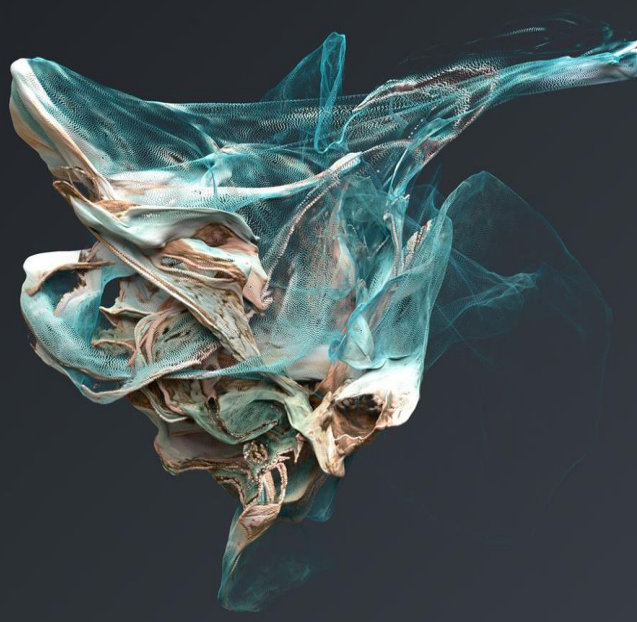

Privacy, Consent, and Trust

Response to the Data Sharing and Release Legislative Reform Discussion Paper

CLAIRE BENN, ROBERT C. WILLIAMSON, BEN RUBINSTEIN AND CHRIS CULNANE

14 OCT 2019

HMI

INTRODUCTION[1]

Integrating our data across government can yield tremendous efficiencies, enabling the kind of smooth service delivery that citizen-consumers have come to expect. In the aggregate, advances in data analytics can now yield unexpected and highly beneficial insights into human behaviour, which the government can harness in the interests of the public. But those advances pose significant risks of harming the very people they are intended to benefit. For the individual, big data enables 'predictive privacy harms'—where personally identifying or sensitive information can be inferred from either 'deidentified' PII or other information.[2] For society, big data enables the prediction and manipulation of our beliefs, preferences and actions through algorithmic news delivery, targeted advertising, and behavioural 'nudges'.[3] Even nudges, which are usually presented as wholly benign, are not always viewed positively by those being nudged.[4]

We are in uncharted territory. As great as the need is for new legislation on the sharing and release of government data, there is an equal need for a collective conversation about how the contextual norms of privacy[5] should be adapted for the era of big data, as well as about which trade-offs to make to realise the benefits of data integration, sharing and release. We welcome the DS&R's contribution to advancing that conversation.

We organise our comments around three values at the heart of the paper: privacy, consent, and trust.

PRIVACY

Why does privacy matter?[6] At its heart, privacy is about control. The DS&R affords a number of important measures to ensure that *access* to government data is controlled:

1. Data will be shared only when it meets several tests (the Purpose Test, the Five Data Sharing Principles, and Data Minimisation).
2. Data will be shared only with those who have been accredited.
3. It will be overseen by the National Data Commissioner.
4. The Data Agreements will be made available to the public.

---

[1] The views presented in this response paper are those of the authors on behalf of the *Humanising Machine Intelligence* (HMI) ANU Grand Challenge project and of Ben Rubenstein and Chris Culnane, and do not represent the views of the Australian National University. We thank Seth Lazar, Michelle Nic Raghnaill and Pamela Robinson for their comments and suggestions on this and earlier drafts.
[2] Kate Crawford and Jason Schultz, 'Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms', *Boston College Law Review* 55 (2014), 93-128.
[3] Karina Vold and Jessica Whittlestone, 'Privacy, Autonomy, and Personalised Targeting: Rethinking How Personal Data Is Used', *IE University's Centre of Governance of Change* (2019).
[4] Gidon Felsen, Noah Casrtelo and Peter B. Reiner, 'Decisional Enhancement and Autonomy: Public Attitudes Towards Overt and Covert Nudges', *Judgement and Decision Making* 8(3) (May 2013), 202-213.
[5] Helen Nissenbaum, 'Privacy as Contextual Integrity', *Washington Law Review* 79(30) (2004), 101-139.
[6] Daniel J. Solove, 'A Taxonomy of Privacy', *University of Pennsylvania Law Review* 154(3) (2006).

