Chapter 5
Incongruities and Dilemmas in Data Donation: Juggling Our 1s and 0s

Kerina H. Jones

Abstract The creation of vast, complex datasets made possible by technological advances over recent decades, has resulted in data becoming big business across many sectors and disciplines world-wide. Everyday life is increasingly networked via a growing array of digital devices to which individuals provide data, passively and actively. The pace of development has led to questions about the role of such 'data donors' and how individuals can be safeguarded when they might not be fully cognisant of the extent or destinations of data provided. We show that the many ways in which individuals provide data about themselves can result in incongruities and dilemmas in apparent decision making. We argue that it is not ethical for the vast swathes of data provided by individuals not to be used for public good. We explore whether we can make truly informed choices with the panoply of issues that may influence our decisions. We conclude without a straightforward yes or no, but propose that if we provide the best available information and engage with information presented, we stand a more reasonable chance. Do that, there is a need for demonstrable trustworthiness and clarity, greater awareness so that trust can be placed wisely, and for us to hone our juggling skills.

Keywords Data donation · Incongruities · Dilemmas · Big data · Tissue/organ donation

5.1 Introduction

The etymology of the word 'donation' is from the latin 'donum' meaning 'gift', with the French 'donner', to give, being a familiar derivative. It has been proposed that it can be easier to donate blood and even organs than to donate our data. This appears incongruous, and raises questions about the contexts and rationales for these positions. Using the donation of general personal data and health data in

K. H. Jones (✉)
Population Data Science, Swansea University Medical School, Swansea, UK
e-mail: k.h.jones@swansea.ac.uk

J. Krutzinna, L. Floridi (eds.), *The Ethics of Medical Data Donation*,
Philosophical Studies Series 137, https://doi.org/10.1007/978-3-030-04363-6_5

example scenarios, this discursive chapter explores areas such as: alternative consent models; the unknown element in data content; trust and trustworthiness in data custodians; and meaningful public engagement, to consider the bioethical balance between individual autonomy, personal exploitation and social responsibility. Ultimately, the question is whether we, as individuals and society, can make truly informed choices with the panoply of issues that may influence our decisions, creating dilemmas as we juggle our 1s and 0s.

5.2 Hast Thou Which Art but *Data*, a Touch, a Feeling?[1]

Digital data at the most fundamental level is represented as a combination of 1s and 0s. Over recent decades, major technological advances have enabled the creation of vast, complex datasets commonly referred to as 'big data'. Big data is big business: it has been estimated that its worth to the UK will exceed £320 billion by 2020, and that in the US, the m-health app market alone will reach almost $60 billion in the same timescale (Greenbaum 2018; City a.m. 2017). There is a global profusion of enterprises seeking to make the best use of person-based data to inform policy, health and other public services, business, marketing and an array of other commercial and non-commercial developments for public good and/or profit. There has never been such a high demand for our personal data to be donated, such that it is often said that individuals are the product, not just the client (Wu 2017). But before we begin considering data donation, it's worth highlighting that data ownership is a tricky concept in law. It's something we often refer to informally e.g. my data, your data, but laws of ownership mainly relate to people owning tangible items, such as objects or property, and data does not fit neatly into these categories. The question arises as to how someone can be said to own data, since in order to be meaningful, ownership should confer a concept of possession. Furthermore, tangible items are generally exhaustible whereas data are not, but can be used repeatedly by multiple parties for multiple purposes *ad infinitum*. With this in mind, it is indeed difficult to see how someone can be said to own their data since once the data are known to others, the person no longer has real control over their fate.

Rather than ownership, data protection legislation and regulations relate to safeguarding the privacy of data subjects and the confidentiality of the information in question. Within the EU, we have seen the recent introduction of the General Data Protection Regulation (GDPR, EU GDPR portal 2018), with concomitant national legislation, marking an overhaul in the way personal data are governed. The GDPR enhances the rights of individuals as data subjects, and it places a greater onus on data controllers and processors to justify proactively their lawful basis for using personal data, including providing suitable privacy notices for data subjects. But it doesn't ultimately change the fundamental focus on data protection rather than data ownership.

[1] Hast thou, which art but air, a touch, a feeling. (Shakespeare, The Tempest).

Article 4 (1) of the GDPR (Intersoft consulting 2017) defines personal data as: *'any information relating to an identified or identifiable natural person'*, and *'an identifiable natural person'* as *'one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'*. Article 6 (1) sets out the six lawful bases for general data processing, and Article 9 sets the provisions for processing special category data, which is defined as: *'personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation '* (Intersoft consulting 2017). Even so, it can be argued that we haven't yet uncovered all the types of data that could be seen as personal, and which could yet necessitate a further update in legislation. We are seeing a rapid increase in connectedness via urban monitoring, the internet of things and smart objects, some within our own homes or even as our clothing (Engineering & Technology 2017). Interestingly, the Article 29 Working Party on Data Protection already broadens the scope of health data, accepting that lifestyle data may constitute health data if they are inherently related to a person's health status (Article 29 Data Protection Working Party 2013). We engage in in-depth (sometimes rather personal) conversations on social media, accept store loyalty cards which track our purchases, and use a variety of lifestyle apps on our devices. The genomic revolution is opening up untold opportunities for research and medicine, that were not available just a few years ago. There is a myriad of occasions where we donate our data, either actively or passively, as we go about our daily business or take part in dedicated activities. Altogether we are creating a rich data footprint, sometimes without knowing which types of data have been collected, by whom, or even having no awareness of them at all.

