust, where the trustee has sole power, to be used solely for the benefit of the beneficiaries, who have renounced or been denied it. So, for instance, if a trustee operates a building in trust for a beneficiary, the beneficiary loses all property rights to the building. The trustee is within his or her rights to prevent the beneficiary from entering or living in the building, and even to evict him or her from it, even though it is the beneficiary's property. It is not clear that the same powers extend to the trustee in a data trust, because the beneficiary's rights are arguably inalienable.

In general, the question of operation of the trust would depend on what rights the beneficiaries had to question the judgment of the trustee. If the trustee somehow held sole power over the data, then he or she could take unilateral decisions as in a normal trust. However, it may be that beneficiaries are able to criticise decisions, and possibly the data trust's constitution may involve some democratic decision-making, giving it some of the properties of a cooperative. Or it could be that beneficiaries could remove their data from the scope of a particular action, opting in or out of the trustee's decisions as they see fit.

Can data trusts help solve the problem of power and information asymmetries? I would raise four questions that would have to be answered before we could answer that question in the affirmative (O'Hara 2020).

Who sets up the trust?

A trust's mission is determined by the person who sets it up, called the *settlor*, whose role is distinct from both that of the trustee and that of the beneficiaries. In the case of an inherited asset held in trust, the settlor is the original owner who sets out the terms of the trust in a will. The trustee then has the responsibility of managing the property both for the benefit of the beneficiaries, *and* consistent with the wishes of the settlor.

The settlor, then, is a very important person, as he or she sets the terms of the data trust. Who would such a person be? For simplicity, let us assume a GDPR context. Under the GDPR, the data controller determines the purpose and means of processing the data (and transferring the data to a data trust almost certainly would count as processing). However, it seems clear that controllers would not queue up to put data in trust. If a controller wished to process some data, then his or her first question will be: does GDPR let me? If the answer is 'yes', then the controller will simply process the data legally, with no incentive to make alternative arrangements. If 'no', then does the controller gain any incentive?

The data would be handed over to a trustee, whose undivided loyalty is directed solely at the beneficiaries. So the controller's interests in processing drop out of the picture altogether (as, incidentally, does the public good). Hence the decision would be unlikely to be taken by a data controller.

The decision to join a trust would therefore be made by data subjects, some of whom could club together to be settlors, and others could join later if the trust proved to be successful. They could either do this with personal data taken under GDPR portability rights, or with personal data they have collected themselves (e.g. from a wellness activity tracker). Note that portability rights only cover data given to a data controller – it does not cover data *created* by the controller, such as a profile or data inferred from the given data. However, it does include transaction data created by the use of a service.

Note also, in passing, that, because of the non-rival nature of data, if data subject get access to personal data via the portability route, the data controller is not thereby deprived of it. So the GDPR portability right, while it may increase the amount of data brought within a data trust, won't have much of an effect on the *status quo ante* of who has access to one's personal data.

a diverse cohort of trustees with management skills and a deep knowledge of data protection law must be found

However that may be, if the settlors are a group of data subjects, then we have a potential dilemma. An important driver of the theory of bottom up data trusts was the information asymmetry between subjects and consumers. One powerful argument from this is that data subjects cannot often give informed consent to the use of their data because the cognitive load of so doing is prohibitively high (Barocas & Nissenbaum 2014).

Many would agree. However, it is also a basic requirement of trust law that the settlor must be fully aware of what the trust will do. Hence the bottom up data trust model seems to require a cohort of data subjects who are simultaneously cognitively incapable of giving informed consent for the use of their data by third parties, *and* cognitively capable of understanding how a trustee will use their data on their behalf. It is possible that the case for this could be made, but in the absence of any existing data trusts, we do not know how complex their operation might be. Data consumers' privacy policies are often as long as Shakespeare plays – but so might data trusts' terms and conditions be.

Who are the trustees?

