Developing a strategy to improve data sharing in health research: A mixed-methods study to identify barriers and facilitators

Michelle A Krahe¹, BA(BiomedSc), PhD
Malcolm Wolski², BBus, MEnvMgt, MInfTech
Sharon Mickan¹, BOccThy, MA(Education), PhD
Julie Toohey³, BA(Info & Lib Studies)
Paul Scuffham⁴,⁵, BA, PhD, GAICD, FAHMS
Sheena Reilly¹, BA(SpPath), PhD, FAHMS, FRCSLT, FSPA, FASSA

¹Health Group, Griffith University, Southport, QLD, Australia
²eResearch Services, Griffith University, Nathan, QLD, Australia
³Library, Griffith University, Logan, QLD, Australia
⁴Centre for Applied Health Economics, Griffith University, Nathan, QLD, Australia
⁵Menzies Health Institute Queensland, Griffith University, Southport, QLD, Australia

Corresponding author: Dr Michelle A. Krahe, Health Executive, Griffith University, Parklands Drive, Southport, 4222, QLD Australia. Email: m.krahe@griffith.edu.au

Short title: Data sharing behaviour in health research
21 Abstract

22 Background: It is well accepted that data sharing presents new opportunities across
the spectrum of research and is vital towards science that is open; where data are
easily discoverable, accessible, intelligible, reproducible, replicable, and verifiable.
25 Despite this, it is yet to become common practice. Global efforts to develop practical
guidance for data sharing and open access initiatives are underway, however
evidence-based studies to inform the development and implementation of effective
strategies are lacking.

29 Objective: This study sought to determine the barriers and facilitators to data sharing
among health researchers and to identify the target behaviours for designing a
behaviour change intervention strategy.

32 Method: Data are drawn from a cross-sectional survey of data management
practices among health researchers from one Australian research institute.

34 Determinants of behaviour are theoretically derived using well established
behavioural models.

36 Results: The data sharing practices are described for 77 researchers. Six barriers
and four facilitators are identified. The primary barriers to data sharing include:
perceived negative consequences, and lack of competency to share data. The
primary facilitators to data sharing include: trust in others using the data, and social
influence related to public benefit. Intervention functions likely to be most effective at
changing target behaviours are also identified.

42 Conclusion: Results of this study provide a theoretical and evidence-based process
to understand the behavioural barriers and facilitators of data sharing among health
researchers.
Implications: Designing interventions that specifically address target behaviours to promote data sharing is important for open researcher practices.

Keywords (Mesh): data sharing, health information management, research

Supplementary keywords: behaviour change, COM-B model, intervention design

Article type: Research
Introduction

While data sharing and reuse is more common in some academic disciplines than others, the argument for and against is remarkably similar. Researchers may be motivated to share data if it results in greater visibility or increased citation of their work, if a supportive data sharing culture exists in their workplace, or if they have had positive experiences with open data reuse (Patel, 2016; Piwowar, 2011; Piwowar et al., 2007; Zuiderwijk and Spiers, 2019). Conversely, researchers are demotivated to share data if they are concerned about privacy, fear of reputation or “scooping” of research, have a desire to retain exclusive rights to data, or are concerned about the inflation, misinterpretation or the usability of data (Alter and Vardigan, 2015; Arzberger et al., 2004; Campbell and Bendavid, 2003; Cheah et al., 2015; Kuipers and van der Hoeven J., 2009; Savage and Vickers, 2009; Tenopir et al., 2011; Zuiderwijk and Spiers, 2019). The perceptions and practices of data sharing and data reuse among researchers was examined by Tenopir et al. (2015), who observed significant increases in the willingness and acceptance of sharing, and “actual” data sharing over a 4-year period. This was also met with an increase in the perception of risk associated with data sharing, including the potential for misuse and misinterpretation and concern about the need to publish data before making it became available. The study also describes that receiving acknowledgement, or a formal citation was important when deciding whether to share data. The Wiley Data Insights survey of more than 4500 researchers from 112 countries indicates a range of motivations for data sharing, including: to increase impact and visibility of research, public benefit, transparency and reuse, or journal requirement (Ferguson, 2014). Australian researchers were most incentivised to share data to ensure preservation, transparency and reuse, or to satisfy a funding requirement. This
survey demonstrates that support for open researcher practices may not directly
translate into data sharing practices; rather influenced by the increased scientific
competitiveness of research, the research environment specific to the country, or the
appeal for commercial application.

In Australia, around 75% of universities have an open access mandate, policy or
statement, including the two major funding agencies: the National Health and
Medical Council and the Australian Research Council (Australian Research Council,
2017; National Health and Medical Research Council, 2018). Despite this support,
the one shortcoming and arguably the most important feature of open practices, is
the lack of participation in data sharing from the research community (Ali-Khan et al.,
2017; Hauessler, 2011). Tenopir and colleagues (2011) report that < 6% of
international researchers make “all” their data available, and while only 36% indicate
that others could easily access their data, approximately 75% agreed that they would
use other researcher’s datasets if easily accessible – these findings are also
supported by others (Federer et al., 2015; Hickson et al., 2016).