5.3 What We Might Be Donating

For the purpose of this chapter, we will stick with the concept of donation as gifting something to another party or parties, even though it is not as straightforward in the donation of data as it is for more tangible objects. Our data subject, let's call him Schrödinger's Pat[2], might passively or actively donate personal data from many sources to various parties, over the course of his life or after his death. This may include data from his: health and administrative records, DNA, social media posts, mobile phone call detail records, apps, and store loyalty cards. He might also donate blood, stem cells, tissue samples and paired organs during his life, then vital organs

[2] Schrödinger's cat: a quantum physics thought experiment where a cat may be simultaneously alive and dead.

or his whole body after his death. Importantly to remember is that whatever is donated, data are being generated.

Pat might donate general and special category data to different parties for different purposes, with or without his full awareness. It can be easy, possibly too easy, to donate data in some instances. For example, we readily sign up for store loyalty cards on the promise of discounts or rewards, but it is important to remember that the data collected as we shop is of more value to the store than any benefits to the customer. Similarly with social media platforms; we gain the benefits of social contact, but our data may be used to target us for marketing and sometimes for other reasons. The recent Facebook debacle over psychological profiling by Cambridge Analytica is a case in point, where it has been alleged that personal data from a personality quiz on Facebook was used to try and manipulate voting intentions. People engaging in the quiz were unwittingly providing detailed information about themselves to be used for illicit purposes (Solon and Graham-Harrison 2018).

Pat might also choose to use health or fitness apps which collect special category data, such as diagnoses, medications and medical symptoms. Or he may choose to engage with a Direct to Consumer (DTC) DNA sequencing company to find out about his genetic susceptibility to certain conditions. DTCs provide information to customers for a fee, but generally use the data for business development, research and to sell to third parties in anonymised or aggregated form. Questions have been raised about the rectitude of DTC services, since individuals might be ill-equipped to deal with the results, and the actual predictive value of the information provided might not live up to company marketing promises (Cussins 2018). Some countries have banned DTCs and others are considering their legislation in this regard (Kalokairinou et al. 2017).

As well as donating data (in the form of data) Pat might donate blood, tissue or organs during the course of his life, or after his death. This, of course, also generates data. Within Wales, and soon to be implemented in England, there is an opt-out consent mechanism whereby a person donates their organs after death, unless they had previously opted out. It is an interesting incongruity that we have an opt out system for organ donation after death, but we do not have a similar arrangement in place for data donation after death. Yet, organs can be used to generate rich data, including full genome sequences of living relatives of the deceased, and thus may uncover highly-sensitive familial information compared to health record data that, paradoxically, cannot be shared in this way. However, it could be argued that it is the opt out consent system for posthumous organ donation that is ethically at fault, since a forced or presumed gift loses the spirit of being a gift (McCartney 2017). The question arises as to whether we can be sure the public genuinely feel they were informed about the process, or if a significant proportion just haven't engaged, and simple inertia has prevented them from going on-line to opt out. We will explore this concept in more detail.

5.4 Are We at Least with Socrates?

Whether or not Socrates actually said he was the wisest man because he was aware he knew nothing, it is an apt sentiment in gauging our own perspectives, and one we can apply in relation to data donation. For the many ways in which Pat might donate data in one form or another, the level to which he is informed and the nature of consent may vary widely. Medical research is generally well-regulated but, across other domains, Pat may give his permission to donate data via everything from properly informed consent to 'agreeing' to lengthy terms & conditions. This is a common problem and one for which there has been a number of social experiments. In a survey completed by 550 DTC customers, most respondents considered themselves aware of privacy issues, and the risk of troubling repercussions of data donation to be negligible. But over 50% men and almost 30% women also said they had not read the terms & conditions (Haeusermann et al. 2017). Among a group of over 500 students signing up to a fictitious social media channel, none of them read the terms & conditions well enough to notice that they had agreed to hand over their first-born child (Technica UK 2016). These studies warrant an exclamation mark(!) and leave us in doubt about the adequacy of consent processes in some spheres. They highlight the dilemma of where the responsibility lies in engaging properly with the public, as to whether the onus should be more on the individual or the data collector. In answer to our own Socratic question, it appears that no, we often don't know that we don't know.

5.5 Legal Position – Data vs Tissue

The EU GDPR, and related new national legislation, might be the hero to save us from some of these difficulties. Under the GDPR, the requirements for valid informed consent have been tightened, such that Recital 32 states that *'consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her'* (Intersoft consulting 2017). This places new limits on the use of opt-out consent mechanisms and on the use of lengthy, tiny wordy, scrolly downy, nobody readsy, terms & conditions. This move, together with the requirement for greater clarity on data processing in privacy notices, the right of individuals to request a copy of data pertaining to them, and the right of data erasure, serves to empower individuals on the use of their personal data. As soon as the GDPR came into force, complaints were brought against Facebook, Google and others, alleging that companies are forcing users to accept targeted advertising, or to delete their accounts (Foxx 2018). If these complaints are upheld, and as more users become aware of their rights, this could result in serious financial and reputational damage for these companies. Hopefully, it will bring about a change in practice to respect data donors and comply properly with the

regulations. Other social media giants, notably Twitter, have introduced clear, granular controls that allow users more choice, including opting out of targeted advertising which relies on user profiling (Foxx 2018).