The second question is where the trustees will come from. The role of trustee is a responsible one, with many liabilities. They have to manage the data well, for the benefit of the beneficiaries, but not for their own benefit. They have to keep on top of fast changing and complex regulation (possibly across a range of jurisdictions). They can find themselves personally liable for costs incurred through negligence or breach of contract (whether by the trustee him- or herself, or of volunteers for or employees of the trust).

The beneficiary is entitled to expenses, and the trust must surely generate enough income to pay its running costs. If there is to be a salary as well, then there must be a substantial income stream. This could be from subscriptions from data subjects – but data subjects have repeatedly shown themselves reluctant in large numbers to pay for their privacy or their empowerment in other contexts. Perhaps it could come from commission from the use of data by data consumers. But then that would depend on sufficient use of the data to pay the salary, which would seem to militate against the privacy of the data subjects. One additional problem is that, in the absence of an attractive salary, the data trust field would fail to attract capable managers.

Delacroix and Lawrence (2019) argue that an ecosystem of such trusts is desirable. In that case, a diverse cohort of trustees with management skills and a deep knowledge of data protection law must be found.

But given the nebulous nature of the responsibilities, the potentially damaging nature of the liabilities, and the small-to-non-existent nature of the incentives, this may be easier said than done.

Where is the ecosystem?

Like many other schemes to empower data subjects, this requires an ecosystem to function well and to create the network effects that would produce value for the subjects. Two decades and more of thinking in this space have produced a number of interesting and original ideas. However, no large ecosystem has yet emerged. There is a bootstrapping problem – without the network effects, there is little incentive to join an ecosystem, but in the absence of the ecosystem, the network effects will be absent.

What about data trusts' network effects?

Finally, what is the data trust endgame? Suppose the first three questions are answered, i.e. we have an ecosystem in place, there are many data trusts run on diverse principles by a large cohort of effective trustees, containing plenty of valuable data. The data trusts would then be in a strong position to negotiate with data consumers, thereby ameliorating the power and information asymmetries that have been identified as the problem of the *status quo*.

a data trust ecosystem would itself be subject to powerful network effects

However, such a data trust ecosystem would itself be subject to powerful network effect**s**. As data subjects choose data trusts, they naturally look for successful ones. The successful ones therefore are likely to have a higher rate of growth as they receive more data. This would enable them to strike even better bargains with data consumers, which would then lead to more success and more growth.

The end result could be a relatively small number of very large data trusts operating as an oligopoly, while smaller trusts, unable to strike the same deals, would wither.

In other words, although the idea of data trusts was in response to the monopsony of data consumers created by network effects, its end result might easily be another oligopoly created by network effects. Without a means to dissipate the network effects in the trust ecosystem, there will still be large power and information asymmetries between trusts and the data subjects. Granted, data trusts have to work in the beneficiaries' interests, but how much choice a data subject actually had in such an oligopolistic ecosystem remains to be seen.

None of these questions is individually fatal to the idea of data trusts. However, cumulatively, they provide a basis for doubt that the data trust ecosystem will achieve the promise held out for it (O'Hara 2020).

Discussion and conclusion

Privacy concerns are increasingly being associated with – and even superseded by – concerns about the size and power of tech giants, with their business models based on walled gardens, network effects and scale (Economist 2020, Kozinets et al 2021, Nielson 2022). Solid is perhaps the most explicit attempt to supply a counterweight to help address both issues, by redecentralising the Web and giving more power to individuals to control how their data is used, to make their consent more meaningful. The data trust community, too, has rhetoric about empowering individuals and removing power and information asymmetries.

I have not, in this report, factored in the resistance of the tech giants to such developments. They are currently working through a wave of antitrust cases in the US, EU, China and elsewhere, with Google, Facebook and Alibaba in regulators' crosshairs. Anti-tech politics seems to be one of the few bipartisan areas one can identify in the US, where even Senators Elizabeth Warren and Josh Hawley, the progressives' progressive and one of the most obdurate Trump loyalists respectively, apparently find common cause. No doubt these cases will play out over a period of years, to the benefit of lawyers and lobbyists by the limousineful.

