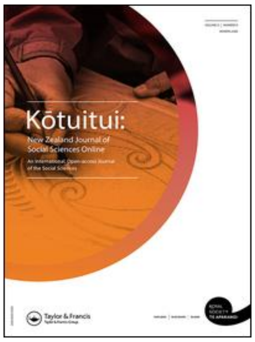

Kōtuitui: New Zealand Journal of Social Sciences Online

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tnzk20

Understanding the data-sharing debate in the context of Aotearoa/New Zealand: a narrative review on the perspectives of funders, publishers/journals, researchers, participants and Māori collectives

Jane Reeves, Gareth J. Treharne, Reremoana Theodore (Ngāpuhi and Te Arawa), Will Edwards (Taranaki Iwi, Ngāruahine, Tāngahoe, Pakakohi, Ngāti Ruanui), Mihi Ratima (Whakatōhea and Ngāti Awa) & Richie Poulton

To cite this article: Jane Reeves, Gareth J. Treharne, Reremoana Theodore (Ngāpuhi and Te Arawa), Will Edwards (Taranaki Iwi, Ngāruahine, Tāngahoe, Pakakohi, Ngāti Ruanui), Mihi Ratima (Whakatōhea and Ngāti Awa) & Richie Poulton (2021): Understanding the data-sharing debate in the context of Aotearoa/New Zealand: a narrative review on the perspectives of funders, publishers/journals, researchers, participants and Māori collectives, Kōtuitui: New Zealand Journal of Social Sciences Online, DOI: 10.1080/1177083X.2021.1922465

To link to this article: https://doi.org/10.1080/1177083X.2021.1922465

KOTUITUI: NEW ZEALAND JOURNAL OF SOCIAL SCIENCES ONLINE
https://doi.org/10.1080/1177083X.2021.1922465

REVIEW ARTICLE

OPEN ACCESS

Understanding the data-sharing debate in the context of Aotearoa/New Zealand: a narrative review on the perspectives of funders, publishers/journals, researchers, participants and Māori collectives

Jane Reeves[a], Gareth J. Treharne[a], Reremoana Theodore (Ngāpuhi and Te Arawa)[b], Will Edwards (Taranaki Iwi, Ngāruahine, Tāngahoe, Pakakohi, Ngāti Ruanui)[c], Mihi Ratima (Whakatōhea and Ngāti Awa)[d] and Richie Poulton[e]

[a]Department of Psychology, University of Otago, Dunedin, Aotearoa/New Zealand; [b]National Centre for Lifecourse Research, University of Otago, Dunedin, Aotearoa/New Zealand; [c]Taumata Associates, New Plymouth, Aotearoa/New Zealand; [d]Te Pou Tiringa, New Plymouth, Aotearoa/New Zealand; [e]Dunedin Multidisciplinary Health and Development Research Unit, Department of Psychology, University of Otago, Dunedin, Aotearoa/New Zealand

ABSTRACT

This review outlines current debates about the sharing of research data, with a focus on relevance for Aotearoa/New Zealand. Recent years have seen increasingly frequent calls for public sharing of data from funders and publishers/journals in particular. Past research has suggested that researchers tend to agree that any detriments of data-sharing are outweighed by benefits for transparency and progress. We summarise trends across past research into perspectives of funders, publishers/journals, and researchers on data-sharing before raising three considerations. Firstly, past research on data-sharing has tended to overlook the potential implications of data-sharing for participants. We review the small body of research on participant perspectives. This research has conceptualised participants as a homogenous group without theorising how participants make sense of data-sharing. Secondly, perspectives on data-sharing vary depending on the methodology being applied, and we raise some specific considerations when data-sharing is proposed in long-term longitudinal research such as the Dunedin Study. Thirdly, Indigenous perspectives on data-sharing must be central to all research into data-sharing with any of the stakeholder groups, and we review existing research on data sovereignty in relation to data-sharing in Aotearoa/New Zealand and globally. We conclude by summarising a series of tensions between stakeholders in the data-sharing debate.

ARTICLE HISTORY
Received 9 July 2020
Accepted 22 April 2021

KEYWORDS
Confidentiality; research ethics; data-sharing; longitudinal research; data sovereignty

Introduction and aims

The data that participants share with researchers are the foundation upon which knowledge and evidence-based forms of practice are built (Molloy 2011). Human participants

---

CONTACT Richie Poulton richie.poulton@otago.ac.nz

take part in hundreds of academic research projects every year in Aotearoa/New Zealand, and millions of pieces of data about these participants are stored in electronic databases. There are established ethical guidelines and legal requirements about access to data, but these are in tension with the increasing number of funders, publishers/journals, and other bodies requesting or requiring that researchers share the raw data upon which their findings are based within many fields, including genetic, clinical and social science research (Fecher et al. 2015). The rate with which data-sharing policies have been enacted has left many unanswered questions about how different stakeholders in academic research interpret and apply data-sharing principles.

The majority of the existing research on data-sharing is framed exclusively from the perspectives of funders or researchers (Aitken et al. 2016), without directly considering how data-sharing may affect participants. Moreover, the limited research that does specifically address the perspectives of participants on data-sharing tends to homogenise participants as a collective group, despite data-sharing being highly contextualised by the methodology applied and the location of the research; for example, qualitative interviews would generally gather data on individual participants to a greater depth and level of detail than a survey and may require more ethical consideration. Contextual considerations mean that data-sharing is particularly complex for ongoing members of longitudinal cohort studies such as the Dunedin Multidisciplinary Health and Development Study (the Dunedin Study; see Poulton et al. 2015) as opposed to one-off studies. In addition, there are specific considerations about data-sharing for Indigenous communities that must account for data sovereignty in light of global directives on Indigenous rights and treaties such as Te Tiriti o Waitangi in Aotearoa/New Zealand.

For this narrative review, we used a combination of methods to retrieve relevant articles. For the coverage of the broad considerations of data-sharing we retrieved articles via search engines such as Medline, Scopus, and Google Scholar; when we searched for relevant literature regarding perceptions of data-sharing held by different stakeholders, we also traced through reference lists of seminal articles. This review begins by outlining the reasons behind increased calls for data-sharing, before considering the alleged benefits and detriments that have been widely publicised. Following this, the bodies of literature related to the perspectives on data-sharing are compared among four different stakeholder groups: funders, publishers/journals, researchers, and participants. We conclude the review by discussing ongoing tensions around data-sharing and the implications for participants as a collective group, and for more specific communities of participants.

The aims of the literature review are:

(1) To synthesise existing research on the perspectives of funders, publishers/journals, and researchers as the three groups of stakeholders in most of the research to-date about data-sharing, and to identify gaps in the current literature about data-sharing.
(2) To amplify the perspectives and concerns about data-sharing from research participants themselves, and note the particular challenges that data-sharing may engender in Aotearoa/New Zealand so as to inform the decision making of researchers, publishers/journals, funders, and wider policy-makers who are considering implementing data-sharing policies.

(3) To challenge Western assumptions about researcher ownership of data by outlining research on data sovereignty, with a specific focus on research led by Māori or other Indigenous researchers and the articulated views of communities related to their experiences and perceptions of data-sharing.

The rationale for data-sharing in social science and health research

The increasing pressures to share data in social science and health research are in part a response to what has been referred to as an academic and public crisis of confidence in the replicability of research findings (Ioannidis 2005; Pashler and Wagenmakers 2012; Roediger 2012; Open Science Collaboration 2015; Vanpaemel et al. 2015; Rouder 2016). These concerns have coincided with high-profile cases of data fraud and/or implausible findings and the acknowledgment that publication bias is exacerbated by a culture of valuing and publishing positive statistical findings over null findings (Bem 2011; Giner-Sorolla 2012; Pashler and Wagenmakers 2012; Yong 2012). In order to combat the replication crisis, data-sharing has been advocated for by an approach known as ‘open science’, which refers to making each component of scientific research as transparent as possible to external researchers and the public (Rouder 2016).