However, ensuring that there is no unauthorised access to government data only ensures privacy as *security*. More needs to be done to ensure that other vital controls are strengthened to protect our interests in other forms of privacy, namely privacy as *secrecy* and privacy as *autonomy*.

PRIVACY AS SECRECY

Privacy is often understood as the control someone possesses over personal and sensitive information about themselves: this is a conception of privacy as *secrecy*. The DS&R as it currently stands does not guarantee privacy as secrecy, offering no in-principle restrictions on the quantity or content of the data that can be shared, inferred or integrated. What is shared could be an entire dataset or a subset of that dataset; it could include only revealed data or all that has already been inferred; it could be a dataset in its original or post-processed form. It could be a dataset where individuals are completely identifiable and the data concerns sensitive or protected attributes. Or on the other end of the spectrum, it could be a highly abstract dataset of anonymised information (such as the demographics of States and Territories).

The issue of access to personal and sensitive data is partially addressed by the 'data minimisation' principle adopted from the *Privacy Impact Assessment*, according to which the government will 'only authorise the sharing of data that is reasonably necessary for a permitted purpose'. However, this principle leaves considerable room for interpretation, and lacks teeth. More data will always be considered necessary for better results.

We suggest that the DS&R should clearly distinguish between different kinds of data that merit different levels of protection. The DS&R should also explicitly state that a higher justificatory bar must be met in order to access more protected data and that such justifications need to be public and subject to independent auditing (that is, a procedure for checking that those who received data have kept to the Agreement).

When the use of identifiable information is deemed unnecessary for the purpose under consideration, it is proposed that the data will be anonymised, protecting privacy as secrecy. However, the success of such a strategy in guaranteeing privacy (understood as secrecy) relies on the claim that such anonymisation is actually possible. There are good reasons to doubt this claim. We detail two.

METHODS OF ANONYMISATION

First, the prevailing methods for anonymisation of data are viewed by many researchers as flawed. We caution against *any* reliance on anonymisation technology, especially technologies that have not undergone an independent technical privacy audit. For example, aggregation does not work as a privacy protection. The US Census Bureau's reconstruction attack of the 2010

census data is an important example of how aggregation can be reversed.[7] The response to these attacks has been to employ differential privacy for the 2020 census.[8] We recommend that existing secrecy constraints be strengthened by explicitly building in requirements for more rigorous methods of anonymisation such as differential privacy.

Furthermore, there is value in making public the methods used to anonymise data. As John Abowd argues in the context of releasing statistics from the US Census data: "being able to say exactly what you did to protect the confidentiality of the data is a benefit. It allows the users of the data to understand how to treat the statistics that you release and it allows the providers of the data to independently verify that the confidentiality protection algorithms work."[9] The secrecy protections offered in the DS&R should ensure that they do not rely on secrecy of the *method* of anonymisation.

Accredited partners who seek to access government data should by default receive differentially private (or equivalently secured) datasets, and only be allowed access to unperturbed datasets if they can demonstrate that the additional gain to the public interest justifies the additional risk of predictive privacy harms. [10] These trade-offs should be made with full transparency and auditability.

Sharing more robustly deanonymized datasets would also provide insurance against inevitable security breaches, and enable greater public trust in government handling of data.

DATA INTEGRATION

Second, determining whether or not a dataset has been anonymised is not something that can be determined in isolation. The degree of anonymisation of a particular data-set is *not* a property of that data set alone, but is instead a property of *all possible datasets that exist in the world either now or in the future.* This is because the *integration* of datasets has been shown to lead to re-identification. For example Chris Culnane, Ben Rubenstein and Vanessa Teague were able to re-identify themselves, their co-travellers and complete strangers from the Myki public transport dataset (released as part of the Melbourne Datathon 2018).[11]

The DS&R (like the *Privacy Act*) limits its concern to cases where it is *reasonable*, not just possible, that re-identification can occur. However, the notion of 'reasonableness' in this domain is