My concern here has been privacy in particular, its value, the harms caused by its unauthorised breach, and the complexities with which the data economy confronts the poor data subject. Certainly data protection rights may be necessary, but seem hardly sufficient to the task of restoring balance. Further technological and legal tools may augment those rights, but they seem all too often to depend on the growth of their own ecosystems, which always seem impossible to bootstrap. The bootstrapping problem also applies to private sector firms which want to build networks, but they can solve this through 'blitzscaling', the costly development of a network at all costs.

Blitzscaling can work, but the blitzscaler's investors must be prepared to haemorrhage funds, while the proportion of large networks that eventually return a profit seems to be relatively small.

It may be that the smart thing for privacy advocates to do is to develop a technology that does not require scale or network effects to work. Solid is designed to sit alongside the wider Web, a linked data platform that deals mainly with a specialist type of data. Its apps need not displace the apps used in the conventional digital economy. It does not have to be large to succeed, but there is a threshold below which it will be an irrelevance. Can it sustain a sufficiently interesting app ecosystem? App design is likely to be trickier, as the designer has to deal with the complex back end, and without enough scale, the results are likely to be less impressive.

the use of data has become a major political and philosophical issue

Bottom up data trusts seem rather less well-placed. Scale does seem to be needed to give them the necessary heft to negotiate with the tech giants – a trustee managing the data of ten or even a hundred data subjects does not seem to be in a much better position than the hapless individuals themselves. But it is not clear what the strategy is for scaling; no-one seems to have much incentive to set one up, however big the demand for them.

Although the use of data has become a major political and philosophical issue, the economics and behavioural incentives for changing business models are relatively weak. People appear to value free services, or at least are relatively reluctant to pay for them.

Paying would remove the need to swap services for personal data, but this is not a bargain many are keen to strike. Individual informational privacy is also on the wrong end of another economic change; there is a greater supply of personalised services and recommendations over authentic, autonomous choice (which is responding to increased demand for the former), and of course personalised services require information, and lots of it (O'Hara 2021). Finally, the individual is a weak position to defend privacy vis-à-vis the tech giants.

It is a common fallacy that privacy is a sort of control

It may be that the current privacy paradigm, based on individual rights and control, is not the best one for unpicking these problems. This is partly because individuals are in such a weak position, but it is also worth pointing out that even if they have (some) control over their personal data, they may not choose privacy. It is a common fallacy that privacy is a sort of control; however, clearly if someone controls their personal data, and publishes it all on the Internet, then they are not private. Instead, they are achieving their privacy preferences – i.e. their preference not to be private. It may also be important to factor in the preservation of social norms (reasonable expectations, contextual integrity), and even political factors, such as the damage that the data economy has wreaked on the democratic process, or the destruction of the professional news media, as first order aspects of the privacy regime, alongside individual preferences.

Similarly, privacy has been conceptualised, at least since the war and the 1948 Universal Declaration of Human Rights, as an individual matter. It is not, however, clear that this is the best way of thinking about it. The privacy of groups – families at a bare minimum, but any informational clusters characterised by confidentiality – also counts, and in the world of big data, the group is often the target of analysis and profiling, rather than the individual (Taylor et al 2017).

Yet the current operative privacy paradigm works to undermine and disintermediate the space, stripping the individual of the support of like-minded people.

In a world where privacy was not left to individual choice, but imposed, then technologies like Solid and legal instruments such as data trusts might begin to find their place. Although in the current opt-in environment, they are unlikely to flourish, it may be that a more paternalistic opt-out environment would more suit their nature. A guaranteed user base would help foster the ecosystem of apps, although data trusts would have to resolve the issues about incentives for trustees and business models outlined earlier.

Indeed, it is conceivable that the combination of a platform such as Solid that limits access to the data by third parties, and a legal instrument such as a data trust to provide strategic data governance would be a powerful combination. It might help create the data unions that some have argued for (Arrietta-Ibarra et al 2018), with the potential advantage that such data trusts will focus around natural social groups and clusters with common and coherent goals. More diverse and not necessarily coherent groups who happen to have chosen a particular data trust model from all the others in the ecosystem, agreeing perhaps in the level of privacy they prefer, may have less agreement on other aspects, such as the purposes for which the processing of data is acceptable.