We have already mentioned posthumous organ donation, but there is also tissue and organ donation from living individuals to consider, with the corresponding implications for the individuals and their kin. The UK, in common with many countries, has specific legislation governing the donation of human tissue from living individuals (UK Human Tissue Act 2004). Organisations processing human tissue must be licensed, abide by strict protocols and are subject to inspections by regulatory authorities. However, with the advancing genomic revolution, tissue donated by one consented individual can generate increasingly rich information about the donor and their kin, as the secrets of DNA are being uncovered. Yet, it would not be practicable or ethical to seek consent from all the possibly relevant individuals. For *bone fide* organisations we rely on good governance regimes for data management and access; but with the explosion of interest in genomic data, and huge multinational companies such as Apple, Google and Amazon entering the health market, it is not known what the future holds or whether the ensuing power-play will yet trump bioethical factors one way or another (Scott 2018).

5.6 No Man Is an Island Entire of Itself[3]

When Pat chooses whether to donate any kind of personal data about himself, he needs to remember that his decision will have implications for others. On the basis of Western philosophy, we lean on the side of individual autonomy in our bioethical principles. The four main principles we commonly rely on being: (i) the rights of individuals to make decisions and to be provided with truthful, complete information to be able to make a properly informed choice, free from coercion (autonomy); (ii) not intentionally harming individuals through acts of commission or omission, and providing standards of care meeting the law and commonly held moral convictions (non-maleficence); (iii) a duty to benefit individuals, and actively preventing harm (beneficence); and (iv) equality and fairness in the provision of care and distribution of resources by seeking to overcome disadvantages (justice) (Beauchamp and Childress 2013). But, if we are to give due consideration to data donation, individuals also have to face the concept of social responsibility, with the added dilemma of whether in the act of donating we are potentially benefitting or harming others associated with us. This can be the case in many contexts, and it not limited to the obvious genomic data example. It has been observed that some apps used on social media platforms seek access to the user's contact list and photographs, and it is worth carefully reviewing privacy settings to be sure we are aware of the data donation 'choices' we are making (Denholm 2016).

[3] No man is an island (John Donne).

Although we have referred to bioethical principles, it is interesting to consider the extent to which our data donation decisions are based on bioethics or on other considerations altogether. When we donate personal data to a research project having a defined protocol, clear aims and anticipated outcomes, with potential risks and benefits for participants, we could reasonably say we choose our position based on our moral perspectives. We might hope for benefits for ourselves, or we might be acting altruistically in undergoing an intrusive process purely for the future benefit of others, based on a sense of social responsibility. But when we donate data in other contexts, such as social media, it is doubtful that we base our decisions on ethics. Similarly with choosing to use an app, a DTC company, or a mobile phone contract. A key difference seems to be in the purpose of the transaction. Participating in a research project carries the concept of 'giving something back' and thus contributing to the good of society. Signing up to use a service or product, however, is rather different, as it is directly associated with obtaining something based on need or desire. As we've observed, we are often presented with a potentially coercive situation where we can only obtain the item if we enter into the agreement. Ironically, we are often taken stepwise through a process to donate data to research for public good, with much less attention paid as we rush to complete the transaction to gain the prize in the latter scenarios. Furthermore, the extent to which the companies with whom we engage are acting on the four ethical principles might be highly variable when profit is their primary purpose. In some cases, this might be more akin to personal exploitation by platforms of largely unaccountable power.

5.7 Data, Data Everywhere, nor any Chance to Think[4]

So if it's only partly about bioethics, it will be valuable to consider what else shapes Pat's choices and the norms of society concerning data donation. We have noted that data might be provided passively or actively and in multiple contexts as we go about our daily lives. In general, we are all subject to vast amounts of potentially influential information from many sources. Unlike in the past, our challenge is not in finding information, but in knowing how to be judicious in selecting what is reliable enough to guide our decisions. Depending how we view the commonly-referred to 'information society', we might see ourselves as the most privileged generation yet, or the one most subject to the attention merchants (Wu 2017). Estimates vary, but reviews indicate that about 2.5 million academic articles are published every year, challenging professionals to keep abreast of cutting edge knowledge to inform their practice (Ware and Mabe 2015). As members of the public with our respective areas of expertise and ignorance, we are bombarded with information on any number of topics from multiple angles and media outlets. It is difficult to ignore information once it is known to us: it is assessed for its value or resides in our subconscious

[4] Water, water everywhere, nor any drop to drink (The Rime of the Ancient Mariner, Samuel Taylor Coleridge).

waiting to become relevant. Altogether, we are subject to a vast, shifting body of knowledge that floats around shaping our social realities.

We might sometimes be relatively disengaged, and perhaps we need to be, from much of the information that comes our way. If we were to ask people travelling on a train, or in another common social setting, whether they have donated any data today, we are likely to receive puzzled looks or responses in the negative. Yet, unless we are disconnected from the digital realm, it is highly unlikely that we have not donated data in some form to someone over the course of a day. But, at times, we may find ourselves in a position where we need to make important decisions about data donation. It is on these days where we hit the personal threshold, that our worldview and bioethical principles come to the fore, and we find ourselves needing to draw upon the information available to us to navigate a moral maze. This might relate to decisions about our own (or a loved one's) health, finances, education or any of an array of issues that may impact on our personal lives, taking us beyond our public personas and into the domain we consider private. At times like this, we are likely to become more concerned that we can trust the recipients of our data and their motives, as we have to move out of our, sometimes, blasé bubbles. But by this time, we are already likely to have a substantial data footprint, which may not have been donated so thoughtfully. Perhaps it is time to re-evaluate.