[7] https://queue.acm.org/detail.cfm?id=3295691 Note that other attacks are possible against aggregations such as differencing attacks: Cynthia Dwork and Aaron Roth, 'The algorithmic foundations of differential privacy' *Foundations and Trends in Theoretical Computer Science* 9.3–4 (2014), 211-407.
[8] https://dl.acm.org/citation.cfm?id=3226070
[9] https://youtu.be/R_8riuhIw-4?t=868
[10] As economists at Harvard and Brown have recently shown, differential privacy need not reduce the utility of a dataset: Raj Chetty and John N. Friedman, 'A practical method to reduce privacy loss when disclosing statistics based on small samples, *AEA Papers and Proceedings* 109 (2019), 414–20, https://www.brown.edu/Departments/Economics/Faculty/John_ Friedman/dp_aea.pdf.
[11] Chris Culnane, Benjamin I. P. Rubinstein and Vanessa Teague, 'Stop the Open Data Bus, We Want to Get Off' (Aug. 2019), arXiv: 1908.05004.

vague. It does not protect against the risks of re-identification, risks that must be borne by the public. For example, when the federal Department of Health published the de-identified longitudinal medical billing records of 10% of Australians, it was judged that it was reasonably not re-identifiable. However, Culnane, Rubenstein and Teague successfully re-identified patients' data records.[12]

It is true that the DS&R does offer a provision to deal with possible integration of datasets, acknowledging the risks of re-identification. It currently states that if the source datasets were subject to non-disclosure, then so is the integrated, enriched dataset. If none of the source datasets were so subject, the penalties of the DS&R apply. However, the DS&R sharing affordances apply in order to overcome existing secrecy provisions. So if I have one dataset and have access to another that falls under a secrecy provision, the default is that the integrated dataset would fall under that secrecy provision. However, the position of the DS&R as it currently stands suggests that the integrated dataset could then be used and shared if it meets the tests set out in the DS&R, given that the purpose of the DS&R is to offer a way round secrecy provisions. In order to make its privacy protections more robust, we suggest that in the legislation the limitations, requirements and safeguards to be placed on the integration of datasets be stated in greater detail, given the known risks that such integration poses to privacy.

THE DATA SHARING PRINCIPLES

The DS&R proposes to deploy five Data Sharing principles, modelled on the ABS "Five-Safes", to help manage risks to privacy. We agree that these principles may mitigate the risks of unauthorised access. However, the Five-Safes are not privacy principles (as noted in the DS&R) and the Data Sharing Principles do not add anything new about privacy except to say that privacy is a "factor" to "consider". Therefore, privacy concerns about the Five-Safes—for example that they suggest risks in one area might be traded off against surety in another—apply to the Data Sharing Principles and still need to be addressed.

In their description of the Five-Safes, the ABS make explicit the need to balance the risk of disclosure against the utility of data sharing. But how will such trade-offs be made? How much utility justifies how much risk? How are those utilities and risks distributed? These trade-offs should be described in the Data Sharing Agreements so they can be publicly assessed. Making the Data Sharing Agreements available for a period of time *prior* to the data exchange would facilitate this public assessment and allow potential and future risks to be identified and rectified.

[12] Chris Culnane, Benjamin I. P. Rubinstein and Vanessa Teague 'Health Data in an Open World', (Dec. 2017), arXiv: 1712.05627.

PRIVACY AS AUTONOMY

We have seen some of the limitations on protecting the secrecy of individuals' data. However, even if it were possible to anonymise data and maintain privacy as secrecy (i.e. preventing the inference of any sensitive or personal information from the dataset), a further dimension of privacy remains unaddressed: *privacy as autonomy*.[13]

Big data enables the prediction of individual and group-level traits and behaviour to an alarming level of accuracy.[14] If we can predict individual and group traits and behaviour, then we can influence people's behaviour without their knowing. This is manipulation; it stands in opposition to autonomy. And it can occur even when the *secrecy* of an individual's data is preserved.

The DS&R is explicit about limiting the purposes for which data can be shared, leaving compliance, assurance, and national security to other legal instruments, and focusing on improving service delivery, policy, and programs. But it also supports sharing data for research and development, and potentially for commercial use. All of these purposes are described as being 'in the public interest', but even what is in the 'public interest' broadly construed can erode the vital interest citizens have in freedom from scrutiny and unnecessary influence, especially from governments.