Even then, it may be more sensible to facilitate and support the parallel business-to-business paradigm of PET deployment, where highly complex methods of data analysis compatible with privacy can be used at scale, and companies such as Privitar and non-profits such as UKAN (the United Kingdom Anonymisation Network – https://ukanon.net/) can advise companies and organisations on remaining GDPR-compatible.

Certain sectors, such as healthcare and telecoms, create quantities of data so vast as to outstrip the capacities of the individuals represented to police them. It may be that data storage infrastructure such as Solid and regulatory approaches such as data trusts could play a vital role in that paradigm (although perhaps at the cost of diluting their stated aims of redecentralising the Web and empowering data subjects respectively).

To conclude: there is a history of two decades or more of the development of PETs for the use of individuals to augment their data protection rights and establish and police a space for their privacy preferences, but the structural impediments to achieving critical mass seem to be insuperable. The Solid project may well provide sufficient technical facility to meet the requirements of an ecosystem, while data trusts adapt an existing legal mechanism for combination and organisation. But in the absence of a legal and rights paradigm that removes the burden from the individual, and protects group interests against disintermediation, it is still hard to see how progress can be made. Since such a paradigm would fly in the face of 70 years of rights jurisprudence, we probably shouldn't hold our breath.

References

Ada Lovelace Institute (2021). *Exploring Legal Mechanisms for Data Stewardship*, https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/.

Mohammad Al-Rubaie & J. Morris Chang (2019). 'Privacy-preserving machine learning: threats and solutions', *IEEE Security & Privacy*, 17(2), 49-58, https://doi.org/10.1109/MSEC.2018.2888775.

Gergely Alpár, Jaap-Henk Hoepman & Johanneke Siljee (2011). 'The identity crisis. security, privacy and usability issues in identity management', *arXiv*, https://arxiv.org/abs/1101.0427.

Imanol Arrieta-Ibarra, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier & E. Glen Weyl (2018). 'Should we treat data as labor? moving beyond "free"', *American Economic Association Papers and Proceedings*, 108, 38-42, https://doi.org/10.1257/pandp.20181003.

Solomon Barocas & Helen Nissenbaum (2014). 'Big data's end run around procedural privacy protections', *Communications of the ACM*, 57(11), 31-33, https://doi.org/10.1145/2668897.

Ulrich Beck (1992). *Risk Society: Towards a New Modernity*, London: Sage.

Ofer Bergman, Ruth Beyth-Marom & Rafi Nachmias (2003). 'The user-subjective approach to personal information management systems', *Journal of the American Society for Information Science and Technology*, 54(9), 872-878, https://doi.org/10.1002/asi.10283.

Tim Berners-Lee (2020). 'The Flanders Government and Solid: "An important milestone in Flemish history"', *Inrupt blog*, https://inrupt.com/flanders-solid.

Ed Bott (2012). 'Why Do Not Track is worse than a miserable failure', *ZDNet*, https://www.zdnet.com/article/why-do-not-track-is-worse-than-a-miserable-failure/.

Dan Brickley & Libby Miller (2014). *FOAF Vocabulary Specification 0.99*, http://xmlns.com/foaf/spec/.

Finn Brunton & Helen Nissenbaum (2015). *Obfuscation: A User's Guide for Privacy and Protest*, Cambridge MA: M.I.T. Press.

Richard Cyganiak, David Wood & Markus Lanthaler (2014). *RDF 1.1 Concepts and Abstract Syntax*, W3C, https://www.w3.org/TR/rdf11-concepts/.

Sylvie Delacroix & Neil D. Lawrence (2019). 'Bottom-up data trusts: disturbing the "one size fits all" approach to data governance', *International Data Privacy Law*, 9(4), 236-252, https://doi.org/10.1093/idpl/ipz014.