Open science is used as an umbrella term, and data-sharing is the specific process of making raw data accessible to researchers outside those collecting the data (Allison and Cooper 1992; Rouder 2016). Nested within data-sharing is the closely related concept of data-linkage, which refers to the practice of integrating shared data from multiple sources about the same individuals, usually for the purposes of extending the initial research findings and creating a more comprehensive dataset (Harron et al. 2017). Data-linkage is increasingly facilitated by nationwide systems in various countries. For example, in Aotearoa/New Zealand the Integrated Data Infrastructure (IDI) is held by Statistics New Zealand Tatauranga Aotearoa. The IDI provides a framework for linking ‘micro-data’ from numerous sources (e.g. the census, justice, healthcare) and the sharing of anonymised data for approved research purposes on a fee-to-access basis (Statistics New Zealand Tatauranga Aotearoa 2018; Theodore et al. 2019).

In this review, we focus on sharing of data from research. Data-sharing of research data can be divided into distinct categories of born-open data and data-on-request (Rouder 2016). Born-open refers to an automated system wherein all collected data are uploaded immediately to an online storage system, including any data from failed or otherwise abandoned pilot studies or test runs. This is contrasted to data-on-request, which refers to the practice of sharing data conditionally after the researcher is explicitly asked to do so. The majority of the articles included in the present review describe practices that reflect data-on-request data-sharing practices.

Perceived benefits of data-sharing

Many studies contributing to the data-sharing debate attempt to present a balanced account of the perceived benefits and disadvantages from the perspective of researchers, but tend to superficially examine the perspectives of other research stakeholders or postulate these perspectives without consultation. Nine studies with broad considerations of data-sharing were retrieved from searches of Medline and Scopus (Pearce and Smith 2011; Tenopir et al. 2011; Cheah et al. 2015; Fecher et al. 2015; Tellam et al. 2015; Hudson 2016; Rouder 2016; Tsai et al. 2016; Martone et al. 2018). Across the nine studies, there was a general consensus regarding researchers' perspectives on data-sharing, wherein researchers identified that data-sharing can promote an open science research culture and increase transparency in the research process. Researchers who participated in research on data-sharing felt that a call to open science encourages the collection and storage of better quality data to ensure judgements about data analysis are justified, as well as identifying mistakes before findings are implemented in clinical practice (Tenopir et al. 2011; Cheah et al. 2015; Tellam et al. 2015; Hudson 2016; Rouder 2016; Tsai et al. 2016; Martone et al. 2018). Other key benefits of data-sharing raised included an improved peer-review process by reviewing datasets in conjunction with respective manuscripts; the usefulness of data-sharing for students or emerging academics; greater opportunities for broader collaborations between researchers; and the potential for benefits to individuals with a variety of illnesses through contribution to accelerated scientific progress. In particular, Martone et al. (2018) highlighted that data-sharing can lead to the inclusion of all relevant information, rather than only including positive results in publications. But what concerns might outweigh these positive aspects of data-sharing?

Perceived disadvantages of data-sharing

A review of studies on researchers' perspectives on data-sharing conducted by Martone et al. (2018) found that identified disadvantages generally fell into one of two categories: the lack of utility in data-sharing and negative outcomes for researchers. In the first category, researchers have argued that due to the nuanced and complex nature of some research, the raw datasets will have no inherent value for re-analysis as they are too difficult to understand and interpret for anyone other than experts and perhaps only the original researchers themselves. Furthermore, researchers have expressed concerns that research may start relying too much on secondary analyses of previously collected data if data-sharing becomes implemented more widely. In the second category, researchers outlined three key arguments against data-sharing. A widely acknowledged disadvantage of data-sharing was the potential loss of reputation for researchers, occurring in one of two ways: (1) by another researcher identifying a genuine mistake in a shared dataset; or (2) another researcher deliberately manipulating the dataset to undermine the original work. It was posited that in more extreme cases, individuals may be targeted based on their research fields. It was also acknowledged that data-sharing could lead to other academics using data before the original researchers have had a chance to publish their own findings, a problem particularly relevant for emerging academics who may take longer to publish early in their career. Finally, Martone et al. (2018) asserted that the amount of effort required to organise and share data may simply outweigh the benefits of doing so.

Only two of the nine articles expressed concerns about the lack of consultation with participants (Pearce and Smith 2011; Hudson 2016). This raises questions about the perspectives of stakeholders other than researchers. Of particular note, none of the identified articles made any mention of Indigenous Data Sovereignty or Indigenous collectives concerns about big data or data-sharing movements. Disadvantages of data-sharing from the perspectives of Indigenous peoples have been clearly identified in terms of Indigenous knowledge and traditions being exploited, especially for the benefit of non-Indigenous researchers or groups; the lack of collective rights and privacy within existing eurocentric data ecosystems and thereby minimal protections for Indigenous collectives; and deliberate or unintentional (mis)interpretations of Indigenous data in conjunction with the absence of relevant historical, cultural, political, and social contextualisation (Rainie et al. 2019; Walter and Suina 2019; George et al. 2020; Kukutai et al. 2020; West et al. 2020).

Competing policies and guidelines of data-sharing

Data-sharing policies are variously defined and implemented by different academic organisations. In Aotearoa/New Zealand, academic bodies such as the New Zealand Psychological Society support researchers and practitioners. However, at the time of writing the New Zealand Psychological Society does not have a statement on their website regarding data-sharing, and their code of ethics is mainly concerned with clinical practice, for example noting that ‘Uninterpreted data from assessments is not normally released to persons who are not specifically trained in the use and interpretation of the instruments concerned’ (New Zealand Psychological Society 2012, 14). The American Psychological Association (APA 2010) currently promotes data-sharing by stating that researchers should share their raw data with editors or other researchers if requested, unless there is a strong reason not to. These guidelines do not provide a definitive framework about why data should or should not be routinely shared, and therefore researchers are required to consider data-sharing within wider ethical guidelines and local regulations.

Sets of ethical principles have been defined in various countries over the past 50 years, such as those outlined in the groundbreaking Belmont Report (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1978) which described three ethical principles which are still applied today. Researchers conducting studies with human participants in Aotearoa/New Zealand are required to apply for ethical approval from recognised ethics committees. Researchers are expected to uphold the principles of Te Tiriti o Waitangi (the Treaty is New Zealand’s foundation document formalising the relationship between the British Crown and Māori tribes) and this is subject to ethical review, though many non-Māori researchers continue to struggle to understand how the Treaty applies to their research (Gibson et al. 2020). The implication is that ethical research in Aotearoa/New Zealand respects the rights of Māori as Treaty partners and this should be reflected in, for example, culturally responsive research processes, contribution of research to equitable outcomes for Māori, and respecting Māori data sovereignty (Hudson and Russell 2009; Gibson et al. 2020). Hudson et al. (2020) noted that relying on ethics committee recommendations to determine access to Indigenous data for data sharing is not appropriate. This is because their principles may not align with ethical principles determined by Indigenous communities themselves, which may be a foundation for Indigenous peoples own approaches to research such as Kaupapa Māori research (Smith 2012). Furthermore, ethics committees often lack representation of Indigenous communities. In this context, it is important to consider guidelines developed by Indigenous peoples that can support decisions about data access and data-sharing in Aotearoa/New Zealand.

In recent years there has been a push for a consistent data-sharing and stewardship framework, resulting in the publication of the FAIR principles (Wilkinson et al. 2016). These principles provide ways of managing data at a local and global levels by addressing four key elements of access to digital data and metadata, defined as data describing the characteristics of other existing data (Wilkinson et al. 2016). The first principle is that (meta)data should be Findable (F), wherein a unique identifier assigned to (meta)data should be easily locatable and the associated descriptions are accurately and explicitly linked to the identifier. The second principle is that (meta)data must be Accessible (A), in that data must be retrievable with the use of the identifier, in a method that is in accordance with existing open-access data principles. The third principle of data being Interoperable (I) states that the language used in (meta)data descriptions is consistent with wider systems of knowledge representations and in accordance with the other FAIR principles. The final principle is that data be Reusable (R), noting that (meta)data should be described with appropriate labels and comprehensive detail given about the collection, stewardship, and dissemination of the (meta)data (Wilkinson et al. 2016). The goal of the FAIR framework is to empower different stakeholders to share and manage their own data, as well as facilitate future data-sharing and digital innovation in data management. Wilkinson et al. (2016) assert that the FAIR principles are a necessary condition for scientific innovation within academic communities. However, the FAIR principles overlook the importance of including research participants as active and equal stakeholders, and also overlook Indigenous collectives altogether. As a response, the Global Indigenous Data Alliance (2019) proposed the CARE principles as a means of complementing the FAIR principles in a way which has relevance for Indigenous individuals and collectives.