**How can researchers be encouraged to share data?**

Encouraging data sharing amongst the scientific community and perhaps more
importantly, understanding what is “ideal” and what is the “reality” of data sharing is a
highly complex and contextual challenge (Bezuidenhout, 2019). In an extensive
review of the literature only one evidence-based incentive for data sharing was
identified (Rowhani-Farid et al., 2017). This was a digital badge created by the
Center for Open Science and tested in the Journal of Psychological Science. The
journal editorial team awarded badges to those authors who voluntarily applied and
proved they met the criteria for open data and open materials upon article
acceptance. Sharing rates greatly improved over three years with the introduction of
For Peer Review

103 the digital badge (1.5% to 39.4%), but upon review, data and materials had
104 somehow disappeared from public view (Kidwell et al., 2016). Several other
105 evidence-based strategies were identified, most were technological strategies
106 (~75%) such as the introduction of data systems to manage and store data, and
107 others included collaborative data environments and workshops, implementation of
108 data sharing policies, and an open data campaign. However, many strategies lacked
109 good quality empirical data of their effectiveness.

110 Conceptual models of how both institutional and individual drivers influence
111 researcher’s data sharing behaviours, indicate the pivotal role that institutions play in
112 influencing open researcher practices (Ali-Khan et al., 2017; Fecher et al., 2015;
113 Kim, 2012). The dynamic relationship between the willingness to share data and the
114 actual sharing of data is largely motivated by rational self-interest rather than pure
115 altruism (Ali-Khan et al., 2017; Fecher et al., 2015; Levin and Leonelli, 2017; Levin et
116 al., 2016). The highly competitive nature of research and the importance of
117 maintaining strong academic metrics, mean that many fear they will miss out on
118 receiving attribution or credit (Anderson and Schonfeld, 2009; Antman, 2014;
119 Federer et al., 2015; Spallek et al., 2019). A review of the principles and norms of
120 data sharing in international health research identified four overarching themes: (1)
121 societal benefits and value; (2) distribution of risk, benefit and burden; (3) respect for
122 individuals and groups; and (4) public trust and engagement (Kalkman et al., 2019).
123 These themes are important to consider for data governance frameworks, data
124 management and sharing policies, and data processes and tools specific to each
125 research context. Observational studies that have identified patterns and barriers to
126 data sharing are plentiful, but whilst these provide useful background knowledge,
127 they do not provide evidence of what can be done to increase data sharing. The
development of successful interventions must be underpinned by local evidence and rigorous theoretical constructs, to facilitate behaviour change and provide explanation for the mechanism of change.

Understanding behaviours that influence data sharing

The use of theoretical frameworks in the design and evaluation of interventions has been increasingly highlighted by implementation researchers (Eccles et al., 2005; Kok et al., 2016; Michie and Johnston, 2012; Michie et al., 2005). For the first time, this study will apply established theories to understand behaviours that influence data sharing of health researchers. The two models to be employed are the Behaviour Change Wheel (BCW) and the Theoretical Domains Framework (TDF).

The BCW is a validated model produced through the synthesis of many behaviour change theories. It is commonly used to analyse, and design behaviour change interventions (Michie et al., 2014). At the centre of the wheel are three core components: capability, opportunity and motivation (collectively known as COM-B). These components interact to produce “behaviour” and may explain the data sharing behaviours of researchers who have different capabilities and opportunities to change behaviour, regardless of their motivation (Michie et al., 2011). The middle ring comprises of nine intervention functions and the outer ring seven policy-driven strategies (Figure 1). The three key steps to using the BCW to design behaviour change interventions includes: (1) understanding the behaviour (COM-B); (2) identifying intervention options (middle ring); and (3) identifying content and implementation options (outer ring).
In addition, the TDF is based on psychological and organisational theories to guide analysis and implementation of evidence-based practice (Michie et al., 2005). The TDF system consists of 14 domains that categorise influences on behaviour and are mapped to the COM-B components of the BCW in the analysis and design of targeted interventions (Box 1) (Atkins et al., 2017; Cane et al., 2012). The BCW and TDF have both been used to analyse and design behavioural change interventions in a variety of health settings, such as: reducing sugar free intake in adults (Rawahi et al., 2018), implementing telemedicine consultations for paediatric care (Bele et al., 2019), refining injury management in hospitals (Curtis et al., 2017), and understanding physical activity in overweight and obese pregnant women (Flannery et al., 2018). Despite the current shift towards more open data in research, there are seemingly no theoretically informed explorations; our understanding is limited to pragmatic reports.