5.8 Up in the Air

All things considered, we propose that there are indeed incongruities in data donation, and that we find ourselves juggling our 1s and 0s between different parties and purposes. Taking Pat as our data subject, he might donate different types of data via the same basic decision-making process, or the same data via very different processes. For example, if he uses a variety of apps, he is able to donate general personal data such as his location via a fitness monitoring app, or his medical data via a health-monitoring app reminding him when to take his medication or recording self-reported symptoms to monitor a chronic condition. Alternatively, and more starkly, he might donate health (including genomic) data to a company by remote agreement to terms & conditions via a web-based agreement form. Or he might donate the very same data to a medical researcher as part of a clinical trial, via one-to-one consultation carried out on a face-to-face basis. As mentioned above, the legitimacy of DTC companies is in question in some domains (Kalokairinou et al. 2017), but perhaps we should also be more cognisant about health-monitoring apps that are not controlled solely by our care provider, but are run by third parties primarily for gain arising from the data harvested. It is possible that when our decisions are made remotely, in the 'privacy' of our own homes, our attention is cocooned by the sense of security created by our familiar environment. But of course, this is irrelevant for digitally-connected transactions even if we may feel 'safe' when signing up on our own device. In some scenarios, Pat could be in the position of not knowing the recipients of his data, exactly what data items have been

collected, or what will be done with them by whom. The original recipient may further process the data and pass it on, or sell it, to other parties albeit in anonymised format.

5.9 Through a Glass Darkly[5]

In these respects, it can be too easy to donate personal data, as even if data have been through a process of anonymisation, it might not be impossible to derive some identifiable information from within them by attribution. This is commonly referred to as 'jigsaw' or 'mosaic' attack and in some cases can lead re-identification (UK Anonymisation Network 2018). Many studies have shown that the removal of commonly recognised identifiers (such as name and address) is insufficient to render a dataset truly anonymous. This is because of residual risks due to the presence of unique records. As a possibly surprising example, 87% of people in the US have been shown to have a unique combination of birth date, sex and zip code. Such information provides a fast-track to uncovering individual identity and has led some authors to lament the broken promises of anonymisation (Ohm 2010). Uncovering individual identity is a sufficient problem in itself, but it doesn't stop there. By using information from public sources and anonymised health data, it has been shown that the confidential health records of specific individuals can be uncovered. Famously, this occurred to a Governor of Massachusetts, where a researcher deduced and sent his health records to his office with 'theatrical flourish' (Ohm 2010)! This problem further extends to genomic data, with its implications for kin as well as for the data donor. Researchers used an open-access genetic database detailing short tandem repeats on the Y-chromosome, and used genetic similarity to infer familiality in the paternal line. By combining these similarities with information on a publically-accessible genealogy database containing surnames, they were able to reveal cases where recorded paternity and genetics did not correspond. This could obviously have serious implications for the personal lives and familial relationships of the data donors (Gymrek et al. 2013). Clearly, there needs to be something more than purported anonymisation if Pat's privacy is to be secured.

All is certainly not lost as there are many *bone fide* enterprises across the world where privacy-by-design is an integral concept with a strong emphasis on good data governance[6]. Privacy-by-design is an important whole system concept where a suite of controls is built in at all stages in working with person-based data (Intersoft consulting 2017). The environment surrounding the data is designed to be conducive to safe data storage and use, providing stronger data governance regimes than relying

[5] 1 Corinthians 13:12 (The Holy Bible).

[6] International examples – Secure Anonymised Information Linkage Databank https://saildatabank.com/; Institute for Clinical and Evaluative Sciences https://www.ices.on.ca/; Population Data BC https://www.popdata.bc.ca/; Population Health Research Network http://www.phrn.org.au/; and Scottish Informatics and Linkage Collaboration http://www.datalinkagescotland.co.uk/

on data curtailment alone. It can be a challenge to strike the optimum balance between data privacy and utility, as controls applied to datasets to limit the risk of disclosure may compromise research utility. Some examples could be aggregating age into 10-year bands, or suppressing outlying variables, depending on the research question of interest. It can be easy to be drawn into what has been termed 'privacy perfectionism', where superfluous controls are applied to datasets diminishing data utility but without providing additional safeguards (Allen et al. 2013). This is where privacy-by-design comes in as it combines physical, technical and procedural controls to provide more robust and flexible data protection (Jones et al. 2014; Pencarrick Hertzman et al. 2013).

As well as addressing the safe use of health data in general, there is considerable debate in the literature over whether genetic data need to be treated as a special case for data protection. This has been termed 'genetic exceptionalism' (Chin and Campbell 2013). As a concept, it flies in the face of some current initiatives, which aim to make genetic data open and publically-accessible. This is the established pattern with platforms for genome referencing and genome-wide association studies (GWAS) (GeneCards 2018), but more recently, there are initiatives where genetic data together with general health data (sometimes plus demographics) are being shared openly. An example of this is the Personal Genome Project (PGP) (Personal Genome Project 2018) where individuals engaging with the project can choose to make their linked health and genetic data publically accessible via a website. The PGP is clear in the information it provides to participants, including the possible risks to their privacy, such that individuals engage with their eyes open (providing they read the information properly!). Sharing data in this way can be seen as an impressive altruistic gesture, but it does also raise risks for the individual and their kin, as we have alluded to earlier. Relatives of a data donor may unwittingly be exposed to others knowing their estimated likelihood of developing a genetic condition, or of the information falling into the hands of parties who might use it to deny them employment or insurance. It is noteworthy that the GDPR does not especially single out genetic data, but classifies them along with general health data under Article 9 (Intersoft consulting 2017). Research has shown that among the general public there is almost an even split among those who think genetic data is different to other health data and those that either do not, or who are unsure (52% yes; 48% no/unsure) (Global Alliance for Genomics and Health 2017). The bioethical debate continues and is likely to do so for some time with new revelations being made about the genome. Without wishing to raise concerns unnecessarily, as individuals considering our options in donating data, we need to move away from the naïve concept that data are anonymised just because we are told this is the case. Again we come back to the judicious use of information from the plethora available to us, and the challenges this presents.

