

1
2
3
4 **1 Developing a strategy to improve data sharing in health research: A mixed-**
5
6 **2 methods study to identify barriers and facilitators**
7
8
9 **3**

10
11 4 Michelle A Krahe¹, BA(BiomedSc), PhD
12

13 5 Malcolm Wolski², BBus, MEnvMgt, MInfTech
14

15 6 Sharon Mickan¹, BOccThy, MA(Education), PhD
16

17 7 Julie Toohey³, BA(Info & Lib Studies)
18

19 8 Paul Scuffham^{4,5}, BA, PhD, GAICD, FAHMS
20

21 9 Sheena Reilly¹, BA(SpPath), PhD, FAHMS, FRCSLT, FSPA, FASSA
22
23
24

25 10

26
27 11 ¹Health Group, Griffith University, Southport, QLD, Australia
28

29 12 ² eResearch Services, Griffith University, Nathan, QLD, Australia
30

31 13 ³Library, Griffith University, Logan, QLD, Australia
32

33 14 ⁴Centre for Applied Health Economics, Griffith University, Nathan, QLD, Australia
34

35 15 ⁵Menzies Health Institute Queensland, Griffith University, Southport, QLD, Australia
36
37
38

39 16

40
41 **Corresponding author:** Dr Michelle A. Krahe, Health Executive, Griffith University,
42
43 Parklands Drive, Southport, 4222, QLD Australia. Email: m.krahe@griffith.edu.au
44
45

46 19

47
48 **Short title:** Data sharing behaviour in health research
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **21 Abstract**
4

5 **22 Background:** It is well accepted that data sharing presents new opportunities across
6
7
8 **23** the spectrum of research and is vital towards science that is open; where data are
9
10 **24** easily discoverable, accessible, intelligible, reproducible, replicable, and verifiable.
11
12 **25** Despite this, it is yet to become common practice. Global efforts to develop practical
13
14 **26** guidance for data sharing and open access initiatives are underway, however
15
16
17 **27** evidence-based studies to inform the development and implementation of effective
18
19 **28** strategies are lacking.

20
21
22 **29 Objective:** This study sought to determine the barriers and facilitators to data sharing
23
24 **30** among health researchers and to identify the target behaviours for designing a
25
26
27 **31** behaviour change intervention strategy.

28
29
30 **32 Method:** Data are drawn from a cross-sectional survey of data management
31
32 **33** practices among health researchers from one Australian research institute.
33
34 **34** Determinants of behaviour are theoretically derived using well established
35
36
37 **35** behavioural models.

38
39
40 **36 Results:** The data sharing practices are described for 77 researchers. Six barriers
41
42 **37** and four facilitators are identified. The primary barriers to data sharing include:
43
44 **38** perceived negative consequences, and lack of competency to share data. The
45
46
47 **39** primary facilitators to data sharing include: trust in others using the data, and social
48
49 **40** influence related to public benefit. Intervention functions likely to be most effective at
50
51 **41** changing target behaviours are also identified.

52
53
54 **42 Conclusion:** Results of this study provide a theoretical and evidence-based process
55
56 **43** to understand the behavioural barriers and facilitators of data sharing among health
57
58
59 **44** researchers.
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

45 Implications: Designing interventions that specifically address target behaviours to
46 promote data sharing is important for open researcher practices.