For example, PM&C has a behavioural economics team: BETA. Their role is to "help people put their good intentions into action" by shaping their behaviour, for example, through 'nudging'.[15] While the intention behind BETA is undoubtedly positive, would the public welcome BETA having access to integrated datasets that may cover all aspects of people's lives and using them to alter people's behaviour through targeted interventions? Many would find the influence made possible by this wealth of data concerning, even when it is benign. And it is reasonable to worry about what could be done with this kind of power under less favourable conditions.

Academic uses of data may also ultimately undermine people's autonomy. Cambridge Analytica had its origins in academic research—Alexsandr Kogan was a post-doc at Cambridge. And of course commercial uses of big data have often been intentionally manipulative.[16] We do not see any grounds for vesting this kind of power in for-profit entities.[17] Relying on existing law to

---

[13] Joseph Kupfer, 'Privacy, Autonomy, and Self-concept', *American Philosophical Quarterly,* 24(1) (1987), 81-9.

[14] Michal Kosinski, David Stillwell, and Thore Graepel, 'Private Traits and Attributes Are Predictable from Digital Records of Human Behavior', *Proceedings of the National Academy of Sciences of the United States of America* 110/15 (2013), 5802-5805. These predictions are not perfect (few predictions are), but, and this is the crucial point for preservation of autonomy, even quite noisy predictions can facilitate effective manipulation: indeed *any* predictability better than random chance can do so.

[15] https://behavioraleconomics.pmc.gov.au/about

[16] Crawford and Schultz, 'Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms', 55 B.C.L. Rev. 93 (2014), https://lawdigitalcommons.bc.edu/bclr/vol55/iss1/4.

[17] Especially to those whose business model is the influence of behaviour. See Karina Vold and Jess Whittlestone, *Privacy, Autonomy, and Personalised targeting: Rethinking How Personal Data is Used*, Forthcoming in the *Report on Data, Privacy, and the Individual in the Digital Age*, by IE University's Center of the Governance of Change, edited by Carissa Véliz (2019).

protect citizens against manipulation may not be sufficient. ONDC should explore public opinion about different kinds of use of data, beyond the broad categories identified.

In summary, privacy can be understood as security, as secrecy, and also as autonomy. And this list is not exhaustive. The DS&R discussion of privacy would be strengthened by recognising and reflecting the complexity inherent in the concept of privacy and our interest in it. Specifically, the provisions it establishes for protecting privacy would benefit by being cognizant of three separate dimensions of concern:

(A) The private nature of the information or context in which the information is revealed (ranging from personal, to intimate, to semi-private, to public). This would be reflected in different standards of justification being required for access to different levels of personal information.
(B) The positive or negative nature of that privacy (as freedom from (e.g. interference) or as freedom to (e.g. self-development)). This would be reflected in limitations on targeting and behavioural influences in addition to focusing on improving individual welfare.
(C) The purpose of privacy (as restricting access or as informational control). This would be reflected in restricting access to data on privileged attributes by default and also allowing a greater public voice throughout the on-going process of data sharing (discussed in greater detail later).[18]

DIFFERENTIAL IMPACT

All technological innovations—in fact all policy decisions—will have differential impacts and the sharing of data is no different. Thus, we should avoid measuring risks by the *average* benefits and harms. The risks of re-identification do not fall on each person in a dataset equally. It is all but guaranteed that they will disproportionately affect the most vulnerable.[19] For example, marginalised members of the population may be at greater risk of their data being re-identified (because they are more likely to have characteristics or collections of characteristics that are unusual), which may in turn have greater negative consequence for them given the other kinds of harms and risks they are already subject to. They are also the group about whom the government has the most data. When it is possible to do so, the more privileged in society are able to choose to have greater privacy instead of engaging with public services, an option not available to the most marginalised and disadvantaged.

[18] Bert-Jaap Koops, Bryce Clayton Newell, Tjerk Timan, Ivan Skorvanek, Tomislav Chokrevski, Masa Galic, 'A Typology of Privacy', 38 U. PA. J. INT'L L. 483 (2017). For more complex accounts of privacy, see also Daniel J. Solove 'A Taxonomy of Privacy', *University of Pennsylvania Law Review* 154(3) (January 2006).
[19] Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York, NY: St. Martin's Press, 2017).