5.10 Life Through a Lens (Or Several)

By the time the information on which we base our data donation choices reaches us, it is likely to have passed through a number of filters affecting its interpretation and presentation to us, as well as to decision makers who may be acting on our behalf. Concepts can be magnified, diminished or fragmented as information passes via a series of intermediaries, with their interpretations acting as lenses, variously refracting the information and influencing the next steps. A decision maker has to be judicious in use of the information available in making choices affecting the use of their, or another's data, but information provenance might not be fully known.

This is a universal problem, not limited to the information sources already considered. Importantly, this also highlights challenges in the interpretation and implementation of privacy legislation and information governance frameworks, which may influence our decisions, and those who provide us with guidance. In a review of harms arising from the use of health and biomedical data, it was shown that the most prevalent cause of data misuse was the maladministration of data governance, rather than wilful data abuse (Laurie et al. 2015; Stevens et al. 2017). This included failures to follow correct procedures, despite guidance and the existence of standard procedures and protocols, and failures to take action to avoid data misuse taking place. The report included recommendations for improved staff training amongst other measures to strengthen information governance practice (Laurie et al. 2015; Stevens et al. 2017). However, as well as protecting proper individual privacy and safeguarding professionals, it can sometimes be the case that rules are over-stringent or their true essence is lost in translation.

The review also covered harm due to the non-use of health data. This is seen as a distinct issue and not merely the reverse of gaining the benefits of data used properly. This aspect of the work showed that there are instances where poor information governance can result in serious repercussions for individuals and society. Some pertinent causes of this problem were: lengthy and duplicative approval processes, conflicting advice, and excessive disclosure controls applied to de-identified data, limiting its utility. The apparent reasons were often quite straightforward but problematic nonetheless. They included: unclear lines of responsibility, fear of making the wrong decisions, and alterations to organisational data governance frameworks in the absence of legislative or regulatory changes. As a result, there can be a skew or deficit in the information available to data donors, like Pat, and the professionals who provide his care (Jones et al. 2017). We will return to the issue of data non-use later in this chapter. As a general rule whether we're acting as individuals or professionals, we should always seek the most definitive information, as close to the primary source as possible, when we make decisions on donating data. But, of course, we may not know the derivation of the information that reaches us, and this may leave us with a dilemma we can't really quantify, whilst needing to proceed one way or the other.

5.11 Minding Our Ps and Qs

So, with all this in mind, let's look at what we can do to raise realistic awareness. One major action is to stop asking questions that cannot be answered. In engaging with the public, there is little point in simply asking whether people think their personal data should be used or not. Sometimes data have to be used for various essential purposes, and at other times data are being used with negligible regard for social acceptability. Public engagement researchers have mostly moved on to asking more focused and answerable questions, such as how data should be used and by whom. In seeking to make public engagement meaningful, there is a need to be upfront about what we know and what we do not know. Keeping with tradition, we can use some alliteration in elaborating this point. These are at least some of the unknown Ps we need to grapple with: the package (the data content); the parties (the data users); the purpose (the data uses); and the places (the data environment).

Unless Pat obtains a copy, he would not be alone in not knowing the full details of data held about him by a given data controller. As an aside, with the introduction of the GDPR, he is now in a position where he may request this is if he wishes to do so (Intersoft consulting 2017). Two key areas where the unknown package is likely to be most pronounced, but for different reasons, are on major social media, search and retail platforms and in genomic data. It is well-known that companies such as Facebook, Google and Amazon use advanced data-scraping algorithms to source as much online information as possible about their users. This puts Pat in a position, whether he's aware of it or not, that he really doesn't know the scope or extent of data held about him and his online activities. On the contrary, with genomic data, it is the full meaning of the dataset itself that is unknown. Even with a copy of the data, Pat would not know what it all means, since that knowledge is just not available. In these scenarios, when we donate data we do so without fully knowing what they contain.

When the public are asked, they generally express differing levels of willingness to donate data to different parties. Unsurprisingly, the trend is usually in favour of non-commercial organisations and less so for the commercial sector. This is true for general health data, administrative data arising from other public services and, unsurprisingly, for genomic data (Global Alliance for Genomics and Health 2017; Cameron et al. 2014; Ipsos MORI 2016). But across all sectors, we might not have full knowledge of the parties themselves or others to whom they may pass the data; and in the act of giving the matter more thought we might even skew our own perspectives. The extent to which this may occur depends on many factors, including the body of information that shapes our personal views and the most pervasive current events in the media. It can be easy to demonise certain sectors wholescale, disregarding that they are not all one entity. This can be seen in public views on the pharmaceutical industry, towards whom distrust is often expressed. Although there have been some high-profile cases where pharma companies have behaved inappropriately with data (Cohen 2014; Goldacre 2013), poor research integrity is not necessarily limited to the private sector. It's also worth remembering that pharma

companies create the majority of our medicines, and the cost of bringing a new drug to market can stretch to $billions (Herper 2017). This requires a major commitment and is one unlikely to be embarked on frivolously; although they profit, they also produce public good. When we engage with the public, or we are the individuals being engaged with, we need to take broader issues into account, beyond the immediate. For example, by remembering that we donate our data to other major organisations, such as via social media, with far less consideration and knowledge than participating in a research study, whether commercial or non-commercial. In this way, we can hope to gain a fuller picture before we make our decisions.