47

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

49

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

51

52 **Article type:** Research

For Peer Review

53 Introduction

54 While data sharing and reuse is more common in some academic disciplines than
55 others, the argument for and against is remarkably similar. Researchers may be
56 motivated to share data if it results in greater visibility or increased citation of their
57 work, if a supportive data sharing culture exists in their workplace, or if they have
58 had positive experiences with open data reuse (Patel, 2016; Piwowar, 2011;
59 Piwowar et al., 2007; Zuiderwijk and Spiers, 2019). Conversely, researchers are
60 demotivated to share data if they are concerned about privacy, fear of reputation or
61 “scooping” of research, have a desire to retain exclusive rights to data, or are
62 concerned about the inflation, misinterpretation or the usability of data (Alter and
63 Vardigan, 2015; Arzberger et al., 2004; Campbell and Bendavid, 2003; Cheah et al.,
64 2015; Kuipers and van der Hoeven J., 2009; Savage and Vickers, 2009; Tenopir et
65 al., 2011; Zuiderwijk and Spiers, 2019). The perceptions and practices of data
66 sharing and data reuse among researchers was examined by Tenopir et al. (2015),
67 who observed significant increases in the willingness and acceptance of sharing, and
68 “actual” data sharing over a 4-year period. This was also met with an increase in the
69 perception of risk associated with data sharing, including the potential for misuse and
70 misinterpretation and concern about the need to publish data before making it
71 became available. The study also describes that receiving acknowledgement, or a
72 formal citation was important when deciding whether to share data. The Wiley Data
73 Insights survey of more than 4500 researchers from 112 countries indicates a range
74 of motivations for data sharing, including: to increase impact and visibility of
75 research, public benefit, transparency and reuse, or journal requirement (Ferguson,
76 2014). Australian researchers were most incentivised to share data to ensure
77 preservation, transparency and reuse, or to satisfy a funding requirement. This

1
2
3 78 survey demonstrates that support for open researcher practices may not directly
4
5 79 translate into data sharing practices; rather influenced by the increased scientific
6
7 80 competitiveness of research, the research environment specific to the country, or the
8
9 81 appeal for commercial application.
10
11
12

13 82 In Australia, around 75% of universities have an open access mandate, policy or
14
15 83 statement, including the two major funding agencies: the National Health and
16
17 84 Medical Council and the Australian Research Council (Australian Research Council,
18
19 85 2017; National Health and Medical Research Council, 2018). Despite this support,
20
21 86 the one shortcoming and arguably the most important feature of open practices, is
22
23 87 the lack of participation in data sharing from the research community (Ali-Khan et al.,
24
25 88 2017; Hauessler, 2011). Tenopir and colleagues (2011) report that < 6% of
26
27 89 international researchers make “all” their data available, and while only 36% indicate
28
29 90 that others could easily access their data, approximately 75% agreed that they would
30
31 91 use other researcher’s datasets if easily accessible – these findings are also
32
33 92 supported by others (Federer et al., 2015; Hickson et al., 2016).
34
35
36
37
38

39 93 ***How can researchers be encouraged to share data?***

40
41 94 Encouraging data sharing amongst the scientific community and perhaps more
42
43 95 importantly, understanding what is “ideal” and what is the “reality” of data sharing is a
44
45 96 highly complex and contextual challenge (Bezuidenhout, 2019). In an extensive
46
47 97 review of the literature only one evidence-based incentive for data sharing was
48
49 98 identified (Rowhani-Farid et al., 2017). This was a digital badge created by the
50
51 99 Center for Open Science and tested in the Journal of Psychological Science. The
52
53 100 journal editorial team awarded badges to those authors who voluntarily applied and
54
55 101 proved they met the criteria for open data and open materials upon article
56
57
58 102 acceptance. Sharing rates greatly improved over three years with the introduction of
59
60

1
2
3 103 the digital badge (1.5% to 39.4%), but upon review, data and materials had
4
5 104 somehow disappeared from public view (Kidwell et al., 2016). Several other
6
7 105 evidence-based strategies were identified, most were technological strategies
8
9 106 (~75%) such as the introduction of data systems to manage and store data, and
10
11 107 others included collaborative data environments and workshops, implementation of
12
13 108 data sharing policies, and an open data campaign. However, many strategies lacked
14
15 109 good quality empirical data of their effectiveness.
16
17
18
19