Tobias Dienlin & Sabine Trepte (2014). 'Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors', *European Journal of Social Psychology*, 45(3), 285-297, https://doi.org/10.1002/ejsp.2049.

The Economist (2020). 'Who owns the Web's data?', *The Economist*, https://www.economist.com/business/2020/10/22/who-owns-the-webs-data.

Lilian Edwards (2013). 'Privacy, law, code and social networking sites', in Ian Brown (ed.), *Research Handbook on Governance of the Internet*, Cheltenham: Edward Elgar, 309-352.

Lothar Fritsch (2008). 'Profiling and location-based services (LBS)', in Mireille Hildebrandt & Serge Gutwirth (eds.), *Profiling the European Citizen: Cross-Disciplinary Perspectives*, Dordrecht: Springer, 147-168, https://doi.org/10.1007/978-1-4020-6914-7_8.

Rüdiger Grimm & Alexander Rossnagel (2000). 'Can P3P help to protect privacy worldwide?', in *MULTIMEDIA '00: Proceedings of the 2000 ACM Workshops on Multimedia*, New York: ACM, 157-160, https://doi.org/10.1145/357744.357917.

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.

Will Douglas Heaven (2020). 'A plan to redesign the internet could make apps that no one controls', *M.I.T. Technology Review*, https://www.technologyreview.com/2020/07/01/1004725/redesign-internet-apps-no-one-controls-data-privacy-innovation-cloud/.

Robert V Kozinets, Daniela Abrantes Ferreira & Paula Chimenti (2021). 'How do platforms empower consumers? Insights from the affordances and constraints of *Reclame Aqui*', *Journal of Consumer Research*, 48(3), 428-455, https://doi.org/10.1093/jcr/ucab014.

Timothy J. Kraft (2017). 'Big data analytics, rising crime, and fourth amendment protections', *University of Illinois Journal of Law, Technology & Policy*, 2017(1), 249-273.

Sophie Kuebler-Wachendorff, Robert Luzsa, Johann Kranz, Stefan Mager, Emmanuel Syrmoudis, Susanne Mayr & Jens Grossklags (2021). 'The Right to Data Portability: conception, status quo, and future directions', *Informatik Spektrum*, 44(4), 264-272, https://doi.org/10.1007/s00287-021-01372-w.

Sophie Lalor-Harbord (2019). *Why You Should Protect Your Personal Data*, Stewarts Law, https://www.stewartslaw.com/news/why-you-should-protect-your-personal-data/.

Farhad Manjoo (2019). 'I visited 47 sites. hundreds of trackers followed me', *New York Times*, https://www.nytimes.com/interactive/2019/08/23/opinion/data-internet-privacy-tracking.html.

Essam Mansour , Andrei Vlad Sambra, Sandro Hawke, Maged Zereba, Sarven Capadisli, Abdurrahman Ghanem, Ashraf Aboulnaga & Tim Berners-Lee (2016). 'A demonstration of the Solid platform for Social Web applications', in *Proceedings of the 25th International Conference*

Companion on World Wide Web, New York: ACM, 223-226, https://doi.org/10.1145/2872518.2890529.

Tom McGrath (2021). 'Can MIT's Tim Berners-Lee save the Web?' *Boston Magazine*, https://www.bostonmagazine.com/news/2021/09/14/tim-berners-lee/.

Steven Melendez (2020). 'DuckDuckGo, EFF, and others just launched privacy settings for the whole Internet', *medium.com*, https://medium.com/fast-company/duckduckgo-eff-and-others-just-launched-privacy-settings-for-the-whole-internet-74e90391795e.

Nandana Mihindukulasooriya & Roger Menday (2015). *Linked Data Platform 1.0 Primer*, W3C, https://www.w3.org/TR/ldp-primer/.

Elizabeth I. Nielson (2022). 'Dislike: Facebook's anticompetitive monopoly on social media and why U.S. antitrust laws must adapt to the technological era', *SMU Law Review Forum*, 75, 120-149, https://doi.org/10.25172/slrf.75.1.2.