SUGGESTED IMPROVEMENTS

1. The Data Sharing Principles should more explicitly detail and acknowledge the importance of preserving different kinds of privacy.
2. Data minimisation principles should be applied not just to the *quantity* of data released, but also to its *quality*. Unadulterated data should be shared only when the public interest case is compelling, and unadulterated data are strictly necessary. Otherwise, well-tested methods for anonymisation, such as differential privacy, should be implemented. And those methods should be transparent and so open to assessment.[20]
3. Trade-offs between privacy and public interest should be transparent. There should be open debate about what kinds of public interest are worth risks to privacy. Privacy should not be risked for the sake of profit, or for research that is not in the public interest.
4. DS&R provides a framework to bypass secrecy provisions (where the Purpose Test and risk management have been done). There should be a requirement that those requesting data consider the motivation for secrecy provisions and provide a justification for why that motivation no longer applies or can be overridden.
5. Assessments of the risks of sharing *this* dataset should take into account which *other* datasets have been shared, or might be shared. It may then be necessary to limit the datasets that can be held in combination by one accredited user—paying attention to the possibility that a user may have access to other relevant datasets through other legislative means.

CONSENT

DATA OWNERSHIP

Throughout the report, data is repeatedly characterised as a 'national resource'. This characterisation makes the position of the DS&R with respect to consent—that individual consent is not required for the sharing of data—more plausible. Individual consent is not relevant if data can be owned and the government owns it. However, this characterisation is misleading in two ways. First, data cannot be owned in the way, for example, copper and oil can be owned. Second, while some data collected by government does 'belong' to it, in some sense, other data extracted from citizens as a necessary condition of their receiving public services are less plausibly common property. We recommend that this terminology be avoided.

There is much work to do in establishing what meaningful consent would mean where data is concerned. We understand the ONDC's argument that people do not have an absolute right to control 'their' data. We do not think that we have such absolute and inalienable rights over data. Individual consent *can* be overridden in certain specific contexts. This is often the argument made about the primary uses of data: consent is not necessarily needed to collect data when that data is necessary for important public goods to be realised.

[20] This is analogous to Kerckhoff's principle that a cryptosystem should be secure even if everything about the system, except the key, is public knowledge. https://en.wikipedia.org/wiki/Kerckhoffs%27s_principle.

However, we do have a *presumptive* claim over data about us, especially when it comes to *secondary* uses of our data. To gather data from us without our consent, and then to use it for purposes quite different from what it was gathered for, again without consent, constitutes a significant restriction on our freedom and demands a commensurately weighty justification. Whether that justification is available depends on the particular use to which that data will be put.

CONTEXTUAL INTEGRITY

Currently, the Australian Privacy Principles severely limit the sharing of data, sharply distinguishing between primary and secondary uses of personal data.[21] This approach more closely respects that personal information is revealed in a specific context for a specific purpose and that the norms of appropriateness and distribution are relative to that context. Respecting these facts is an essential part of what Nissenbaum calls *contextual integrity*.[22] This entails that "personal information revealed in a particular context is always tagged with that context and never 'up for grabs'."[23] Thus, the Data Availability and Use Report (2017) that sparked this legislative change offers an overly simplistic view when it claims that: (1) people already share their data with businesses with almost no concern, (2) people trust the government much more than businesses and that therefore (3) this gives the government a mandate to share the data it has on its citizens. This disguises the contexts in which this information is given and the norms of distribution that applied when people gave up this information. For example, someone might very happily leave the Google Maps location tracker on their phone turned on while visiting a gay bar and also, in general, trust the government to have their best interests at heart, at least in comparison with Google. However, this does not entail that they must therefore be comfortable revealing their sexual orientation to all government departments. Similarly, my agreeing to share my data with a government service for a specific purpose does not mean that I have implicitly consented to share that data with any service for any of the purposes outlined in the DS&R. This is especially so given the breadth of these purposes, including as they do to 'inform' government policy and 'enable' research.