20 110 Conceptual models of how both institutional and individual drivers influence
21
22 111 researcher's data sharing behaviours, indicate the pivotal role that institutions play in
23
24 112 influencing open researcher practices (Ali-Khan et al., 2017; Fecher et al., 2015;
25
26 113 Kim, 2012). The dynamic relationship between the willingness to share data and the
27
28 114 actual sharing of data is largely motivated by rational self-interest rather than pure
29
30 115 altruism (Ali-Khan et al., 2017; Fecher et al., 2015; Levin and Leonelli, 2017; Levin et
31
32 116 al., 2016). The highly competitive nature of research and the importance of
33
34 117 maintaining strong academic metrics, mean that many fear they will miss out on
35
36 118 receiving attribution or credit (Anderson and Schonfeld, 2009; Antman, 2014;
37
38 119 Federer et al., 2015; Spallek et al., 2019). A review of the principles and norms of
39
40
41 120 data sharing in international health research identified four overarching themes: (1)
42
43 121 societal benefits and value; (2) distribution of risk, benefit and burden; (3) respect for
44
45 122 individuals and groups; and (4) public trust and engagement (Kalkman et al., 2019).
46
47 123 These themes are important to consider for data governance frameworks, data
48
49 124 management and sharing policies, and data processes and tools specific to each
50
51 125 research context. Observational studies that have identified patterns and barriers to
52
53 126 data sharing are plentiful, but whilst these provide useful background knowledge,
54
55 127 they do not provide evidence of what can be done to increase data sharing. The
56
57
58
59
60

1
2
3 128 development of successful interventions must be underpinned by local evidence and
4
5 129 rigorous theoretical constructs, to facilitate behaviour change and provide
6
7
8 130 explanation for the mechanism of change.
9

10 11 131 ***Understanding behaviours that influence data sharing***

12
13 132 The use of theoretical frameworks in the design and evaluation of interventions has
14
15 133 been increasingly highlighted by implementation researchers (Eccles et al., 2005;
16
17 134 Kok et al., 2016; Michie and Johnston, 2012; Michie et al., 2005). For the first time,
18
19
20 135 this study will apply established theories to understand behaviours that influence
21
22 136 data sharing of health researchers. The two models to be employed are the
23
24 137 Behaviour Change Wheel (BCW) and the Theoretical Domains Framework (TDF).
25
26
27 138 The BCW is a validated model produced through the synthesis of many behaviour
28
29 139 change theories. It is commonly used to analyse, and design behaviour change
30
31
32 140 interventions (Michie et al., 2014). At the centre of the wheel are three core
33
34 141 components: capability, opportunity and motivation (collectively known as COM-B).
35
36
37 142 These components interact to produce “behaviour” and may explain the data sharing
38
39 143 behaviours of researchers who have different capabilities and opportunities to
40
41 144 change behaviour, regardless of their motivation (Michie et al., 2014; Michie et al.,
42
43 145 2011). The middle ring comprises of nine intervention functions and the outer ring
44
45
46 146 seven policy-driven strategies (Figure 1). The three key steps to using the BCW to
47
48 147 design behaviour change interventions includes: (1) understanding the behaviour
49
50 148 (COM-B); (2) identifying intervention options (middle ring); and (3) identifying content
51
52
53 149 and implementation options (outer ring).
54
55

56 150

57
58
59 151 [Insert Figure 1 about here]
60

152

153 In addition, the TDF is based on psychological and organisational theories to guide
 154 analysis and implementation of evidence-based practice (Michie et al., 2005). The
 155 TDF system consists of 14 domains that categorise influences on behaviour and are
 156 mapped to the COM-B components of the BCW in the analysis and design of
 157 targeted interventions (Box 1) (Atkins et al., 2017; Cane et al., 2012). The BCW and
 158 TDF have both been used to analyse and design behavioural change interventions in
 159 a variety of health settings, such as: reducing sugar free intake in adults (Rawahi et
 160 al., 2018), implementing telemedicine consultations for paediatric care (Bele et al.,
 161 2019), refining injury management in hospitals (Curtis et al., 2017), and
 162 understanding physical activity in overweight and obese pregnant women (Flannery
 163 et al., 2018). Despite the current shift towards more open data in research, there are
 164 seemingly no theoretically informed explorations; our understanding is limited to
 165 pragmatic reports.