Like the FAIR principles, the CARE principles outline four key focuses, with each aspect related to the promotion and maintenance of Indigenous data sovereignty (Research Data Alliance International Indigenous Data Sovereignty Interest Group 2019). Collective benefit (C) requires that any value derived from Indigenous data should benefit those Indigenous communities in an equitable manner, and that the management and storage of such data should be organised in a way that is accessible for Indigenous peoples. The Authority to control (A) outlines that consent must be gained at the appropriate level and that the parameters of data-sharing (if allowed) are clearly demarcated by the Indigenous collectives participating in research; furthermore, this principle promotes Indigenous collectives asserting sovereignty over their data by determining and defining the governance of their data. The third principle of Responsibility (R) emphasises the importance of establishing connections and building relationships between Indigenous communities and non-Indigenous researchers *before* discussions about data-sharing and reuse occur, and that (meta)data and research output must be grounded within Indigenous frameworks. This principle also highlights that in order to create and promote spaces for Indigenous peoples and knowledge, the decolonisation of higher research agencies is a necessary step. The last CARE principle is Ethics (E) and describes that potential harm should be minimised and every opportunity maximised; conclusions should not be deficit-based; benefits should be directed to the Indigenous communities participating in the research; Indigenous collectives should be included in ethics decision-making; and potential future uses of (meta)data should be considered within the relevant Indigenous framework (Research Data Alliance International Indigenous Data Sovereignty Interest Group 2019).

In Aotearoa/New Zealand, understanding Māori and Indigenous data sovereignty is fundamental to discussions about data that have been collected on Māori participants, whānau (families), communities, hapū (subtribes) and iwi (tribes) (Taylor and Kukutai 2017). Indigenous data sovereignty is defined as '… the right of Indigenous Peoples to own, control, access and possess data that derive from them, and which pertain to their members, knowledge system, customs or territories' (International Work Group for Indigenous Affairs (IWGIA) 2020). Indigenous data sovereignty applies across dimensions of ethical, legal, and practical considerations including intellectual property, collective rights and benefits, and guardianship (Taylor and Kukutai 2017). Te Mana Raraunga, the Māori Data Sovereignty Network, state that Māori data sovereignty recognises that Māori data should be subject to Māori governance and be used to support the realisation of Māori and iwi aspirations (Te Mana Raraunga n.d). To date, data collected on Māori individuals or collectives have been primarily used to undertake research about Māori as opposed to 'by Māori for Māori'. This often results in deficit-based approaches and analyses, and the exploitation of culture, knowledge and resources (Smith 2012).

Data-sharing in research practice

While data-sharing policies continue to be introduced by funders and publishers/journals, widespread uptake of data-sharing is yet to be reflected in practice. It is crucial to understand how the perspectives of funders, publishers/journals, and researchers intersect with each other in order to better understand the dialogue between the perceptions and practices around data-sharing.

Funders as drivers of data-sharing

We identified four articles directly related to funders' perspectives on data-sharing. The first of these four articles described the outcomes of a data-sharing workshop, a culmination of efforts by funders and researchers to understand and reduce barriers to data-sharing in 'low and middle income' countries (Carr and Littler 2015, 315). The second article presented the shared perspective of four major funding institutes in the United States (US) and the United Kingdom (UK) that were all active proponents of data-sharing, and critical that data-sharing was not more widely implemented. However, the funding institutes acknowledged that the incentives to encourage data-sharing among researchers were inadequate (Kiley et al. 2017). These findings suggest that funders are aware of some of the barriers to data-sharing that researchers experience (Carr and Littler 2015; Kiley et al. 2017). However, the majority of data-sharing policies already in place do not seem to address these barriers.

In spite of the increasingly frequent top-down pressure from funders for data-sharing to be implemented more widely, policies differ across funders and the majority do not address data-sharing comprehensively enough to ensure its regular implementation (Molloy 2011; Shah et al. 2019). A review of data-sharing policies of 66 US funders found that 80% of these policies required, or encouraged, researchers to submit written proposals detailing their prospective data-sharing plans (Williams et al. 2017). Analysis of the policy statements identified 43 components required to comprehensively address the different aspects of data-sharing; compared against the original policies, none of the funders had policies that addressed all 43 components, with the highest addressing only 72% of the criteria (Williams et al. 2017).

In sum, funding bodies who require consideration of data-sharing in research proposals seem to place more emphasis on broad statements, rather than requiring specific, detailed steps to be met (Williams et al. 2017). This lack of specific policy around data-sharing has the potential to leave key decisions about data-sharing up to the discretion of individual publishers/journals and researchers. Research has revealed a lack of consistency in data-sharing statements among funders, which leaves the implementation primarily in the hands of researchers (Kaye and Hawkins 2014; Shah et al. 2019). The net result is a set of processes that do not serve the interests of Māori or other Indigenous peoples. Simultaneously, broad statements about data-sharing are open to interpretation, and this raises questions about the views of other research stakeholders who are generally not consulted about data-sharing.

Publishers/journals as drivers of data-sharing

Many academic journals now have data-sharing policies in their instructions for authors that may encourage, or even require, data-sharing and such policies are being introduced by more publishers/journals. However, widespread uptake of data-sharing is yet to occur (Nosek et al. 2015; Federer et al. 2018). There is mixed evidence about whether journal policies effectively facilitate data-sharing, with some research indicating that increased data-sharing is dependent on the pragmatism of journals and their policies across different research fields (Sturges et al. 2015; Crosas et al. 2018). For example, studies have indicated that in gene science research the highest rate of data-sharing occurs when articles are published in: (a) open-access journals (free for anyone to access online); (b) journals with clear and enforced data-sharing policies; and (c) journals with high citation rates (Piwowar and Chapman 2008; Piwowar and Chapman 2010; Piwowar 2011; Keerie et al. 2018).

Research on the data-sharing policies of top-ranked psychology journals has also indicated that the higher the ranking, the greater the likelihood that the journal enforces data-sharing (Crosas et al. 2018). Among these top-ranked psychology journals, 22% have policies requiring data-sharing, 38% strongly encourage data-sharing, while 40% outline no data-sharing policies (Crosas et al. 2018). Journal initiatives that seek to acknowledge researchers who share data have significantly increased the rates of publicly available datasets, compared to similar journals without such initiatives (Giofrè et al. 2017; Kidwell et al. 2016). However, it was noted that for journals piloting data-sharing policies, the initial increase in the rate of sharing stalled at some point, indicating that such initiatives may not address all the barriers to data-sharing (Giofrè et al. 2017). Moreover, research into the data-sharing behaviours of psychology researchers has indicated that neither the ease of data storage nor increased pressure from journals had an association with willingness to adopt data-sharing initiatives (Harper and Kim 2018). It appears that publishers/journals play a role in engendering data-sharing, but the complex interplay between journal policies and researchers' data-sharing practices is poorly understood.

Quantitative findings about researchers' perspectives on data-sharing

Despite the increase in data-sharing policies and the increased ease of data-sharing through technological advances, it appears that data-on-request models of data-sharing do not result in all requested datasets being received. Researchers have been expressing reticence about sharing data even before data-sharing began being implemented more widely, with studies indicating that as few as 24%–38% of requested datasets were received by researchers aiming to run replications (Wolins 1962; Craig and Reese 1973; Wicherts et al. 2006). More recently, Vanpaemel et al. (2015) evaluation has focused upon whether researchers have become more willing to share data since the more widespread calls for open science; however, research has indicated similar levels of hesitancy in sharing data. Vanpaemel et al. (2015) requested raw datasets that contained at least one p value and could be included in a Bayesian reanalysis from the authors of 394 articles published in four American Psychological Association journals. The request was framed as an attempt to run replication checks on the published findings. Affirmative responses were received from 38% of authors; 18% of contacted authors were unable or unwilling to share their data; and 41% did not respond to any data-sharing request. Similarly, Federer et al. (2018) analysed the data-sharing in the Public Library of Science journal PLoS One and found that only 20% of over 47,000 papers evaluated provided full datasets.