This leads nicely to the purposes for which our donated data are used, of which there will be far too many to consider here. But we can take a basic division between primarily for-profit and not-for-profit. Again, we are likely to be more permissive towards not-for-profit uses of our data. But even so, it is sometimes difficult to know exactly how data will be used. This is less likely to be an issue with data we provide as part of our receipt of healthcare. However, this is not failsafe, as there have been rare cases where large volumes of National Health Service data have been passed to third parties without due governance and leading to an outcry in the media (Hern 2017). Properly-governed research from any sector should ideally include a research protocol with defined questions and data requirements. Clinical trials of medicinal products tend to have tightly-controlled specifications from the outset. But in other research designs, it is not always desirable or even possible to be completely definitive at the start. Often research studies need to build in a degree of flexibility whilst operating within the bounds of regulatory approvals, including having a relevant lawful basis for processing the donated data. Where participant consent is relied upon, it must be properly informed and freely given. Thus we may have a conundrum in some research scenarios: unknown elements vs the need to inform. When this occurs, it is the duty of researchers to be upfront in the recruitment process so that participants have the best information available on which to base their decisions. Across other for-profit domains where Pat may donate his data, the purposes of data use could be vaguer and more exploitative, as we've noted earlier.

As well as a measure of unknowns in the package, parties and purposes, we may also be faced with uncertainties in the places: that is, in the data environment. Data could be stored and managed in a myriad of ways, just some of which are outlined briefly since this is a vast subject beyond the scope of this chapter. They might reside in anything from a simple, locally-held database under the control of a single individual to a large-scale platform with privacy-by-design. The data could be stored on a single PC, on a local server, or on a cloud-based storage system operating across jurisdictions. They could be publically accessible, or subject to access restrictions, and they could be released externally or retained within a data safe haven. How and where the data are to be held is an essential data governance issue for the safe, secure use of data. It calls for assurance that security measures have been applied to mitigate risks, and that the data custodians can be considered trustworthy. This is needful, not just to satisfy regulatory authorities, but also in communicating with, and respecting, individuals donating their data. It is part of conveying transparent information to promote informed choice. Of course, it might

not be appropriate to describe the security model and control measures in technical detail, but it is important to do so in a way that enables individuals to understand the principles of how their data will be handled and protected. Again, it's about clarity and the limits thereof. To complete our consideration of Ps and Qs, we ask what we can best do, and we propose that: as professionals we should be honest about uncertainty to the best of our knowledge; and as individuals, we should recognise that there are sometimes limits in knowledge when working with data. Even so, all must be conducted with integrity and trustworthiness on all sides.

5.12 The Need for Innovation in Data Governance

Having established some of the many complexities and uncertainties to be taken into account in data donation, we propose that there is a strong need for innovation in data governance so that the best use of data can be made in safe, socially-acceptable ways. Pat's data are subject to a range of factors influencing their transformation into information, not limited to the specifications of legislation and regulations. There is a complex interplay between legislation and regulations, how they are interpreted and the body of knowledge that influences this process, such that the realisation of information from Pat's data, in all its manifestations, is dependent on a variety of factors. These instruments give rise to broader ethical, legal and societal issues (ELSI), implementation frameworks and due diligence processes. In line with the famous adage that 'in theory there is no difference between theory and practice: in practice there is', implementation is dependent on interpretation. Different individuals and organisations will have their own perceptions, risk appetites and motivations in coming to an opinion on a data governance issue. As well as that, these perceptions are coloured by the body of opinion coming from stakeholders and the wider media. In these respects, this is a prime example of our life through a lens illustration. As we've noted, the 'big data' landscape is one that is expanding rapidly, with increases not just in data volumes but also in data types being donated. These emerging types, such as genomic, imaging, free-text, social media and smart object data can stretch current data governance regimes. They call for innovative solutions to avoid undue bureaucracy and support the safe, socially-acceptable use of donated data. It is a positive step that the GDPR has introduced measures to strengthen individual rights over their personal data, and to limit exploitative activities such as automated profiling (Intersoft consulting 2017). But even so, data analytics often run in advance of ELSI-based solutions, particularly in areas where data are seen as a lucrative commodity. Thus, there is a need for innovation to enable safe, effective data use without undue bureaucracy.

5.13 Who Are You, Who, Who, Who, Who?[7]

By now, we might really want to know what our role as individuals and professionals can be or should be in relation to the complex issue of data donation. This, of course, will depend on a multiplicity of factors, and is really only a question we can answer for ourselves in relation to each data donation instance. But it is worth considering some of the factors that may dictate our role. Firstly, our views are likely to be strongly influenced depending whether we ourselves are the data donor, whether they relate to our close kin, and the perceived sensitivity of the likely data content. Whether we realise it fully or not, our thinking is shaped by our worldview and the vast, shifting body of knowledge that floats around shaping our social realities that we mentioned earlier. We also have the urgency of the issue versus the distance from the issue, and whatever defines that concept in a given instance. Furthermore, as we have shown, we are often likely to be dealing with incomplete information on which to base our choices. Altogether, it's not surprising that Pat may find himself in a quandary when he has to come out of his comfort zone and make important decisions about his data. Though painful, this can actually be a good thing, on the basis that at least he can question his position and the information being presented to him in making his choices. For Pat, and any of us, there are likely to be occasions where we need to take on different roles to fulfil our familial and social responsibilities. We might be leading, shaping a situation, or basically a follower guided by others. The key, in all cases is to seek and provide the best available information to inform decisions. There are various versions of the phrase: 'without data, all decisions are guesswork', and ultimately, our aim is the safe, socially-acceptable, increasing use of data for public good. The alternative is unthinkable.