Box 1. Mapping of the TDF domains to the COM-B model (Atkins et al., 2017).

<table>
<thead>
<tr>
<th>TDF domain</th>
<th>COM-B component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioural Regulation; Cognitive and</td>
<td>Psychological Capability</td>
</tr>
<tr>
<td>Interpersonal Skills; Knowledge; Memory,</td>
<td></td>
</tr>
<tr>
<td>Attention and Decision Processes</td>
<td></td>
</tr>
<tr>
<td>Physical Ability (Skills)</td>
<td>Physical Capability</td>
</tr>
<tr>
<td>Environmental Context and Resources</td>
<td>Physical Opportunity</td>
</tr>
<tr>
<td>Social Influences</td>
<td>Social Opportunity</td>
</tr>
</tbody>
</table>
Beliefs About Capabilities; Beliefs About
Consequences; Emotion; Goals; Intentions;
Optimism; Social/Professional Role and Identity

Reflective Motivation

Reinforcement; Emotion

Automatic Motivation

Study objectives
The objective of this study was to determine the key behaviours that influence data
sharing practices among researchers, to inform the design of targeted behaviour
change interventions. The specific aims include:

a) Describe the current data sharing behaviours of health researchers using a
qualitative descriptive design.

b) Conduct a behavioural diagnosis of researcher data sharing behaviours using
the TDF and COM-B model.

c) Identify target behaviours and characterise intervention functions likely to be
most effective at changing target behaviours.
Method

Study design and recruitment

A cross-sectional survey study was conducted to evaluate the data management practices of health researchers from one research institute at Griffith University, Australia. Full study methods and results have been previously published (Krahe et al., 2019a). The present study employed mixed-methods analysis to a selection of the survey data (not previously reported) (Sandelowski, 2000). A copy of the survey questions used in the analysis are available as Online Supplemental Material. This study and all documentation received ethical approval from the Griffith University Human Research Ethics Committee on 3 July 2017 (Reference number 2017/457).

Participants were eligible if they were employed at the university in a permanent, temporary or casual appointment and were actively involved in research. Potential participants (n = 401) were identified via the membership database and approached by means of email with a request to participate in our online survey. The initial email was sent with a link to the survey. If no response was received within two weeks, one reminder email was sent. All participants were recruited between July and August 2017. LimeSurvey (GmbH, v1.9X, Hamburg Germany) was used to administer the survey.

Data sharing behaviours of researchers

The main outcome measures were derived from survey questions about researcher’s current data sharing practice, their willingness to share data, and responses to statements designed to explore components of data sharing behaviours. Descriptive statistics were used to express categorical variables as counts and percentages. Response options for questions about willingness to share data are dichotomised into two categories: willing (and somewhat willing); and unwilling (somewhat
unwilling or not at all willing). These were coded as: willing = 1, and unwilling = 2.

While the original survey provided a neutral option (neither willing nor unwilling) it was treated as “missing” in this analysis, as it does not provide any information about the level of agreement to the statement. A paired-samples t-test was conducted to compare willingness to share data by researchers before and after publishing. The mean and 95% CI for the change in willingness is also given. Quantitative survey data are imported into SPSS (IBM version 25.0) and analysed using descriptive statistics.

**Behavioural analysis and diagnosis**

Survey data are coded based on the definitions of the TDF domains (Cane et al., 2012) and exemplar quotes illustrating these were identified in this process. Quantitative survey data are deductively mapped to the TDF domains and then to behavioural components of the COM-B model. Discussion occurred with members of the research team to ensure appropriate coding. Behavioural themes were identified and categorised as barriers (requiring modification) or facilitators (needing enhancing).

**Identify intervention functions to change target behaviours**

Applying the BCW, a well validated and widely used framework of behaviour change (Michie et al., 2014), intervention functions that correlate with the TDF and COM-B component for each behaviour were identified. These intervention functions are the “active ingredients” of intervention strategies.


Results

Participant characteristics

A total of 81 researchers participated in the study. Only 77 (19.2%) sufficiently completed section four of the survey (i.e. related to data sharing) and were included in the analysis. The majority were members of academic staff (58.4%), working in the discipline of medical sciences (32.5%), and had worked in research for more than 10 years (57.0%) (Table 1). Almost one-quarter of respondents self-identified as early career researchers (n = 16, 24.6%), with a median of 2.5-years since completing their PhD (range = 0 to 12 years).

To understand the data characteristics that researchers were working with, respondents were asked to indicate where they typically sourced their research data (Krahe et al., 2019b). A list of primary (information collected and processed by the researcher) and secondary (information retrieved through pre-existing sources) data sources were provided, with the most common being surveys (66.2%), interviews (54.5%), and experiments (46.8%) (Figure 2).