Kieron O'Hara (2020). 'Data trusts', *European Data Protection Law Review*, 6(4), 484-491, https://doi.org/10.21552/edpl/2020/4/4.

Kieron O'Hara (2021). 'Personalisation and digital modernity: deconstructing the myths of the subjunctive world', in Uta Kohl & Jacob Eisler (eds.), *Data-Driven Personalisation in Markets, Politics and Law*, Cambridge: Cambridge University Press, 37-54, https://doi.org/10.1017/9781108891325.004.

Kieron O'Hara & Wendy Hall (2021). *Four Internets: Data, Geopolitics and the Governance of Cyberspace*, New York: Oxford University Press.

Marion Oswald (2014). 'Seek, and ye shall not necessarily find: the Google Spain decision, the surveillant on the street and privacy vigilantism', in Kieron O'Hara, M.-H. Carolyn Nguyen & Peter Haynes (eds.), *Digital Enlightenment Yearbook 2014: Social Networks and Social Machines, Surveillance and Empowerment*, Amsterdam: IOS Press, 99-115.

Marion Oswald (2017). 'Jordan's dilemma: can large parties still be intimate? Redefining public, private and the misuse of the digital person', *Information and Communications Technology Law*, 26(1), 6-31, https://doi.org/10.1080/13600834.2017.1269870.

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

Simon Pitt (2020). 'Data privacy exhaustion is real', *medium.com*, https://onezero.medium.com/data-privacy-exhaustion-is-real-9ca868068f2b.

Teresa Scassa (2020). 'Designing data governance for data sharing: lessons from Sidewalk Toronto', *Technology and Regulation*, 2020, 44–56, https://doi.org/10.26116/techreg.2020.005.

Royal Society (2019). *Protecting Privacy In Practice: The Current Use, Development and Limits Of Privacy Enhancing Technologies In Data Analysis*, London: Royal Society, https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/.

Jeramie D. Scott (2017). 'Social media and government surveillance: the case for better privacy protections for our newest public space', *Journal of Business & Technology Law*, 12(2), 151-164.

Aaron Smith (2017). *Americans and Cybersecurity*, Pew Internet and Technology Research Center, https://www.pewresearch.org/internet/2017/01/26/americans-and-cybersecurity/.

Daniel J. Solove (2013). 'Introduction: privacy self-management and the consent dilemma', *Harvard Law Review*, 126(7), 1880-1903.

Koleman Strumpf (2016). 'Unreasonable searches: the abuse of open records laws', *Academic Questions*, 29(1), 24-30, https://doi.org/10.1007/s12129-016-9550-3.

Linnet Taylor, Luciano Floridi & Bart van der Sloot (eds.) (2017). *Group Privacy: New Challenges of Data Technologies*, Cham: Springer.

Liam Tung (2020). 'Berners-Lee's Solid project: Schneier joins team to give you back control over data', *ZDNet*, https://www.zdnet.com/article/berners-lees-solid-project-schneier-joins-team-to-give-you-back-control-over-data/.

Max Van Kleek & Kieron O'Hara (2014). 'The future of social is personal: the potential of the Personal Data Store', in Daniele Miorandi, Vincenzo Maltese, Michael Rovatsos, Anton Nijholt & James Stewart (eds.), *Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society*, Cham: Springer, 2014, 125-158, https://doi.org/10.1007/978-3-319-08681-1_7.

Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor & Brent Hecht (2021). 'Data leverage: a framework for empowering the public in its relationship with technology companies', in *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, New York: ACM, 215-227, https://doi.org/10.1145/3442188.3445885.

Oscar Williams-Grut (2018). 'Hackers stole a casino's high-roller database through a thermometer in the lobby fish tank', *Business Insider*, https://www.businessinsider.in/Hackers-stole-a-casinos-high-roller-database-through-a-thermometer-in-the-lobby-fish-tank/articleshow/63769685.cms.

Jonathan Zittrain (2008). *The Future of the Internet: And How to Stop it*, New Haven: Yale University Press.

Shoshana Zuboff (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, London: Profile.