Researchers have explored why data-sharing is not yet common practice, with findings indicative of systemic problems in academia more broadly. Wicherts et al. (2011) found that researchers who were unwilling to share datasets were more likely to have published articles with weaker levels of significance or with methodological or statistical errors. Moreover, researchers appear to prefer publishing in journals where the data-sharing mandates are less specific (Fecher et al. 2015; Nosek et al. 2015). Researchers have also indicated that unwillingness to share requested data is related to the lack of financial compensation, insufficient time and resources to withdraw data, having incomplete or destroyed data, and a lack of knowledge about why the data were being requested (Craig and Reese 1973; Tenopir et al. 2011).

Researchers' experiential accounts of data-sharing

In addition to the insights into the limitations of open science and data-sharing provided by quantitative research, qualitative research provides insight into the experiences of researchers and participants relating to data-sharing (Shaw et al. 2008). Much of the past quantitative research has treated data-sharing as a homogenous practice in which researchers either share data or do not (Kurata et al. 2017). However, researchers' individualised perspectives and approaches to sharing of quantitative data have been found to be heterogeneous across seven qualitative studies on data-sharing that were retrieved from our literature searches. These studies can be split into two groups: (1) independent studies (Wallis et al. 2013; Levin et al. 2016; Kurata et al. 2017; Ho et al. 2018); and (2) studies within one large multisite research collaboration on data-sharing in low-/middle-income countries (Cheah et al. 2015; Denny et al. 2015; Hate et al. 2015).

Across these qualitative studies, researchers who participated reported that data-sharing could improve and promote scientific understanding and advancement. However, Levin et al. (2016) found that researchers in the UK perceived the data-sharing solution as 'over-hyped' (132) and chose to describe open data by what it could not achieve rather than what it could achieve, contradicting the perception that researchers are wholly positive about the process. Researchers also described wider data-sharing being impeded by the time required to adequately prepare data, the lack of funding to cover the time commitment, and the complications of accessing and uploading data, with other researchers identifying that current infrastructures acted as barriers by tending to isolate datasets rather than allowing researchers to link them easily (Levin et al. 2016; Ho et al. 2018). These concerns clash with the notion that increased data-sharing would exclusively lead to more comprehensive datasets through data-linkage (Harron et al. 2017).

The majority of the qualitative accounts described primary researchers' concerns over retaining 'ownership' of data they had collected (Wallis et al. 2013; Cheah et al. 2015; Denny et al. 2015; Levin et al. 2016; Kurata et al. 2017). Moreover, it was evident that there are cultural differences in researchers' perspectives about data-sharing (Kurata et al. 2017). For example, accounts from the large multisite study indicated that researchers were concerned about data-sharing reflecting neo-colonialist research practices, wherein data collected in low-/middle-income countries are shared but exclusively benefits researchers and populations in high-income countries. A lack of trust between researchers across geographical locations, and across cultures, were also identified as barriers by researchers based in Thailand and India (Cheah et al. 2015; Denny et al. 2015; Hate et al. 2015). These accounts are mirrored by UK researchers who reported that they should be helping researchers in lower-income countries by sharing their own data (Levin et al. 2016). Across the seven qualitative studies, researchers held diverse views on how data-sharing should be implemented; thus, data-sharing does not seem to be a straightforward solution to the replication crisis. Furthermore, this challenges how appropriate the implementation of data-sharing policies may be in research which has additional considerations.

The research we have summarised demonstrates how data-sharing is currently perceived and implemented by three groups of stakeholders: funders, publishers/journals, and researchers. Funders of research have varied policies about data-sharing and this means that publishers/journals do not have to implement data-sharing in any one specific way. Moreover, the relationships between journal policies about data-sharing and researcher practices are not linear. Although researchers play a central role in enacting data-sharing, most of the quantitative research reviewed indicates a reluctance to do so. The qualitative research reviewed suggests that there are more barriers to data-sharing than previously thought. However, it is important to note that up to this point participants have generally been neglected in data-sharing research.

Participants as overlooked stakeholders in the data-sharing debate

Quantitative research exploring views on data-sharing among participants has generally used their preferences for informed consent as a proxy of willingness to have their data shared. There are several consent options for individuals entering into a study as participants. However, these options may not always be offered; ethical regulations are country-specific, as are considerations of Indigenous communities who participate in research (Lemke et al. 2010; Treharne and Marx 2018). These include broad consent options, in which a prospective participant agrees for their data to be used in any unknown future project, decided by the researcher overseeing data management (Garrison et al. 2016). Within broad consent options, either opt-in or opt-out policies have been applied. Opt-in refers to participants actively giving consent for their data to be shared in future projects, whereas opt-out policies assume consent unless actively declined by the participant (Giesbertz et al. 2012). A third approach involves categorical consent, which refers to participants giving conditional agreement for prospective data-sharing, usually restricted to studying specific diseases (Garrison et al. 2016). Hudson (2016) emphasised that most of the literature on participants' perspectives on data-sharing has focused on suppositions made by researchers; it is crucial to separate these suppositions from what participants have indicated about data-sharing. Researchers have also been encouraged to view informed consent as a reciprocal relationship in which trust must be built to some extent *before* research participants give their informed consent to a study (Nairn et al. 2020).

Quantitative findings about participants' perspectives on data-sharing

The literature review carried out by Garrison et al. (2016) revealed that participants tend to be open to the idea of widespread data-sharing of their health data, but this is based on studies where broad consent towards data-sharing was the sole choice presented. When the participants were given alternative options, such as categorical consent, they were much less likely to favour broad consent to their health data being shared. For example, some participants who had already agreed to broad consent conditions have been found to be more reluctant to share their data under the same consent conditions in hypothetical scenarios (Oliver et al. 2012), while others have suggested that broad consent to data-sharing is appropriate as long as participants privacy is protected (Mello et al. 2018). Taken together, the research suggests that there is a need to support participants in coming to an understanding about what data-sharing and broad consent involves.

Participants have also indicated that factors such as medical history can influence their willingness to share data, with individuals who have a history of cancer being significantly more willing to share their genomic data compared to individuals with no history of cancer or relatives of individuals with cancer (Goodman et al. 2017). Moreover, participants are more open to their data being shared with academics at the same institute as the original research or with not-for-profit organisations, but less willing to have their data shared with external researchers, private companies, or nationwide federal databases in the US (Goodman et al. 2017; Goodman et al. 2018). Additionally, most of the quantitative literature on participants' perspectives on data-sharing has not accounted for how factors such as socioeconomic status, work status, or ethnicity may influence willingness to share health data (Garrison et al. 2016). Ethnicity is especially relevant in Aotearoa/New Zealand, where Māori participants and collectives are more reticent about data-sharing than Pākehā (a person of non-Māori descent, typically referring to a person of European descent) participants, due to previous instances of Māori sovereignty over data (among many things) being disregarded (Hudson, Southey, et al. 2016; Beaton et al. 2017).

Overall, past quantitative findings indicate that participants are conditionally positive about data-sharing, but that the current processes used to assess views about data-sharing may not be the most appropriate. The past quantitative research also highlights that conceptualisations of consent for data-sharing differ between participants and researchers. It is, therefore, useful to compare these quantitative findings to the available qualitative research to provide further insight into the perspectives of participants, and to identify whether all the barriers to data-sharing have been acknowledged.

Qualitative findings about participants' perspectives on data-sharing

A common trend across most qualitative studies we retrieved from our searches was that participants were generally altruistic and open to the idea of sharing their individual health data; however, they shared similar concerns to researchers about aspects of privacy (McGuire et al. 2008; Lemke et al. 2010; Mählmann et al. 2017). Participants in Zarate et al.'s (2016) study expanded on these concerns, citing hypothetical risks such as discrimination on the basis of health status from employers or insurance companies if data were shared. Other participants felt that governmental agencies overseeing the management of their data were inappropriate and could lead to misuse of data through breaches of privacy (Lemke et al. 2010).

There are mixed findings about whether participants feel they should retain some control over the data they contribute to research. Some participants have suggested that their right to disclose information about themselves to others would be lost if researchers shared health-related information in the public sphere (Jamal et al. 2014), while other participants have conceptualised data they contributed as belonging to the researcher who collected it (McGuire et al. 2008; Jamal et al. 2014). A trend evident across the qualitative studies was participants' desire for researchers to increase their transparency around data-sharing processes, consent options, and data ownership. Participants identified transparency as being concurrent with the need for increased public awareness about what constitutes data-sharing, as well as ensuring participants understand the nuances of consenting to research (McGuire et al. 2008; Jamal et al. 2014).