Data sharing behaviours of researchers

Almost two-fifths of respondents (37.7%, n = 29) had ever shared data outside their immediate research/project team. Of these, 65.5% (n = 19/29) shared data only after
the research data had been published (Figure 3). Only 15.6% (n = 12/77) of
researchers had ever shared their data publicly as: supplementary material in a
journal (58.3%), through an institutional (41.7%) or discipline-specific (25%)
repository, at a conference (33.3%), on a personal or project website (16.7%), or
through direct data requests (8.3%) (results not tabulated). The motivations for
researchers that shared their data with others outside of the research project and
more publicly (i.e. open access) are presented in Figure 4(a) and (b).

The attitudes of researchers towards data sharing was further explored through their
willingness to share data at two time points: (1) before data had been published; and
(2) after data had been published. Figure 5 illustrates that overall researchers were
most willing (97.2% ± 1.97, n = 69/71) to share data before it was published with
members of their research group, and least willing (24.5% ± 6.2, n = 12/49) to share
data with the wider public. After data had been published, all respondents (100%, n =
71/71) were willing to share data within their research group, 91% ± 3.8 (n = 51/56)
were willing to share with colleagues in their organisations, 89% ± 4.2 (n = 49/55)
with colleagues in their field, and just over one-half (55.1% ± 7.1, n = 27/49) with the
wider public.
A significant increase in the willingness to share data with colleagues in the organisation (14.2%; 95% CI 4.8 to 23.7; \( P < 0.01 \)), colleagues in the field (29%; 95% CI 16.6 to 41.4; \( P < 0.0001 \)), and with the wider public (30.6%; 95% CI 17.2 to 43.9; \( P < 0.0001 \)) after data was published, was observed.

### Barriers and facilitators to data sharing

Six barriers and four facilitators were identified across seven of the 14 domains of the TDF, including: knowledge, skills, environmental context and resources, social influences, professional role and identity, beliefs about capabilities and beliefs about consequences (Table 2). Domains that were not identified included optimism, intention, behavioural regulation, goals, emotion, memory or attention and decision processes. The primary barriers identified were coded to knowledge, and beliefs about consequence. Many participants indicated that protecting the confidentiality of their data to ensure ethical sharing (82.4%) and having a lack of understanding of how to share data (71%), were barriers to data sharing. The primary facilitators were coded to social influences and beliefs about consequence, where a clear understanding of the benefit of data sharing (77.9%), and trust in those requesting or using the data (77.9%), promoted data sharing. These are described in more detail below.

[Insert Table 2 about here]
Responses to capability measures

Researcher's psychological capabilities influenced their data sharing practices, defined within the COM-B model as the capacity to engage in the necessary thought processes, such as comprehension and reasoning (Atkins et al., 2017).

Respondents identified that a lack of understanding in the data sharing process deterred them from data sharing (71.0%) and they would be more motivated to share data if they had a better understanding of the process. Another capability barrier was the lack of skill to conduct the data sharing process, with 53.2% reporting that they do not know how to share data and 62.3% did not know where or who to share it with. In both instances, one-fifth of researchers were unsure whether they had the skills or knowledge related to data sharing. ‘I don't really understand the processes around data sharing, and I have many concerns about how the data is treated, where it ends up, how it is interpreted, etc’ (Participant 43).

Responses to opportunity measures

Barriers and facilitators within the social and physical research environments of the COM-B model shaped the opportunities for researchers to share their data. Social opportunity refers to factors that influence the way that we think about things (i.e. cultural norms and social cues), and physical opportunity is afforded by the environment (i.e. time and location). To measure opportunity factors, participants were asked questions about capacity, regulative pressures and social influences.

Many respondents (67.5%) advised that they did not have the appropriate physical opportunity (i.e. time or resources) to prepare their data for sharing; however, many would if it were a funding requirement (72.7%), institutional requirement (63.6%) or journal requirement (59.7%). ‘Preparing data to ensure it is not identifiable can be
complex and problems may not be immediately obvious, e.g. institutions can be
identified by bed numbers, yet this can be important data’ (Participant 71).

The notion that data should be shared to create greater impacts for the wider public
was explored through a measure of social opportunity. Respondents were more
willing to share data if they knew it had public benefit (77.9% agreed, 3.9%
disagreed and 18.2% were unsure).

Responses to motivation measures

Several barriers and facilitators tapped into researcher’s motivations, which are
defined as the brain processes which direct our decisions and behaviours. The
COM-B model differentiates between automatic motivation (i.e. emotions and
impulses), and reflective motivation (i.e. evaluations and plans). Survey questions
measured factors influencing motivation related to scholarly gain or risk,
responsibility, integrity of practice, and integrity of others. The first two questions
related to automatic motivation, involving emotional reactions and desires (i.e. wants
and needs) that drive states and reflex responses. Seventy percent of respondents
advised that they would be motivated to share their data if they were guaranteed
proper credit for its use; 66.2% if it increased the visibility or impact of their research.