166

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

TDF domain	COM-B component
Behavioural Regulation; Cognitive and Interpersonal Skills; Knowledge; Memory, Attention and Decision Processes	Psychological Capability
Physical Ability (Skills)	Physical Capability
Environmental Context and Resources	Physical Opportunity
Social Influences	Social Opportunity

Beliefs About Capabilities; Beliefs About Consequences; Emotion; Goals; Intentions; Optimism; Social/Professional Role and Identity	Reflective Motivation
Reinforcement; Emotion	Automatic Motivation

168

169 ***Study objectives***

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

- 173 a) Describe the current data sharing behaviours of health researchers using a
174 qualitative descriptive design.
- 175 b) Conduct a behavioural diagnosis of researcher data sharing behaviours using
176 the TDF and COM-B model.
- 177 c) Identify target behaviours and characterise intervention functions likely to be
178 most effective at changing target behaviours.

1
2
3 **179 Method**

4
5 **180 Study design and recruitment**

6
7
8 181 A cross-sectional survey study was conducted to evaluate the data management
9
10 182 practices of health researchers from one research institute at Griffith University,
11
12 183 Australia. Full study methods and results have been previously published (Krahe et
13
14 184 al., 2019a). The present study employed mixed-methods analysis to a selection of
15
16 185 the survey data (not previously reported) (Sandelowski, 2000). A copy of the survey
17
18 186 questions used in the analysis are available as Online Supplemental Material. This
19
20 187 study and all documentation received ethical approval from the Griffith University
21
22 188 Human Research Ethics Committee on 3 July 2017 (Reference number 2017/457).
23
24
25
26
27 189 Participants were eligible if they were employed at the university in a permanent,
28
29 190 temporary or casual appointment and were actively involved in research. Potential
30
31 191 participants (n = 401) were identified via the membership database and approached
32
33 192 by means of email with a request to participate in our online survey. The initial email
34
35 193 was sent with a link to the survey. If no response was received within two weeks,
36
37 194 one reminder email was sent. All participants were recruited between July and
38
39
40
41 195 August 2017. LimeSurvey (GmbH, v1.9X, Hamburg Germany) was used to
42
43 196 administer the survey
44
45

46 **197 Data sharing behaviours of researchers**

47
48 198 The main outcome measures were derived from survey questions about researcher's
49
50 199 current data sharing practice, their willingness to share data, and responses to
51
52 200 statements designed to explore components of data sharing behaviours. Descriptive
53
54 201 statistics were used to express categorical variables as counts and percentages.
55
56 202 Response options for questions about willingness to share data are dichotomised
57
58 203 into two categories: willing (and somewhat willing); and unwilling (somewhat
59
60

1
2
3 204 unwilling or not at all willing). These were coded as: willing = 1, and unwilling = 2.
4
5 205 While the original survey provided a neutral option (neither willing nor unwilling) it
6
7 206 was treated as “missing” in this analysis, as it does not provide any information about
8
9
10 207 the level of agreement to the statement. A paired-samples t-test was conducted to
11
12 208 compare willingness to share data by researchers before and after publishing. The
13
14 209 mean and 95% CI for the change in willingness is also given. Quantitative survey
15
16 210 data are imported into SPSS (IBM version 25.0) and analysed using descriptive
17
18 211 statistics.

22 ***Behavioural analysis and diagnosis***

23
24 213 Survey data are coded based on the definitions of the TDF domains (Cane et al.,
25
26 214 2012) and exemplar quotes illustrating these were identified in this process.
27
28 215 Quantitative survey data are deductively mapped to the TDF domains and then to
29
30 216 behavioural components of the COM-B model. Discussion occurred with members of
31
32 217 the research team to ensure appropriate coding. Behavioural themes were identified
33
34 218 and categorised as barriers (requiring modification) or facilitators (needing
35
36 219 enhancing).