Presently, there is a lack of consideration of the additional complexities that specifying data ownership engenders, particularly with regards to working with Māori in Aotearoa/New Zealand. Mātauranga Māori and tikanga Māori have often been overlooked in Western research practices, and non-Māori researchers have neglected to consult with Māori collectives when studying data from or about Māori (Taniguchi et al. 2012). Research has indicated that within an ao Māori worldview, consenting to one's biological samples being used in genomic research should be framed as tākoha: a gift in which both parties involved recognise the inherent tapu and spiritual importance of data and to respect that certain restrictions will be imposed on the samples, particularly regarding who should be kaitiaki (guardians) of the data, challenging the Western view of data ownership (Hudson, Russell, et al. 2016; Beaton et al. 2017). Further research has indicated that reticence towards participating in genetic research is diminished when direct benefits may be seen for whānau, hapū or iwi (Hudson, Beaton, et al. 2016).

Understanding Indigenous perspectives about the use of genetic material and all aspects of data from or about Indigenous peoples, and prioritising Indigenous consultation and input are important features of a process of working with newly implemented data-sharing frameworks such as the CARE principles and existing definitions of data sovereignty (as defined by Indigenous peoples).

Participants have described the need for researchers to be culturally sensitive about data-sharing and consent processes, and that specific communities should have tailored data-sharing plans, with recommendations given about requesting different communities to review data-sharing plans prior to their implementation (McGuire et al. 2008; Lemke et al. 2010; Murad et al. 2017). This insight reinforces that a one size fits all approach to data-sharing will not be applicable across the whole population. For example, Indigenous peoples hold views that are different and even in opposition to those of majority populations and require alternative approaches. Understanding different communities and how they may be affected by data-sharing is crucial to informing decision-making about the implementation of data-sharing policies in Aotearoa/New Zealand as well as other locations.

Complexities of data-sharing for participants in longitudinal research

Participants in long-term longitudinal research such as the Dunedin Study (Poulton et al. 2015) occupy perhaps one of the most complex positions within the data-sharing debate. As funders and publishers/journals implement data-sharing policies more frequently and increase pressure to comply, directors of ongoing longitudinal studies (often covering periods of many decades) will have to critically evaluate if, and when, it is ethically appropriate to share their participants' data. A key consideration is the possibility of anonymised data allowing for re-identification of participants; the risk of re-identification increases the longer individuals are members of the same study (El Emam et al. 2011).

Eleven questions were identified as we began our literature search related to longitudinal research participants' perspectives of data-sharing:

(1) Can participants even remember what data was obtained about them 30–40 years prior?
(2) How should researchers treat data that was obtained from informants such as parents, school teachers, and/or significant others versus data collected from participants themselves?
(3) Should all data collected over many decades be made available on request or should data be restricted to that required to test specific hypotheses?
(4) Are all data to be treated as equally sensitive or are there different levels of sensitivity?
(5) How can the concept of Māori data sovereignty be applied in the context of ongoing longitudinal research projects like the Dunedin Study?
(6) Should self-reported data be treated the same as biological data?
(7) What *bona fides* are required for researchers to be granted access to data?
(8) Is any hypothesis sufficient to justify data access or should there be a requirement to demonstrate public benefit? And who would make this judgment?

(9) How can research leaders ensure high quality academic output standards are maintained in an ongoing (i.e. 'living') longitudinal study?
(10) How can researchers balance the needs and rights of long-term (e.g. of 30+ years duration) researchers versus new data requestors?
(11) What are the current study-specific policies and commitments to study participants around data access issues?

Researchers conducting longitudinal studies also must recognise that the public may be resistant to data-sharing, in contrast to the current trends in scientific communities. Some public policy research originating from Statistics New Zealand Tatauranga Aotearoa has indicated that people living in Aotearoa/New Zealand are open to data-sharing and integration when the processes are transparent and the benefits are clearly outlined, and this did not differ across ethnicities (Opus International Consultants 2015). However, more recent research has noted that people living in Aotearoa/New Zealand are reluctant to specifically have their personal health data shared with, and by, the government (Moore and Niemi 2016), drawing upon beliefs that personal health data belongs to the individual it represents. From our search of available literature, there have been no published articles exploring the perspectives of participants in longitudinal research about data-sharing, and studies that have explored comparable contexts generally focus on official administrative data like personal health data rather than research data. It is crucial to address this gap in the literature.

Some cross-sectional research has sought to identify whether participants of genomic and biobank studies feel they need to be re-contacted before data-sharing if the samples have already been collected (Ludman et al. 2010; McCormack et al. 2016). Biobanks represent an opportunity for accelerated investigation through the sharing of previously collected samples, but the nature of what and how biobank data are shared is intrinsically linked to participant trust and public understanding. Research conducted by Ludman et al. (2010) and McCormack et al. (2016) found that biobank participants, parents of children with rare diseases, and participants who had re-consented to their data being used for new research avenues are open to reanalyses. However, participants indicated a shared belief that researchers should always ask before reanalysing historical samples. This was emphasised when the new investigation was not outlined in the original consent agreement. Approaches where participants had to opt-out of their biobank data being reanalysed were generally not seen as appropriate alternatives. Participants felt that the decision-making for new analyses of biobank data should remain the patient or parent's decision alone, with no additional pressure being applied about the rarity of the disease or assurances of privacy (McCormack et al. 2016). Discussions also led to two schools of thought about whether children should be asked to re-consent to a study that their parents had previously given consent for. Some participants felt that the original consent would remain valid once the child became legally an adult, while others argued that not re-contacting compromised the autonomy of the children involved (McCormack et al. 2016). The preferences for consultation across these studies suggest that participants believe that maintaining autonomy over their samples is equally as important as the potential benefits that may arise from new analyses.

Studies exploring the perspectives of participants as stakeholders in the data-sharing debate have been conducted (Ludman et al. 2010; McCormack et al. 2016); however, none of these studies have examined data-sharing considerations in the context of ongoing longitudinal research, which may be very different to perspectives of individuals participating in a cross-sectional or one-off study. At present, there has not been enough research with members of specific communities to account for all the complexities of data-sharing from the perspective of participants. When considering the context of Aotearoa/New Zealand, there is no research which qualitatively explores how data-sharing is perceived by members of Indigenous communities involved in research projects, particularly where the data-sharing policies may change over time.

The evidence presented thus far is not comprehensive enough to yet be able to define best practices about data-sharing for participants in longitudinal research, and researchers have had to rely exclusively on tangentially related findings. Further to this, it would be premature to assume that the reviewed research encompasses all relevant perspectives, particularly as most of the presented research has treated the participant samples as relatively homogenous and there has generally been a lack of analysis by ethnicity or other factors reflective of participant diversity. There is much work to be done if data-sharing is to move beyond a one-size-fits-all approach.

Conclusions

The literature we have reviewed indicates that the data-sharing debate has implicit tensions that make it more difficult to judge whether data-sharing is appropriate to implement in any particular research project. We have identified six tensions that we believe are important for any researcher planning a study to consider, but especially for those in Aotearoa/New Zealand. We conclude by summarising these tensions and indicating how they might be addressed by identifying specific avenues of future research to increase understanding and by suggesting changes to current research practices that might effectively mitigate some of the existing tensions.

Tensions around data-sharing between funders, publishers/journals, Indigenous and non-Indigenous researchers

The trends in the literature indicate that funders, publishers/journals, Indigenous and non-Indigenous researchers are treated as mutually exclusive stakeholders in the data-sharing debate and almost always consulted in separate research projects on data-sharing, if at all (Pearce and Smith 2011; Tenopir et al. 2011; Cheah et al. 2015; Fecher et al. 2015; Tellam et al. 2015; Hudson 2016; Rouder 2016; Tsai et al. 2016; Martone et al. 2018). There is no consensus in the literature about whether funders or publishers/journals have the most sway in enacting data-sharing policies, and the lack of clarity leaves most of the practical implementation of data-sharing in the hands of researchers who experience barriers that clash with existing data-sharing policies. Further to this, Indigenous researchers and others committed to Indigenous data sovereignty will not be in a position to share all data despite potential implications for access to funding and avenues for publication.

The tension between and within these groups may be addressed by (1) increased communication between these stakeholders about how data-sharing is most appropriately implemented (including acknowledging when data sharing is not appropriate) for different types of research and (2) funders and publishers/journals exploring and incorporating outside perspectives on data-sharing, such as those from Indigenous peoples, researchers and participants, into their data-sharing policies.