Just over one-half of respondents (55.8%) did not know whether it was their
responsibility to share data. ‘Why would I need to share my data, other than with my
team and potentially in publication…current practices appear sufficient?’ (Participant
41).

Respondents had strong beliefs about the consequences of data sharing. This
referred to the integrity of practice, where 85.7% reported concern about protecting
the confidentiality of their data and 79.2% were concerned about the ethics of data
Similarly, 77.9% indicated that they would be more willing to share data with someone that they trusted. ‘Where it was not considered ethically appropriate to do so, or I was compelled to do so by the institution without a valid explanation of why; I would be deterred from sharing’ (Participant 50).

Lastly, reflective motivation was assessed. About one-half were concerned that their research would be stolen (53.2%), 79.2% were concerned it might be misinterpreted or misused, and 75.3% wanted to protect their intellectual property. ‘People, when pushed to the extreme for research and publications are happy to steal other’s ideas and would be quite capable of stealing their data’ (Participant 80).

**Identifying intervention functions and target behaviours**

Following application of the BCW, intervention functions were identified to address barriers and facilitators (Table 2). It is possible to see how one intervention function may serve to influence more than one behaviour. For example, incorporating modelling techniques (i.e. demonstration of the behaviour) would address five of the seven behaviours. Intervention functions that were not aligned to the target behaviours, include incentivisation and coercion.

Considering that behaviours do not occur in isolation of one another, the research team critically reviewed each behaviour and where appropriate grouped co-dependent behaviours. These were: (1) understanding the data sharing process (knowledge) and knowing how, where and who to share with (physical skills); and (2) protecting the confidentiality of data and safeguarding intellectual property from being stolen or data being misinterpreted or misused (both beliefs about consequences). This resulted in eight key behaviours that influence data sharing and should be addressed through targeted intervention strategies (see Box 2).
Box 2. Target behaviours to promote data sharing in health research.

1. The researcher understands and is competent to conduct data sharing practices.
2. The researcher prioritises their time and resources to prepare data for sharing.
3. The researcher shares data to meet funding, institutional and/or journal requirements.
4. The researcher shares data for public and/or patient benefit.
5. The researcher receives credit and increased visibility of their research following data sharing.
6. The researcher understands data sharing responsibilities.
7. The researcher understands the way in which the confidentiality and safety of data are protected and how to share data ethically.
8. The researcher shares data with trusted and credible people.
Discussion

This study applied a systematic and theoretical process to identify barriers and facilitators that influence data sharing practices among researchers. It has also filled a gap in the research by conceptualising the findings more broadly within researcher’s capability, opportunity, and motivation, and provides a more granular understanding of behaviours which influence data sharing. Application of the BCW and TDF added substantial strength to the study, grounding it in evidence-based theory.

A key observation of this study was that while several individual behaviours can be targeted to influence data sharing, the interaction between these behaviours is complex, and single solutions to single behaviours are unlikely to suffice. Strategies to resolve specific barriers may not advance data sharing if related barriers are not addressed, or if fundamental barriers are unchanged; many of the barriers and facilitators were opposites of one another or dependent on one another. For example, in terms of the societal benefits and value, the public and patient benefit to data sharing was an important facilitator, while a lack of understanding the duty to share data inhibits the behaviour. To this end, it is commonly recognised that there can be no “one size fits all” when it comes to a data sharing strategy. In most cases, disciplinary or user communities need to develop their own data sharing standards and practices that suit their needs, and the combinations of different factors need to be considered rather than looking at individual barriers and facilitators in a stand-alone fashion (Zuiderwijk and Spiers, 2019). The best practices from one discipline cannot simply be transferred to other disciplines without considering the unique characteristics and addressing the individual challenges. For instance, atmospheric science and oceanography have well-developed traditions of open access, whereas...
ecology, medicine and social sciences do not (Tenopir et al., 2011; Zuiderwijk and Spiers, 2019).

In developing the next steps of our data sharing strategy, we recognise that information generated or collected from health research may contain large volumes of highly sensitive data, have explicit privacy and/or security considerations, or a degree of commercialisation. In order to protect privacy, confidentiality and respect the terms under which participants consented to take part in the original study, data needs to be planned, collected and stored in such a way that is appropriate. In some cases, it may not be possible to share or reuse data. This is explored within the larger cohort of this study, where the majority (65.5%) of researchers obtain specific consent for their studies and collect data in a non-identifiable format (69.0%), limiting it for use in the immediate project and not permitting data linkage or sharing for future projects (Krahe et al., 2019a). Considering this, it is now more important than ever to educate researchers about planning their research to enable sharing and reuse which involves the incorporation of considered and appropriate consent (Cheah et al., 2015; O'Keefe and Connolly, 2010). This is one example of a contextual challenge that presents within the health research discipline that must be reflected within the strategy; perhaps learning from other disciplines where sharing is more commonplace.