5.14 Finding the Black Cat in the Dark Room

There is no doubt that data saves lives, and this phrase itself is the rallying cry of a major public engagement campaign to highlight the value of health data, and to encourage people to pledge their support for their safe re-use in research for the benefit of individuals and society (Farr Institute of Health Informatics Research 2018). Earlier, we looked briefly at the influence of poor information governance practice on the non-use of health data. Let's look at the implications of data non-use in a little more detail. The aforementioned international case study covered the non-use of health data across clinical records and research domains, as well as in relation to governance regimes (Jones et al. 2017). From this study, it became evident that there are multiple reasons for data non-use, compounding each other and resulting in serious harm to individuals and society. As a result, health data non-use has been strongly implicated in hundreds of thousands of deaths and £billions in financial

[7] Who are you? (Pete Townsend, The Who).

burdens to societies (Jones et al. 2017). It is a challenging issue to study, as harm due to data non-use is difficult to attribute unequivocally, but there is no doubt that this black cat certainly is there. The study concluded that, although there are many initiatives seeking to address this problem, much more needs to be done. The most effective moves are likely to be those that that: (i) uncover the sources, types and reasons for data non-use in a given domain; and (ii) recognise the multiple aspects to this complex issue across other domains in seeking solutions to move steadily towards socially responsible reuse of data becoming the norm. Pat's life is at stake here, and it has even been argued that harm due to data non-use is a greater risk than data misuse (St. Clair 2008). Unlike Schrodinger's pet, it is most certainly alive and manifests itself globally as a large, agile, polymorphic, lethal, black cat that must be captured and tamed.

5.15 Conclusion

Having considered some of the incongruities, dilemmas, risks and benefits in data donation, we argue that on balance it is not ethical for the vast, increasing swathes of data donated in good faith by individuals not to be used for public good. Determining what constitutes public good is another issue in itself and one that has not been explored here. It appears that, in common with other aspects of life, individuals might simultaneously hold conflicting beliefs with regards to data donation. The challenge lies in finding a bioethical balance between individual autonomy, personal exploitation and social responsibility, when our knowledge is incomplete and powerful actors have their own agendas. But to use one more analogy, let's not throw out the baby with the bathwater, but endeavour to pursue the best information we can obtain.

The demand for personal data is massive and multi-faceted, and it is in our interests to be guarded in our influences and to invest our trust with caution. Our ultimate question was whether we, as individuals and society, can make truly informed choices about data donation with the panoply of issues that may influence our decisions. This might not be a question that can be answered with a straightforward yes or no, for all the reasons we have discussed in this chapter, and more besides. It might well be impossible for any individual to comprehend the breadth of data use and its implications, but if we provide the best available information and engage properly with the information presented to us, we stand a more reasonable chance. To do that, there is an obvious need for carefully placed trust, demonstrable trustworthiness, and for Pat to hone his juggling skills.

References[8]

Allen, J., C.D.J. Holman, E.M. Meslin, and F. Stanley. 2013. Privacy protectionism and health information: Any redress for harms to health? *Journal of Law and Medicine* 21 (2): 473–485. https://www.researchgate.net/profile/Judy_Allen3/publication/260561318_Privacy_protectionism_and_health_information_Is_there_any_redress_for_harms_to_health/links/55489a9e0cf2e2031b388b1a.pdf.

Article 29 Data Protection Working Party. 2013. Opinion 02/2013 on apps on smart devices. https://autoriteitpersoonsgegevens.nl/sites/default/files/downloads/int/wp202_en_opinion_on_mobile_apps.pdf.

Beauchamp, T., and J. Childress. 2013. *Principles of biomedical ethics*. 7th ed. New York: Oxford University Press.

Cameron, D., S. Pope, and M. Clemence. 2014. Dialogue on data: Exploring the public's views on using administrative data for research purposes. http://www.esrc.ac.uk/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/.

Chin, J.J.L., and A.V. Campbell. 2013. What – if anything is special about "Genetic Privacy"? In *Genetic privacy: An evaluation of the ethical and legal landscape*, ed. T.S.-H. Kaan and C.W.-L. Ho. London: Imperial College Press.

City a.m. 2017. Scottish Development International. http://www.cityam.com/261165/can-you-get-your-teeth-into-uks-322-billion-data-sector-22/3/17.

Cohen, D. 2014. Dabigatran: How the drug company withheld important analyses. *BMJ* 349: g4670. https://www.bmj.com/content/349/bmj.g4670.

Cussins, J. 2018. Direct-to-consumer genetic tests should come with a health warning. In *Beyond bioethics: Toward a new biopolitics*, ed. O.K. Obasogie and M. Darnovsky. Oakland: California University Press.

Denholm, E. 2016.Why we should worry about WhatsApp accessing our personal information. *The Guardian* (10th November 2016). https://www.theguardian.com/commentisfree/2016/nov/10/whatsapp-access-personal-information-privacy-facebook-consumers-information-commission.

Engineering & Technology. 2017. Clothing embedded with smart fabric used as digital key to open doors. https://eandt.theiet.org/content/articles/2017/11/clothing-embedded-with-smart-fabric-used-as-digital-key-to-open-doors/.

EU GDPR portal. 2018. https://www.eugdpr.org/eugdpr.org-1.html.

Farr Institute of Health Informatics Research. 2018. Data Saves Lives. http://www.farrinstitute.org/public-engagement-involvement/datasaveslives.

Foxx, C. 2018. Google and Facebook accused of breaking GDPR laws. *BBC News* (25th May 2018). http://www.bbc.co.uk/news/technology-44252327.

GeneCards. 2018. https://www.genecards.org/.