41 ***Identify intervention functions to change target behaviours***

42
43 221 Applying the BCW, a well validated and widely used framework of behaviour change
44
45 222 (Michie et al., 2014), intervention functions that correlate with the TDF and COM-B
46
47 223 component for each behaviour were identified. These intervention functions are the
48
49 224 “active ingredients” of intervention strategies.
50
51
52
53
54
55
56
57
58
59
60

225 **Results**

226 ***Participant characteristics***

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

234

235 [Insert Table 1 about here]

236

237 To understand the data characteristics that researchers were working with,
238 respondents were asked to indicate where they typically sourced their research data
239 (Krahe et al., 2019b). A list of primary (information collected and processed by the
240 researcher) and secondary (information retrieved through pre-existing sources) data
241 sources were provided, with the most common being surveys (66.2%), interviews
242 (54.5%), and experiments (46.8%) (Figure 2).

243

244 [Insert Figure 2 about here]

245

246 ***Data sharing behaviours of researchers***

247 Almost two-fifths of respondents (37.7%, n = 29) had ever shared data outside their
248 immediate research/project team. Of these, 65.5% (n = 19/29) shared data only after

1
2
3 249 the research data had been published (Figure 3). Only 15.6% (n = 12/77) of
4
5 250 researchers had ever shared their data publicly as: supplementary material in a
6
7 251 journal (58.3%), through an institutional (41.7%) or discipline-specific (25%)
8
9
10 252 repository, at a conference (33.3%), on a personal or project website (16.7%), or
11
12 253 through direct data requests (8.3%) (results not tabulated). The motivations for
13
14 254 researchers that shared their data with others outside of the research project and
15
16
17 255 more publicly (i.e. open access) are presented in Figure 4(a) and (b).
18
19

20 256

21
22
23 257 [Insert Figure 3 about here]

24
25
26 258 [Insert Figure 4 about here]

27
28
29 259

30
31
32 260 The attitudes of researchers towards data sharing was further explored through their
33
34 261 willingness to share data at two time points: (1) before data had been published; and
35
36 262 (2) after data had been published. Figure 5 illustrates that overall researchers were
37
38 263 most willing ($97.2\% \pm 1.97$, n = 69/71) to share data before it was published with
39
40 264 members of their research group, and least willing ($24.5\% \pm 6.2$, n = 12/49) to share
41
42 265 data with the wider public. After data had been published, all respondents (100%, n =
43
44 266 71/71) were willing to share data within their research group, $91\% \pm 3.8$ (n = 51/56)
45
46 267 were willing to share with colleagues in their organisations, $89\% \pm 4.2$ (n = 49/55)
47
48 268 with colleagues in their field, and just over one-half ($55.1\% \pm 7.1$, n = 27/49) with the
49
50 269 wider public.
51
52

53
54
55 270

56
57
58 271 [Insert Figure 5 about here]
59
60

1
2
3 272
4
5

6 273 A significant increase in the willingness to share data with colleagues in the
7
8 274 organisation (14.2%; 95% CI 4.8 to 23.7; $P < 0.01$), colleagues in the field (29%;
9
10 275 95% CI 16.6 to 41.4; $P < 0.0001$), and with the wider public (30.6%; 95% CI 17.2 to
11
12 276 43.9; $P < 0.0001$) after data was published, was observed.

13
14
15
16 277 ***Barriers and facilitators to data sharing***

17
18 278 Six barriers and four facilitators were identified across seven of the 14 domains of
19
20 279 the TDF, including: knowledge, skills, environmental context and resources, social
21
22 280 influences, professional role and identity, beliefs about capabilities and beliefs about
23
24 281 consequences (Table 2). Domains that were not identified included optimism,
25
26 282 intention, behavioural regulation, goals, emotion, memory or attention and decision
27
28 283 processes. The primary barriers identified were coded to knowledge, and beliefs
29
30 284 about consequence. Many participants indicated that protecting the confidentiality of
31
32 285 their data to ensure ethical sharing (82.4%) and having a lack of understanding of
33
34 286 how to share data (71%), were barriers to data sharing. The primary facilitators were
35
36 287 coded to social influences and beliefs about consequence, where a clear
37
38 288 understanding of the benefit of data sharing (77.9%), and trust in those requesting or
39
40 289 using the data (77.9%), promoted data sharing. These are described in more detail
41
42
43
44
45 290 below.