Applying the COM-B model raised our awareness of the significant barriers to perceived capability, opportunity and motivation to data sharing, but also provided guidance on the most appropriate methods needed to elicit behaviour change. We believe that this behavioural approach to intervention design will allow more informed decisions about which “active ingredients” to include in the intervention strategy and will assist in selecting the right mix of strategies to drive behaviour change. The
capability, opportunity and motivations of individual researchers to share data are strongly influenced by contextual factors (i.e. institutional policies and regulations on sharing data or the degree to which sharing data is encouraged by supervisors and colleagues) and hence cannot be isolated from institutional context.

The next step in this study is to select intervention functions aimed at overcoming the barriers and enhancing the facilitators identified. Research team members with expertise in data sharing best practices, intervention design and behaviour change, will consider which intervention functions are most appropriate or have the best potential chance of success, and how they would best be delivered. This process will be guided by the APEASE criteria (Affordability, Practicability, Effectiveness/Cost-Effectiveness, Acceptability, Safety/Side Effects, Equity) to make context-based decisions (Michie et al., 2014). Interventions will be designed that aim to reduce the barriers and increase the likelihood (facilitators) that researchers will partake in data sharing practices. While several interventions could be applied, a pragmatic approach will be taken, in combination with further stakeholder consultation prior to piloting. The effectiveness of the selected interventions will be tested using mixed methods to evaluate which components are the most effective in changing behaviour.

**Implications of our findings**

The advancement of science thrives on the sharing and accessibility of data, but despite this, it is yet to become common practice in research. A major strength of this study is that a systematic approach was applied which strengthens the theoretical foundations towards the development of data sharing interventions in the research setting. It contextualises two important factors of research best practice and
442 behaviour change, and while the process was time intensive, it provides a foundation
443 for the next stage of intervention design, implementation and evaluation.

444 Limitations of this study
445 In this study a possible limitation is the potential lack of generalisability of the results,
446 which in part is due to the limited survey completion rate. It is feasible that factors
447 perceived as important for this cohort could be different in other settings with varying
448 attitudes, capabilities and/or research culture. While our outcomes are likely to be
449 representative, or at least comparable to similar health research settings, it is
450 important to consider the contextual intricacies that influence behaviour. For
451 example, our cohort is predominantly (57.0%) representative of researchers with 10
452 or more years of experience; where compared to researchers with less experience
453 (i.e. 1-5 years or 5-10 years), may likely hold different attitudes and practices based
454 upon their exposure to and agility towards data sharing. Central to the recent Wiley
455 Survey is growing researcher demand for open access and open research, where
456 older, more experienced researchers require different forms of open research
457 centred around guidelines for the sharing of data and improved measurement of the
458 impact of research (Roscoe, 2020). We recommend using this study as a guide to
459 understand data sharing behaviours but advise developing localised intervention
460 strategies for different contexts. Another limitation is the iterative nature of the
461 process, which is often navigated in a bi-directional state due to ongoing consultation
462 and further consideration or refinement. Despite this, it is the first study to provide a
463 behavioural analysis of the data sharing practices of health researchers and the
464 rigorous method used merits consideration.
Conclusion

This is to our knowledge, the first study that has assessed and quantified barriers and facilitators to data sharing in health research. Using a theoretically informed methodology, we have identified that while the barriers to improving data sharing behaviours are considerable, opportunities for improvement do exist. Interventions that address specific behaviours are important towards motivating participation in data sharing and open researcher practices. It is hoped that the process we have described may be used by others working in research, health information/informatics, library and information services, and/or those interested in the adoption and acceleration of research practices, capabilities and solutions.
Author’s contributions

All authors meet the criteria for authorship and contributed to the conceptualisation and design of this study. Michelle Krahe (MK) was responsible for the acquisition, analysis of data and drafted the article. Malcolm Wolski (MW), Sharon Mickan (SM) and Julie Toohey (JT) aided in the interpretation of the data. Paul Scuffham (PS) and Sheena Reilly (SR) provided critical editing, and final approval.

Declaration of conflicting interests

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Data availability statement

The data that support the findings of this study are openly available in Figshare at: https://doi.org/10.6084/m9.figshare.11521890.v1
References


Australian Research Council (2017) ARC open access policy. Canberra, Australia: Australian Government.