Tensions around data-sharing within the global community of researchers

The geographical location of researchers can create tension when data-sharing is required or requested. Country-specific legislation may conflict, and researchers from countries with higher income generally have better access to funding, resources, and infrastructure support, while researchers from lower-income countries may benefit significantly from wider data-sharing but may lack the necessary resources to do so (Wallis et al. 2013; Levin et al. 2016; Kurata et al. 2017; Ho et al. 2018). The current literature would be enriched by cross-cultural research that examines perspectives on data-sharing within a social justice framework that considers who benefits from data-sharing.

Data-sharing can be contentious among researchers as there is often a reluctance to share data with peers. The competitiveness within academia, perceived notions of ownership, and a lack of resources have all been cited as reasons why researchers do not want to share their data with others (Wallis et al. 2013; Cheah et al. 2015; Levin et al. 2016; Cheah et al. 2017; Martone et al. 2018). These concerns are indicative of a wider cultural system in academia, and it is likely that systemic changes to academic publishing would have to occur to assuage concerns. Indigenous researchers are more likely to be opposed to data-sharing without first gaining consent from Indigenous collectives (Hudson, Southey, et al. 2016) and being assured of the benefits to Indigenous communities. Important first steps might include: introduction of guidelines around data ownership which are more closely aligned with Indigenous data sovereignty guidelines (Research Data Alliance International Indigenous Data Sovereignty Interest Group 2019); increased resources made available to researchers by those actioning data-sharing policies; and an increased focus on publishing non-significant or negative findings.

Tensions around data-sharing due to lack of consultation of participants or Indigenous collectives

The evidence presented in this review suggests a tension exists around the degree of consultation between researchers and participants or Indigenous collectives when it comes to issues around data-sharing. Researchers must simultaneously interpret top-down data-sharing policies from funders or publishers/journals and act as the sole point of contact with people or communities who participate in their research projects. Moreover, there exists a power imbalance within discussions around data-sharing in that researchers' power compared to participants has been ignored. Most of the extant research about researchers' perspectives on data-sharing suggests that participants are not considered beyond common practices such as de-identification, and there is a concurrent lack of research investigating the views of Indigenous researchers who work with Indigenous communities. Hudson (2016) postulated most participants would be open to data-sharing, but emphasised it to be a supposition due to the lack of research focusing on participants' views. Research about data-sharing that amplifies the perspectives of participants and Indigenous collectives is an important next step in minimising the tension between these stakeholders in the debate on data-sharing.

Tensions around data-sharing due to current consent processes

The evidence to-date suggests that participants do not have a clear understanding of the consent options that are commonly used in research pertaining to data-sharing (Giesbertz et al. 2012; Garrison et al. 2016; Hudson, Southey, et al. 2016; Beaton et al. 2017; Goodman et al. 2017; Goodman et al. 2018). There is a need for research that explores perceptions or experiences of consent options outside of blanket consent or categorical consent to build on promising pilot research exploring dynamic consent methods (Spencer et al. 2016). This gap in the literature would be addressed by further research that qualitatively explores perceptions of dynamic consent to explore whether it is an optimal way to gain consent from participants.

When considering data-sharing with Māori and/or other Indigenous collectives and participants, extant research has clearly identified that hegemonic consent processes are inappropriate for specific types of Indigenous data (e.g. genetic material) especially with regards to the unspecified future use of data, and that consent should be sought at the collective level. This tension could be mitigated by strengthening protections around collective privacy and collective data-sharing.

Tensions around data-sharing between Māori and non-Māori knowledge/value paradigms

As has been found for Indigenous participants around the globe, the literature relating to data-sharing and data sovereignty in Aotearoa/New Zealand has highlighted that Māori participants are reticent about data-sharing and its implementation more widely (Smith 2012), in part due to previous negative experiences when engaging with non-Māori researchers and compounded by biological data samples being tapu (of a sacred and special nature; contains notions of restriction) (Hudson, Russell, et al. 2016; Beaton et al. 2017). Engaging with appropriate guidelines such as the CARE principles (Research Data Alliance International Indigenous Data Sovereignty Interest Group 2019) is an important ethical step in research with people in Aotearoa/New Zealand. These guidelines also provide useful insights when considering the appropriateness of data-sharing, and these insights should be supplemented by consultation with participants and other stakeholders in the data-sharing debate.

The Māori and Indigenous data sovereignty movement challenges the hegemonic assumption that all data should be readily available to the Academy and other higher research agencies. These sites have perpetuated colonisation processes on Indigenous peoples by controlling what constitutes ‘legitimate’ knowledge, under the guise that Western knowledge and value paradigms are ‘universal’ (Smith and Smith 2019). There exists the potential for these sites to create meaningful change by meeting the aspirations of Indigenous peoples; such changes may be brought about by active, sustained decolonisation of higher research agencies, and will require systemic change at multiple places within these agencies (Smith and Smith 2019).

Tensions around data-sharing in longitudinal research

There is a gap in the evidence-base about data-sharing because participants in different types of research have been treated as a homogenous group. In particular, participants in longitudinal research have been neglected as stakeholders in the data-sharing literature. Research exploring how participants in longitudinal research feel about their data being shared with established stakeholders would help to broaden the currently limited evidence-base on participants' perspectives on data-sharing when they have had a short-term involvement in research or a life-long commitment to a project such as the Dunedin Study (Poulton et al. 2015).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Office of the Privacy Commissioner Te Mana Mātāpono Matatapu. Reremoana Theodore was supported by a Health Research Council of New Zealand (HRC) Fellowship [grant number 18/644].

References

Aitken M, de St. Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. 2016. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics. [accessed 2018 Jun 29]. 17(1). http://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-016-0153-x.

Allison JR, Cooper WW. 1992. Data disclosure and data sharing in scientific research. Account Res. 2(2):93–132.

American Psychological Association. 2010. Publication manual of the American Psychological Association. 6th ed. 2nd printing [corr. rev. ed.]. Washington, DC: American Psychological Association.

Beaton A, Hudson M, Milne M, Port RV, Russell K, Smith B, Toki V, Uerata L, Wilcox P, Bartholomew K, Wihongi H. 2017. Engaging Māori in biobanking and genomic research: a model for biobanks to guide culturally informed governance, operational and community engagement activities. Genet Med. 19(3):345–351.

Bem DJ. 2011. Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. J Pers Soc Psychol. 100(3):407–425.

Carr D, Littler K. 2015. Sharing research data to improve public health: a funder perspective. J Empir Res Hum Res Ethics. 10(3):314–316.

Cheah PY, Day NPJ, Parker M, Bull S. 2017. Sharing individual-level health research data: experiences, challenges and a research agenda. Asian Bioeth Rev. 9:393–400.

Cheah PY, Tangseefa D, Somsaman A, Chunsuttiwat T, Nosten F, Day NPJ, Bull S, Parker M. 2015. Perceived benefits, harms, and views about how to share data responsibly: a qualitative study of experiences with and attitudes toward data sharing among research staff and community representatives in Thailand. J Empir Res Hum Res Ethics. 10(3):278–289.

Craig JR, Reese SC. 1973. Retention of raw data: a problem revisited. Am Psychol. 28(8):723–723.

Crosas M, Gautier J, Karcher S, Kirilova D, Otalara G, Schwartz A. 2018. Data policies of highly-ranked social science journals. 1–28.

Denny SG, Silaigwana B, Wassenaar D, Bull S, Parker M. 2015. Developing ethical practices for public health research data sharing in South Africa: the views and experiences from a diverse sample of research stakeholders. J Empir Res Hum Res Ethics. 10(3):290–301.

El Emam K, Buckeridge D, Tamblyn R, Neisa A, Jonker E, Verma A. 2011. The re-identification risk of Canadians from longitudinal demographics. BMC Med Inform Decis Mak. [accessed 2018 Oct 8]. 11(1). http://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-11-46.

Fecher B, Friesike S, Hebing M. 2015. What drives academic data sharing? PLoS One. 10(2):e0118053.

Federer LM, Belter CW, Joubert DJ, Livinski A, Lu Y-L, Snyders LN, Thompson H. 2018. Data sharing in PLOS ONE: an analysis of data availability statements. PLoS One. 13(5):e0194768.