BIAS AND THE PUBLIC INTEREST

Individual consent would provide a direct assessment of approval for data gathered in one context to be shared in another and thus whether the public agrees that the purposes in question

[21] https://www.oaic.gov.au/privacy/australian-privacy-principles-guidelines/ especially with respect to personal information.
[22] Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, (Stanford University Press, 2010).
[23] Helen Nissenbaum, 'Privacy as Contextual Integrity', *Washington Law Review* (2004), 125.

are in the public interest. However, the DS&R rejects a requirement for consent grounding their argument in two technocratic claims: (1) that allowing people to withdraw their data would lead to biased datasets and (2) that the interests of the public outweigh the interests of the individual. There is reason to doubt both of these claims. No doubt making data collection voluntary can sometimes undermine the statistical value of a dataset, but this is clearly not always true. Furthermore, it is less clear that the risks of bias outweigh the interests of individuals.

In other countries a different balance has been reached. The NHS in the UK allows patients to opt out of their medical data being used for 'purposes other than their own care and treatment.'[24] Both the NHS and the GDPR[25] allow for exceptions to the requirement for consent, but they are tightly constrained: the public interest basis requires that the processing of data is *necessary* for the performance of a task or function carried out in the public interest that has a clear basis in law. Specifically, if the controller could reasonably perform the tasks or exercise its powers in a way that does not require the processing of this data then the 'public interest' basis does not apply.[26] This is different from arguing that consent can be disregarded in general, because allowing people to opt out would skew datasets.

DIFFERENTIAL REQUIREMENTS FOR CONSENT

We take seriously the concerns that have led the DS&R to reject the need for consent: that securing individual consent for every secondary use would be infeasible, especially if consent is to be meaningful, and that it would often lead to highly impoverished datasets that would be unusable for meaningful improvements to public service. However, we reject that consent should be done away with entirely. Instead of a 'one-size-fits-all' approach to consent in the context of data sharing —that is, that it is never needed—we propose the following *differential* approach.

First, we distinguish the contexts of sharing within the APS and sharing with third-parties.

Second, within the APS, we recommend as a minimum that there is a default prohibition on using integrated datasets, as this is where the greatest risks to privacy as secrecy and privacy as autonomy lie. If permission to use integrated datasets is granted *without consent* it should only be for purposes more narrowly defined than the current Purpose Test: namely, for improving the impacts of policies and programs on whole cohorts and to improve service delivery for individuals.[27] Furthermore, we encourage further specificity as to the scope and nature these

[24] https://www.nhs.uk/using-the-nhs/about-the-nhs/sharing-your-health-records/
[25] https://gdpr-info.eu/art-6-gdpr/
[26] Note also that in the other legal basis in the case of the GDPR 'legitimate interest' individuals have a right of erasure of their data.
[27] Note that some of the authors of this paper maintain that consent should be required for *all* secondary purposes. However, if this government goes ahead with the proposal that consent is not always required for the sharing of data, then the following requirements add needed protection for individual privacy understood as secrecy and autonomy as well as involvement and oversight.

purposes. For instance, they should exclude the use of data to influence people's behaviour until a time when this specific purpose can be subject to greater public scrutiny and consensus. This would allow for the protection of people's autonomy in the absence of their consent.

Third, when the government shares data with third parties, the purposes for which this data might be put should be broken into broad categories. Relative to these *categories* (i.e. not on a case-by-case basis) consent should be sought. For high value purposes with clear benefits to individuals, such as important improvements to health and well-being, we recommend implementing a simple opt-out scheme.[28] For the use of integrated data for more contentious purposes, opting in should be required.

This proposal respects contextual integrity as it gives consideration to the degree of difference between the purposes and contexts of the *collection* of data and those of *sharing* the data, placing greater requirements on sharing data *outside* government than on sharing *inside* government. Furthermore, it acknowledges that the public should be part of the on-going and dynamic process of data sharing.

SUGGESTED IMPROVEMENTS

1. DS&R should acknowledge that people do have a presumptive claim over *some* data about them. More work needs to be done on *which* data that includes, as well as what kind of claim it is—to be informed, to access, to rectification, to erasure, to data portability, and so on.[29]
2. DS&R should acknowledge a higher bar for justifying the sharing of data that was involuntarily collected (including contexts where data was compelled and those where opting out was not feasible given the government's monopoly on services) from that which is voluntarily given.
3. Further explanation and justification needs to be provided of in exactly what circumstances allowing opt out would impoverish the data to such an extent that consent is not needed. We simply do not accept that the public interest always outweighs the data rights that a data subject might have.
4. The Differential Requirement Approach to consent should be adopted whereby:
    a. Intra-government sharing for purposes more tightly constrained than the current purpose test do not need consent;
    b. For purposes outside of those detailed in (a), an opt out mechanism for intra-government sharing;
    c. For purposes of high value and clear or direct benefit, the sharing of data with third parties must allow for opting out;
    d. Sharing data with third parties must require opting in for more contentious purposes.