Ferguson L (2014) How and why researchers share data (and they don't). In: The Wiley Network. Available at: www.wiley.com/network/researchers/licensing-


### Tables and figures

#### Table 1. Characteristics of respondents (n = 77).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appointment type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td>45</td>
<td>58.4</td>
</tr>
<tr>
<td>Adjunct</td>
<td>4</td>
<td>5.2</td>
</tr>
<tr>
<td>Research Fellow</td>
<td>12</td>
<td>15.6</td>
</tr>
<tr>
<td>Research Assistant</td>
<td>4</td>
<td>5.2</td>
</tr>
<tr>
<td>Postgraduate Student (PhD)</td>
<td>12</td>
<td>15.6</td>
</tr>
<tr>
<td><strong>Health discipline</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allied Health Science</td>
<td>10</td>
<td>13.0</td>
</tr>
<tr>
<td>Applied Psychology</td>
<td>14</td>
<td>18.2</td>
</tr>
<tr>
<td>Human Services and Social Work</td>
<td>5</td>
<td>6.5</td>
</tr>
<tr>
<td>Medical Sciences</td>
<td>25</td>
<td>32.5</td>
</tr>
<tr>
<td>Medicine</td>
<td>4</td>
<td>5.2</td>
</tr>
<tr>
<td>Nursing and Midwifery</td>
<td>11</td>
<td>14.2</td>
</tr>
<tr>
<td>Pharmacy</td>
<td>2</td>
<td>2.6</td>
</tr>
<tr>
<td>Other(^a)</td>
<td>6</td>
<td>7.8</td>
</tr>
<tr>
<td><strong>Research Experience (years)</strong>(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>9</td>
<td>13.8</td>
</tr>
<tr>
<td>5-10</td>
<td>19</td>
<td>29.2</td>
</tr>
<tr>
<td>&gt; 10</td>
<td>37</td>
<td>57.0</td>
</tr>
</tbody>
</table>

\(^a\) researcher working for the Institute, not one specific area

\(^b\) excluding postgraduate students (n = 65)
Table 2. Behavioural analysis of data sharing among health researchers using the TDF and BCW.

<table>
<thead>
<tr>
<th>TDF domain</th>
<th>COM-B model</th>
<th>Perceived barrier / facilitator to data sharing among health researchers</th>
<th>Intervention function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>Psychological Capability</td>
<td>Barrier: Lack of understanding how to share data</td>
<td>Education and training</td>
</tr>
<tr>
<td>Physical Skills</td>
<td>Physical Capability</td>
<td>Barrier: Lack of experience (i.e. how, where and who)</td>
<td>Training and enablement</td>
</tr>
<tr>
<td>Environmental Context and Resources</td>
<td>Physical Opportunity</td>
<td>Barrier: Resources (i.e. time) are inadequate to prepare data for sharing</td>
<td>Training, restriction, environmental restructuring and enablement</td>
</tr>
<tr>
<td>Social Influences</td>
<td>Social Opportunity</td>
<td>Facilitator: Understanding the public or patient benefits of data sharing</td>
<td>Restriction, environmental restructuring, modelling and enablement</td>
</tr>
<tr>
<td>Professional Role and Identity</td>
<td>Automatic Motivation</td>
<td>Facilitator: Improving visibility of research and ensuring credit (for use) is received</td>
<td>Education, persuasion and modelling</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Beliefs About Capabilities</td>
<td>Barriers: Uncertainty about who is responsible for data sharing</td>
<td>Beliefs About Consequences: Acceptance of the truth, reality, or validity about outcomes of a behaviour in a given situation. TDF domains not covered: optimism; reinforcement; intentions; goals; memory, attention and decision processes; emotion, behavioural regulation.</td>
<td>Barriers: Protecting data confidentiality in the absence of adequate sharing guidelines</td>
</tr>
<tr>
<td>Beliefs About Consequences</td>
<td>Reflective Motivation</td>
<td>Reflective Motivation: Trust in those requesting or using the data</td>
<td>Barriers: Safeguarding intellectual property from being stolen, or data being misinterpreted or misused</td>
</tr>
</tbody>
</table>

TDF = theoretical domains framework; BCW = behaviour change wheel; COM-B = capability, opportunity and motivation.

- Definition of TDF domains adapted from Atkins et al. (2017): **Knowledge**: An awareness or existence of something; **Physical Skills**: An ability or proficiency acquired through practice; **Environmental Context and Resources**: A person’s situation or environment; **Social Influences**: Process that can change their thoughts, feelings, or behaviours; **Beliefs About Capabilities**: Acceptance of the truth, reality, or validity about an ability, talent or facility that can be used constructively; **Beliefs About Consequences**: Acceptance of the truth, reality, or validity about outcomes of a behaviour in a given situation. TDF domains not covered: optimism; reinforcement; intentions; goals; memory, attention and decision processes; emotion, behavioural regulation.
intervention functions (adapted from Michie et al. (2011)): **Education**: Increase knowledge or understanding; **Persuasion**: Using communication to induce positive or negative feelings or stimulate action; **Training**: Imparting skills; **Restriction**: Using rules to reduce the opportunity to engage in the target behaviour; **Environmental Restructuring**: Changing the physical or social context; **Modelling**: Providing an example for people to aspire to or imitate; **Enablement**: Increasing means/reducing barriers to increase capability or opportunity.
Figure 1. The Behaviour Change Wheel. Adapted from Michie et al. (2011)
Figure 2. Places where researchers typically source their data.