Global Alliance for Genomics and Health. 2017. Your DNA, Your Say. https://societyandethicsresearch.wellcomegenomecampus.org/sites/default/files/media/item/your-dna-your-say-at-the-global-alliance-for-genomics-and-health-plenary/files/20171017-orlando-florida-usa-ga4gh-plenary-video-of-slides.mp4.

Goldacre, B. 2013. *Bad Pharma: How medicine is broken and how we can fix it*. London: Fourth Estate publishers.

Greenbaum, D. 2018. Avoiding overregulation in the medical internet of things. In *Big data, health law and bioethics*, ed. I.G. Cohen, H.F. Lynch, E. Vayena, and U. Gasser. Cambridge: Cambridge University Press.

Gymrek, M., A.L. McGuire, D. Golan, et al. 2013. Identifying personal genomes by surname inference. *Science* 339: 321 https://pdfs.semanticscholar.org/c8f1/44712004bdfeb779b17d0b9c31e08ed2b0ef.pdf.

[8] All online references were checked, May 2018.

Haeusermann, T., B. Greshake, A. Blasimme, et al. 2017. Open sharing of genomic data: Who does it and why? *PLoS One* 12 (5). http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0177158.

Hern, A. 2017. Royal Free breached UK data law in 1.6m patient deal with Google's DeepMind. *The Guardian* (3rd July 2017). https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act.

Herper, M. 2017. The Cost of Developing Drugs Is Insane. *Forbes* (16th October 2017). https://www.forbes.com/sites/matthewherper/2017/10/16/the-cost-of-developing-drugs-is-insane-a-paper-that-argued-otherwise-was-insanely-bad/#160396842d45.

Intersoft Consulting. 2017. EU GDPR – final text neatly arranged. https://gdpr-info.eu/.

Ipsos MORI. 2016. The One-Way Mirror: Public attitudes to commercial access to health data. Wellcome Trust. https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf

Jones, K.H., D.V. Ford, C. Jones, et al. 2014. A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation. *Journal of Biomedical Informatics: Special Issue on Medical Data Privacy* 50: 196 https://www.sciencedirect.com/science/article/pii/S1532046414000045.

Jones, K.H., G. Laurie, L.A. Stevens, C. Dobbs, D.V. Ford, and N. Lea. 2017. The other side of the coin: Harm due to the non-use of health-related data. *Indian Journal of Medical Informatics* 97: 43–51 https://www.sciencedirect.com/science/article/pii/S1386505616302039.

Kalokairinou, L., H.C. Howard, S. Slokenberga, et al. 2017. Legislation of direct-to-consumer genetic testing in Europe: A fragmented regulatory landscape. *Journal of Community Genetics* 9 (2): 117–132. https://link.springer.com/article/10.1007/s12687-017-0344-2.

Laurie, G., K.H. Jones, L. Stevens, and C. Dobbs. 2015. A review of evidence relating to harm resulting from uses of health and biomedical data. Nuffield Council on Bioethics. http://nuffieldbioethics.org/wp-content/uploads/FINAL-Report-on-Harms-Arising-from-Use-of-Health-and-Biomedical-Data-30-JUNE-2014.pdf.

McCartney, M. 2017. When organ donation isn't a donation. *BMJ* 356: j1028. https://doi.org/10.1136/bmj.j1028.

Ohm, P. 2010. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review* 57: 1701. https://www.uclalawreview.org/pdf/57-6-3.pdf.

Pencarrick Hertzman, C., N. Meagher, and K.M. McGrail. 2013. Privacy by Design at Population Data BC: A case study describing the technical, administrative, and physical controls for privacy-sensitive secondary use of personal information for research in the public interest. *Journal of the American Medical Informatics Association* 20 (1): 25 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3555322/.

Personal Genome Project. 2018. https://www.personalgenomes.org/gb.

Scott, D. 2018. Why Apple, Amazon and Google are making big health moves. https://www.vox.com/technology/2018/3/6/17071750/amazon-health-care-apple-google-uber.

Solon, O., and E. Graham-Harrison. 2018. The six weeks that brought Cambridge Analytica down. *The Guardian* (3rd May 2018). https://www.theguardian.com/uk-news/2018/may/03/cambridge-analytica-closing-what-happened-trump-brexit.

St. Clair, D. 2008. Non-use of patient clinical data a greater risk than misuse. *Managed Healthcare Executive*. http://managedhealthcareexecutive.modernmedicine.com/managed-healthcare-executive/news/non-use-patient-clinical-data-greater-risk-misuse?page=full.

Stevens, L.A., G. Laurie, C. Dobbs, and K.H. Jones. 2017. Dangers from within? Looking inwards at the role of maladministration as the leading cause of health data breaches in the UK. In *Data protection and privacy: (In)visibilities and infrastructures*, ed. R. Leenes et al. Cham: Springer International Publishing https://link.springer.com/chapter/10.1007/978-3-319-50796-5_8.

Technica UK. 2016. TOS agreements require giving up first born—And users gladly consent. https://arstechnica.co.uk/tech-policy/2016/07/nobody-reads-tos-agreements-even-ones-that-demand-first-born-as-payment/.

UK Anonymisation Network. 2018. http://ukanon.net/about-us/ukan-activities/.

UK Human Tissue Act. 2004. https://www.hta.gov.uk/policies/human-tissue-act-2004.
Ware, M., and M. Mabe. 2015. *The STM Report: An overview of scientific and scholarly journal publishing*, 4th ed. https://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf.
Wu, T. 2017. *The attention merchants: The epic struggle to get inside our heads*. London: Atlantic Books.