[28] This is the model the NHS uses, whereby individuals are able to decide whether their data can be used for medical research.
[29] https://ico.org.uk/your-data-matters/

TRUST

Clearly public acceptance of these reforms requires trust and the DS&R frequently mentions the importance of 'building trust'. However, as Onora O'Neill points out, what is needed is not more *trust* from the public but for our institutions to be more *trustworthy*.[30] Trust without trustworthiness is foolish. The duty is then on the system of data collection and sharing to be trustworthy and to demonstrate this in order that the public can form *appropriate* trust.

TRANSPARENCY, ACCOUNTABILITY AND AUDITABILITY

We welcome the emphasis on transparency in the DS&R—in particular making Data Sharing Agreements public. We agree that this is a necessary condition for public trust. But it is not sufficient for accountability, that is, for robust trust*worthiness* of the system. The DS&R is currently missing key elements for trustworthiness. It does not allow, for example, for a simple way for individuals to review the data the government has on them to ensure that what is being shared is accurate. It does not allow for the review of information that is *inferred* by departments, which is especially important when that inferred information is also shared. We need to know more about *what kinds of data* the government holds about us, *what can be inferred from that data*, and *what use they can and will be put to*, and *by whom*. Transparency in terms of outputs (what Data Sharing Agreements have been reached) is not sufficient: it is also important to have accountability, in the form of greater public debate and input into processes (for example, those by which 'accredited users' are selected) as well as mechanisms in place for holding decision-makers to account and for challenging their decisions. In order to increase accountability, *auditability* should be more robustly built in as well.

Auditability of information use requires a chain of provenance built from specific links:

(1) Information about the origins of the data: when, by whom, and why the data was originally collected;
(2) Information about the movement of the data: when, by whom, and why the data was shared, who got to decide, and what restrictions governed that decision;
(3) Information about the processing of the data: what was done to the data by whom and why; and
(4) Information about safeguards and oversight: who ensures that commitments regarding the data are upheld and that restrictions are not breached.

We recommend putting in place a 'policy-aware transaction log'[31] detailing each information-use event that may be relevant to the assessment of accountability.

[30] For a brief and informal explanation of her argument see Onora O'Neill's TED talk: https://www.ted.com/talks/onora_o_neill_what_we_don_t_understand_about_trust/discussion.
[31] Weitzer et al. 'Information Accountability', Communications of the ACM 51(6) (June 2008), 86.

ELEMENTS OF TRUST

David Danks et al. defines appropriate trust in this way: the Trustor has justified beliefs that the Trustee has suitable dispositions to care for the Trustor's values.[32]

Crucially, appropriate trust requires *justified* beliefs, that is, beliefs grounded in adequate evidence about the dispositions of the trustee. This can be achieved through openness and transparency across all aspects of the data collection, sharing and processing progression, but must include transparency about the value judgments made by the trustee—for example, when determining that a given use of our data is beneficial enough to warrant the risks of sharing.

To be confident not only about what the trustee actually does, but about their dispositions, we also need suitable guard-rails in place. This of course means that any instance of misconduct would lead to data access being withdrawn, but also implies the importance of oversight at every stage by an independent party. One area of serious concern is that the National Data Commissioner has no ability to challenge the decisions of the Data Custodians in a specific instance. This is no doubt beyond the scope of the DS&R paper, but is nonetheless a cause for concern.

Finally, trust is only appropriate when the trustee cares for *the trustor's values*. Stating that sharing of people's data is in the public interest is a valuable start. But we need to show that it is in the interests of, in particular, those who are put at risk and who are denied a say. This would put stringent limits around the purposes for which data could be shared. It should not be assumed that trust is transitive: just because you trust your boss, doesn't mean that you should similarly trust everyone they trust.