Garrison NA, Sathe NA, Antommaria AHM, Holm IA, Sanderson SC, Smith ME, McPheeters ML, Clayton EW. 2016. A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. Genet Med. 18(7):663–671.

George L, Macdonald LTAT, Tauri J. 2020. An introduction to Indigenous research ethics. In: George L, Tauri J, MacDonald LTAT, editors. Indigenous research ethics: claiming research sovereignty beyond deficit and the colonial legacy. Vol. 6. [place unknown]: Emerald Publishing Limited; p. 1–15. doi:10.1108/S2398-601820200000006001.

Gibson L, Mercier OR, Kiddle R. 2020. Engaging with 'that treaty question' on a university ethics committee in Aotearoa New Zealand. In: George L, Tauri J, MacDonald LTAT, editors. Indigenous research ethics: claiming research sovereignty beyond deficit and the colonial legacy. Vol. 6. [place unknown]: Emerald Publishing Limited; p. 95–108. doi:10.1108/S2398-601820200000006007.

Giesbertz NAA, Bredenoord AL, van Delden JJM. 2012. Inclusion of residual tissue in biobanks: opt-in or opt-out? PLoS Biol. 10(8):e1001373.

Giner-Sorolla R. 2012. Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. Perspect Psychol Sci. 7(6):562–571.

Giofrè D, Cumming G, Fresc L, Boedker I, Tressoldi P. 2017. The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices. PLoS One. 12 (4):e0175583.

Global Indigenous Data Alliance. 2019. CARE principles of Indigenous data governance. Glob Indig Data Alliance [Internet]. [accessed 2021 Jan 12]. https://www.gida-global.org/care.

Goodman D, Johnson CO, Bowen D, Smith M, Wenzel L, Edwards K. 2017. De-identified genomic data sharing: the research participant perspective. J Community Genet. 8(3):173–181.

Goodman D, Johnson CO, Bowen D, Smith M, Wenzel L, Edwards KL. 2018. A comparison of views regarding the use of de-identified data. Transl Behav Med. 8(1):113–118.

Harper LM, Kim Y. 2018. Attitudinal, normative, and resource factors affecting psychologists' intentions to adopt an open data badge: an empirical analysis. Int J Inf Manag. 41:23–32.

Harron K, Dibben C, Boyd J, Hjern A, Azimaee M, Barreto ML, Goldstein H. 2017. Challenges in administrative data linkage for research. Big Data Soc. 4(2). doi:10.1177/2053951717745678.

Hate K, Meherally S, Shah More N, Jayaraman A, Bull S, Parker M, Osrin D. 2015. Sweat, skepticism, and uncharted territory: a qualitative study of opinions on data sharing among public health researchers and research participants in Mumbai, India. J Empir Res Hum Res Ethics. 10(3):239–250.

Ho HKK, Görges M, Portales-Casamar E. 2018. Data access and usage practices across a cohort of researchers at a large tertiary pediatric hospital: qualitative survey study. JMIR Med Inform. 6 (2):e32.

Hudson M. 2016. Data sharing – peeping inside Pandora's box. Eur Sci Ed. [accessed 2018 Jun 29]. http://europeanscienceediting.eu/articles/data-sharing-peeping-inside-pandoras-box/.

Hudson M, Beaton A, Milne M, Port W, Russell KJ, Smith B, Toki V, Wilcox P, Uerata L, Te Mata Hautū Taketake – The Māori & Indigenous Governance Centre. 2016. Te Mata Ira: guidelines for genomic research with Māori. [place unknown]; [accessed 2019 Jul 19]. http://natlib-primo.hosted.exlibrisgroup.com/NLNZ:NLNZ:NLNZ_ALMA11274890460002836.

Hudson M, Garrison NA, Sterling R, Caron NR, Fox K, Yracheta J, Anderson J, Wilcox P, Arbour L, Brown A, et al. 2020. Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. Nat Rev Genet. 21(6):377–384.

Hudson M, Russell K. 2009. The treaty of Waitangi and research ethics in Aotearoa. J Bioethical Inq. 6:61–68.

Hudson M, Russell K, Uerata L, Milne M, Wilcox P, Port RV, Smith B, Toki V, Beaton A. 2016. Te Mata ira—faces of the gene: Developing a cultural foundation for biobanking and genomic research involving Māori. Altern Int J Indig Peoples. 12(4):341–355.

Hudson M, Southey K, Uerata L, Beaton A, Milne M, Russell K, Smith B, Wilcox P, Toki V, Port W. 2016. Key informant views on biobanking and genomic research with Māori. N Z Med J. 129 (1447):14.

Ioannidis JPA. 2005. Why most published research findings are false. PLoS Med. 2(8):e124.

IWGIA (International Work Group for Indigenous Affairs). 2020. Indigenous world 2020: indigenous data sovereignty. [accessed 2021 Jan 12]. https://www.iwgia.org/en/ip-i-iw/3652-iw-2020-indigenous-data-sovereignty.html.

Jamal L, Sapp JC, Lewis K, Yanes T, Facio FM, Biesecker LG, Biesecker BB. 2014. Research participants' attitudes towards the confidentiality of genomic sequence information. Eur J Hum Genet. 22(8):964–968.

Kaye J, Hawkins N. 2014. Data sharing policy design for consortia: challenges for sustainability. Genome Med. 6(1):4.

Keerie C, Tuck C, Milne G, Eldridge S, Wright N, Lewis SC. 2018. Data sharing in clinical trials – practical guidance on anonymising trial datasets. Trials. [accessed 2018 Jun 29]. 19(1). https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-017-2382-9.

Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, Falkenberg L-S, Kennett C, Slowik A, Sonnleitner C, Hess-Holden C, *et al.* 2016. Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. PLOS Biol. 14(5):e1002456.

Kiley R, Peatfield T, Hansen J, Reddington F. 2017. Data sharing from clinical trials – a research funder's perspective. N Engl J Med. 377(20):1990–1992.

Kukutai T, Carroll SR, Walter M. 2020. Indigenous data sovereignty. In: Indigenous World 2020. 34th ed.: IWGIA; p. 654–662. http://iwgia.org/images/yearbook/2020/IWGIA_The_Indigenous_World_2020.pdf.

Kurata K, Matsubayashi M, Mine S. 2017. Identifying the complex position of research data and data sharing among researchers in natural science. SAGE Open. 7(3). doi:10.1177/2158244017717301.

Lemke AA, Wolf WA, Hebert-Beirne J, Smith ME. 2010. Public and biobank participant attitudes toward genetic research participation and data sharing. Public Health Genomics. 13(6):368–377.

Levin N, Leonelli S, Weckowska D, Castle D, Dupré J. 2016. How do scientists define openness? Exploring the relationship between open science policies and research practice. Bull Sci Technol Soc. 36(2):128–141.

Ludman EJ, Fullerton SM, Spangler L, Trinidad SB, Fujii MM, Jarvik GP, Larson EB, Burke W. 2010. Glad you asked: participants' opinions of re-consent for DbGap data submission. J Empir Res Hum Res Ethics. 5(3):9–16.

Mählmann L, Schee gen. Halfmann S, von Wyl A, Brand A. 2017. Attitudes towards personal genomics and sharing of genetic data among older Swiss adults: a qualitative study. Public Health Genomics. 20(5):293–306.

Martone ME, Garcia-Castro A, VandenBos GR. 2018. Data sharing in psychology. Am Psychol. 73 (2):111–125.

McCormack P, Kole A, Gainotti S, Mascalzoni D, Molster C, Lochmüller H, Woods S. 2016. You should at least ask'. The expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research. Eur J Hum Genet. 24(10):1403–1408.

McGuire AL, Hamilton JA, Lunstroth R, McCullough LB, Goldman A. 2008. DNA data sharing: research participants' perspectives. Genet Med. 10(1):46–53.

Mello MM, Lieou V, Goodman SN. 2018. Clinical trial participants' views of the risks and benefits of data sharing. N Engl J Med. 378(23):2202–2211.

Molloy JC. 2011. The open knowledge foundation: open data means better science. PLoS Biol. 9 (12):e1001195.

Moore D, Niemi M. 2016. The sharing of personal health data – a review of the literature. [place unknown]. http://datafutures.co.nz/assets/Uploads/The-Sharing-of-Personal-Health-Data-Sapere-FINAL.pdf.