Note: Respondents could select multiple responses.
Figure 3. The proportion (%) of researchers who share their data external to the research project, as distributed by research data timepoints.

Data are calculated for researchers who had ever shared data external to the research project (n = 29).
Figure 4. What motivates researchers to share their data with: (a) external individuals to the research project and, (b) the wider public.
Figure 5. Researcher willingness (mean ± SEM) to share research data.

- Includes internal and external colleagues; n = represents the denominator excluding “neutral”.
- *correlation is significant at the 0.05 level (2-tailed); **correlation is significant at the 0.001 level (2-tailed).
Supplemental Material

Summary of research data management survey questions and response options used in this study.

<table>
<thead>
<tr>
<th>Survey Question</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Have you ever shared research data outside of your project team?</td>
<td>Yes / No / Unsure</td>
</tr>
<tr>
<td>^Q1.1. If yes, what kind of data was it?</td>
<td>4 options: Draft or preliminary data (pilot experiments); early-stage data (initial experiments); late-stage data (publication ready); post-publications data (shared after manuscript was accepted) + open-ended</td>
</tr>
<tr>
<td>^Q1.2. If yes, why did you choose to share your research data?</td>
<td>7 options: advancing knowledge; collaboration with colleagues; increase the impact and visibility of my research; preservation of the data; public benefit; recommended or required by the funder; transparency and reuse + open-ended</td>
</tr>
<tr>
<td>Q2. Have you ever shared research data publicly?</td>
<td>Yes / No / Unsure</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>^Q2.1. If yes, where did you make your data publicly available?</td>
<td>6 options: as supplementary material in a journal; institutional repository; discipline-specific repository; personal, institutional or project website; informal paths or upon request; at a conference + open-ended</td>
</tr>
<tr>
<td>^Q2.2. If yes, why did you choose to share your research data?</td>
<td>7 options: advancing knowledge; collaboration with colleagues; increase the impact and visibility of my research; preservation of the data; public benefit; recommended or required by the funder; transparency and reuse + open-ended</td>
</tr>
<tr>
<td>Q3. <strong>Before you publish your research, how willing are you to share your data with...?</strong></td>
<td>5-point Likert scale: definitely willing; somewhat willing; neither willing nor unwilling; somewhat willing; not at all</td>
</tr>
<tr>
<td>a) Members of your research team/group</td>
<td></td>
</tr>
<tr>
<td>b) Colleagues at the university</td>
<td></td>
</tr>
<tr>
<td>c) Colleagues in your field</td>
<td></td>
</tr>
<tr>
<td>Q4. After you publish your research, how willing are you to share your data with…?</td>
<td>5-point Likert scale: definitely willing; somewhat willing; neither willing nor unwilling; somewhat willing; not at all willing (dichotomised into two categories: willing; unwilling)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>a) Members of your research team/group</td>
<td></td>
</tr>
<tr>
<td>b) Colleagues at the university</td>
<td></td>
</tr>
<tr>
<td>c) Colleagues in your field</td>
<td></td>
</tr>
<tr>
<td>d) The public at large</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q5. How much do you agree with the following statements…?</th>
<th>5-point Likert scale: strongly agree; somewhat agree, neither agree or disagree; somewhat agree; strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) I do not know where to share my data or who to share it with</td>
<td></td>
</tr>
<tr>
<td>b) I do not know how to share my data</td>
<td></td>
</tr>
<tr>
<td>c) I don’t have the time or resources to prepare my data for sharing</td>
<td></td>
</tr>
<tr>
<td>d) I don’t know if it is my responsibility to share the data</td>
<td></td>
</tr>
<tr>
<td>e) I am concerned about given proper credit for the use of the data</td>
<td></td>
</tr>
<tr>
<td>f) I am concerned that my research will be stolen</td>
<td></td>
</tr>
<tr>
<td>g) I am concerned about the misinterpretation or misuse of the data</td>
<td></td>
</tr>
</tbody>
</table>
h) I am concerned whether it is unethical to share the data
i) I want to protect my intellectual property
j) I want to protect the confidentiality of the data

Q6. Would you be motivated to share your research data if…?
   Yes / No / Unsure
   a) It was a funding requirement
   b) It was an institutional requirement
   c) It was a journal requirement
   d) There was a public/patient benefit
   e) The impact and visibility of my research increased
   f) I was guaranteed proper credit for its use
   g) I had absolute trust in the person requesting the data
   h) I had a better understanding of the data sharing process.

^Multiple responses to question could be selected by respondent