INFORMATIONAL FRICTION AND RISK-SENSITIVITY

An important means by which to build trust is to take more seriously the concerns about privacy and consent. This does not entail that privacy can never be risked for the public interest nor that consent must always be required in order to authorise the sharing of data. But it does mean that greater space in the process must be given to a critical perspective, especially over those parts of the system that we, as individuals, cannot see and cannot control.

By providing a mechanism for 'overcoming' existing secrecy provisions, by demanding a move towards a 'pro-disclosure culture', and by enjoining the National Data Commissioner to campaign for greater sharing, the DS&R leaves little provision for institutional and systematic forces

[32] Danks articulated this particular formulation in his public lecture *Humanising Machines and Mechanised Humans: Challenges and Opportunities*, ANU, 16 September 2019. See also: David Danks, *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, p521-2; Emily LaRosa and David Danks, 'Impacts on Trust of Healthcare AI, *AIES 2018*, 10.1145/3278721.3278771; Heather Roff and David Danks, '"Trust but Verify": The difficulty of trusting autonomous weapons systems', Journal of Military Ethics, 17(1), 2-20, 10.1080/15027570.2018.1481907; Annette Baier 'Trust and Antitrust' *Ethics* 96(2) (Jan., 1986), 231-260.

that oppose the flow of information. Such forces—forces of what Luciano Floridi calls 'informational fiction'—are essential for protecting informational privacy, especially in the absence of consent.[33] Providing a role for such voices acknowledges risk-*sensitivity* as an *appropriate* attitude towards sharing data. The era of big data has taught us, again and again, that there will always be unintended consequences, things will often go wrong, and when they go wrong, they will be worst for the most vulnerable. Thus, while we agree with moving away from a model where *all* parties in the data sharing process are risk-averse, the other extreme whereby *everyone* in a system is encouraged to focus on the benefits of data sharing will almost inevitably lead to vulnerabilities in the process being overlooked.

SUGGESTED IMPROVEMENTS

1. The emphasis should be on creating trustworthy institutions first, in order to build public trust.
2. Greater emphasis should be placed on transparency, especially with respect to trade-offs being made.
3. Means of achieving accountability should be added.
4. A chain of provenance and a transaction log should be provided to secure the auditability of data sharing.
5. There should be an acknowledgement that sensitivity and attentiveness to risk can be not only appropriate but also a necessary part of safeguarding individuals' rights to privacy, especially given that under the proposed legislation there will be great pressure to share. Thus, risk-sensitivity should be built into the legislation to increase informational friction and thus informational privacy.
6. Greater *independent* oversight.
7. There should be more robust procedures for when things go wrong internally. Provision is made for notification of data breaches, but provision should also be made for internal abuses. For example, whistleblowing ought to be encourage and a procedure in place for dynamic review rather than only a system of approval.
8. Space should be provided for the public to engage in the ongoing process of data sharing, even in the absence of individual consent: for example, a way for the public to appeal or challenge accreditation or Data Sharing Agreements.

CONCLUSION

The three values raised in this response are intertwined. Consent is one way we can control our data and thus one way in which we seek to protect our privacy. We are only happy to give up such control to a system when that system demonstrates itself to be trustworthy. However, the very relations of trust are normally established through interactions of informed consent.[34] The

[33] Luciano Floridi defines informational privacy as informational *friction*: the forces that oppose the flow of information. Luciano Floridi, The ontological interpretation of informational privacy. *Ethics and Information Technology* 7(4) (2005), 185-200.
[34] Onora O'Neill, 'Informed consent is one of the hallmarks of trust between strangers', Reith Lectures, 22.

DS&R should therefore not reject wholesale the need for informed consent when it comes to the sharing of data. A one-size-fits-all approach must be avoided and a dynamic and subtle system put in place: one that is robust, transparent, accountable, and auditable, where different levels of justifications, safeguards, restrictions as well as requirements for consent apply that reflect the differences in the contexts in which the data is collected, the levels of risk different kinds of data pose, the kinds of harm at stake, the nature of the organisations making the request, the purposes for which the data is used and the value of sharing the data. It is only by acknowledging and being sensitive to these differences that trust can be built in the system as a whole—and the government needs the approval and trust of the public for this entire edifice to be built and to succeed.

To discuss or clarify any aspect of this submission, please contact Dr Claire Benn at Claire.Benn@anu.edu.au or 0481023387.