Murad AM, Myers MF, Thompson SD, Fisher R, Antommaria AHM. 2017. A qualitative study of adolescents' understanding of biobanks and their attitudes toward participation, re-contact, and data sharing. Am J Med Genet A. 173(4):930–937.

Nairn K, Showden CR, Sligo J, Matthews KR, Kidman J. 2020. Consent requires a relationship: rethinking group consent and its timing in ethnographic research. Int J Soc Res Methodol. 23(6):1–13.

The National Commission for the Protection of Human Subjects of, Biomedical and Behavioral Research. 1978. The Belmont Report. [place unknown]. https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf.

New Zealand Psychological Society. 2012. Code of ethics for psychologists working in Aotearoa/New Zealand [Internet]. Wellington; [accessed 2021 Jan 5]. http://www.psychologistsboard.org.nz/cms_show_download.php?id=235.

Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, Buck S, Chambers CD, Chin G, Christensen G, et al. 2015. Promoting an open research culture. Science. 348 (6242):1422–1425.

Oliver JM, Slashinski MJ, Wang T, Kelly PA, Hilsenbeck SG, McGuire AL. 2012. Balancing the risks and benefits of genomic data sharing: genome research participants' perspectives. Public Health Genomics. 15(2):106–114.

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. Science. 349(6251):aac4716.

Opus International Consultants. 2015. Public attitudes to data integration: report prepared for statistics New Zealand. [place unknown].

Pashler H, Wagenmakers E. 2012. Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? Perspect Psychol Sci. 7(6):528–530.

Pearce N, Smith AH. 2011. Data sharing: not as simple as it seems. Environ Health. [accessed 2018 Jun 29]. 10(1). http://ehjournal.biomedcentral.com/articles/10.1186/1476-069X-10-107.

Piwowar HA. 2011. Who shares? Who doesn't? Factors associated with openly archiving raw research data. PLoS One. 6(7):e18657.

Piwowar HA, Chapman WW. 2008. A review of journal policies for sharing research data. 14.

Piwowar HA, Chapman WW. 2010. Public sharing of research datasets: a pilot study of associations. J Informetr. 4(2):148–156.

Poulton R, Moffitt TE, Silva PA. 2015. The Dunedin multidisciplinary health and development study: overview of the first 40 years, with an eye to the future. Soc Psychiatry Psychiatr Epidemiol. 50(5):679–693.

Rainie SC, Kukutai T, Walter M, Figueroa-RodrÃGuez OL, Walker J, Axelsson P. 2019. Indigenous data sovereignty. In: Davies T, Walker SB, Rubinstein M, Perini F, editors. The state of open data: histories and horizons. Cape Town: African Minds and the International Development Research Centre (IDRC); p. 300–319. https://hdl.handle.net/10289/12918.

Research Data Alliance International Indigenous Data Sovereignty Interest Group. 2019. CARE principles of Indigenous data governance. Glob Indig Data Alliance. [accessed 2021 Jan 12]. https://www.gida-global.org/care.

Roediger HL. 2012. Psychology's woes and a partial cure: the value of replication. APS Obs. 25(2).

Rouder JN. 2016. The what, why, and how of born-open data. Behav Res Methods. 48(3):1062–1069.

Shah N, Coathup V, Teare H, Forgie I, Giordano GN, Hansen TH, Groeneveld L, Hudson M, Pearson E, Ruetten H, Kaye J. 2019. Sharing data for future research—engaging participants' views about data governance beyond the original project: a DIRECT study. Genet Med. 21 (5):1131–1138.

Shaw RL, Senior C, Peel E, Cooke R, Donnelly LS. 2008. Ethical issues in neuroimaging health research: an IPA study with research participants. J Health Psychol. 13(8):1051–1059.

Smith LT. 2012. Decolonizing methodologies: research and Indigenous peoples. London: Zed Books. [accessed 2019 Jul 18]. http://ebookcentral.proquest.com/lib/otago/detail.action?docID=1426837.

Smith GH, Smith LT. 2019. Doing Indigenous work: decolonizing and transforming the academy. In: McKinley EA, Smith LT, editors. Handbook of Indigenous education. Singapore: Springer; p. 1075–1101. doi:10.1007/978-981-10-3899-0_69.

Spencer K, Sanders C, Whitley EA, Lund D, Kaye J, Dixon WG. 2016. Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: a qualitative study. J Med Internet Res. 18(4):e66.

Statistics New Zealand Tatauranga Aotearoa. 2018. Integrated data infrastructure | Stats NZ. [accessed 2020 Oct 15]. https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure.

Sturges P, Bamkin M, Anders JHS, Hubbard B, Hussain A, Heeley M. 2015. Research data sharing: developing a stakeholder-driven model for journal policies: research data sharing. J Assoc Inf Sci Technol. 66(12):2445–2455.

Taniguchi NK, Taualii M, Maddock J. 2012. A comparative analysis of Indigenous research guidelines to inform genomic research in Indigenous communities. Int Indig Policy J. [accessed 2018 Jun 29]. 3(1). http://ir.lib.uwo.ca/iipj/vol3/iss1/6/.

Taylor J, Kukutai T. 2017. Indigenous data sovereignty: toward an agenda. Canberra: ANU Press.

Tellam RL, Rushton P, Schuerman P, Pala I, Anane D. 2015. The primary reasons behind data sharing, its wider benefits and how to cope with the realities of commercial data. BMC Genomics. [accessed 2018 Jun 29]. 16(1). http://www.biomedcentral.com/1471-2164/16/626.

Te Mana Raraunga. n.d. Te Mana Raraunga. [accessed 2020 Jun 25]. https://www.temanararaunga.maori.nz.

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M. 2011. Data sharing by scientists: practices and perceptions. PLoS One. 6(6):e21101.

Theodore R, Ratima M, Edwards W, Sporle A, Te Morenga L, Kiro C, Hond R. 2019. How a life-course approach can promote long-term health and wellbeing outcomes for Māori. J Indig Wellbeing. 4(1):15–25.

Treharne GJ, Marx J. 2018. Introduction: encounters with systems within which critical research is conducted. In: Macleod CI, Marx J, Mnyaka P, Treharne GJ, editors. The Palgrave handbook of ethics in critical research. Cham: Springer International Publishing; p. 17–28. doi:10.1007/978-3-319-74721-7_2.

Tsai AC, Kohrt BA, Matthews LT, Betancourt TS, Lee JK, Papachristos AV, Weiser SD, Dworkin SL. 2016. Promises and pitfalls of data sharing in qualitative research. Soc Sci Med. 169:191–198.

Vanpaemel W, Vermorgen M, Deriemaecker L, Storms G. 2015. Are we wasting a Good crisis? The availability of psychological research data after the storm. Collabra. [accessed 2018 Jun 29]. 1(1). http://collabra.org/article/view/collabra.13/.

Wallis JC, Rolando E, Borgman CL. 2013. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLoS ONE. 8(7):e67332.

Walter M, Suina M. 2019. Indigenous data, indigenous methodologies and indigenous data sovereignty. Int J Soc Res Methodol. 22(3):233–243.

West K, Hudson M, Kukutai T. 2020. Data ethics and data governance from a Māori world view. In: George L, Tauri J, MacDonald LTAT, editors. Indigenous research ethics: claiming research sovereignty beyond deficit and the colonial legacy. Vol. 6. [place unknown]: Emerald Publishing Limited; p. 67–81. doi:10.1108/S2398-601820200000006005.

Wicherts JM, Bakker M, Molenaar D. 2011. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. PLoS One. 6(11): e26828.

Wicherts JM, Borsboom D, Kats J, Molenaar D. 2006. The poor availability of psychological research data for reanalysis. Am Psychol. 61(7):726–728.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 3(1):160018.

Williams M, Bagwell J, Nahm Zozus M. 2017. Data management plans: the missing perspective. J Biomed Inform. 71:130–142.

Wolins L. 1962. Responsibility for raw data. Am Psychol. 17(9):657–658.

Yong E. 2012. Replication studies: bad copy. Nature. 485:298–300.

Zarate OA, Brody JG, Brown P, Ramirez-Andreotta MD, Perovich L, Matz J. 2016. Balancing benefits and risks of immortal data: participants' views of open consent in the personal genome project. Hastings Cent Rep. 46(1):36–45.