46
47
48
49 291

50
51 292 [Insert Table 2 about here]

52
53
54 293
55
56
57
58
59
60

1
2
3 294 *Responses to capability measures*
4

5 295 Researcher's psychological capabilities influenced their data sharing practices,
6
7 296 defined within the COM-B model as the capacity to engage in the necessary thought
8
9
10 297 processes, such as comprehension and reasoning (Atkins et al., 2017).
11
12 298 Respondents identified that a lack of understanding in the data sharing process
13
14 299 deterred them from data sharing (71.0%) and they would be more motivated to share
15
16 300 data if they had a better understanding of the process. Another capability barrier was
17
18 301 the lack of skill to conduct the data sharing process, with 53.2% reporting that they
19
20 302 do not know how to share data and 62.3% did not know where or who to share it
21
22 303 with. In both instances, one-fifth of researchers were unsure whether they had the
23
24 304 skills or knowledge related to data sharing. *'I don't really understand the processes*
25
26 305 *around data sharing, and I have many concerns about how the data is treated,*
27
28 306 *where it ends up, how it is interpreted, etc'* (Participant 43).
29
30
31
32
33

34 307 *Responses to opportunity measures*
35

36 308 Barriers and facilitators within the social and physical research environments of the
37
38 309 COM-B model shaped the opportunities for researchers to share their data. Social
39
40 310 opportunity refers to factors that influence the way that we think about things (i.e.
41
42 311 cultural norms and social cues), and physical opportunity is afforded by the
43
44 312 environment (i.e. time and location). To measure opportunity factors, participants
45
46 313 were asked questions about capacity, regulative pressures and social influences.
47
48 314 Many respondents (67.5%) advised that they did not have the appropriate physical
49
50 315 opportunity (i.e. time or resources) to prepare their data for sharing; however, many
51
52 316 would if it were a funding requirement (72.7%), institutional requirement (63.6%) or
53
54 317 journal requirement (59.7%). *'Preparing data to ensure it is not identifiable can be*
55
56
57
58
59
60

1
2
3 318 *complex and problems may not be immediately obvious, e.g. institutions can be*
4
5 319 *identified by bed numbers, yet this can be important data' (Participant 71).*
6
7

8 320 The notion that data should be shared to create greater impacts for the wider public
9
10 321 was explored through a measure of social opportunity. Respondents were more
11
12 322 willing to share data if they knew it had public benefit (77.9% agreed, 3.9%
13
14 323 disagreed and 18.2% were unsure).
15
16

17
18 324 *Responses to motivation measures*
19

20 325 Several barriers and facilitators tapped into researcher's motivations, which are
21
22 326 defined as the brain processes which direct our decisions and behaviours. The
23
24 327 COM-B model differentiates between automatic motivation (i.e. emotions and
25
26 328 impulses), and reflective motivation (i.e. evaluations and plans). Survey questions
27
28 329 measured factors influencing motivation related to scholarly gain or risk,
29
30 330 responsibility, integrity of practice, and integrity of others. The first two questions
31
32 331 related to automatic motivation, involving emotional reactions and desires (i.e. wants
33
34 332 and needs) that drive states and reflex responses. Seventy percent of respondents
35
36 333 advised that they would be motivated to share their data if they were guaranteed
37
38 334 proper credit for its use; 66.2% if it increased the visibility or impact of their research.
39
40 335 Just over one-half of respondents (55.8%) did not know whether it was their
41
42 336 responsibility to share data. *'Why would I need to share my data, other than with my*
43
44 337 *team and potentially in publication...current practices appear sufficient?' (Participant*
45
46 338 *41).*
47
48
49
50
51

52
53 339 Respondents had strong beliefs about the consequences of data sharing. This
54
55 340 referred to the integrity of practice, where 85.7% reported concern about protecting
56
57 341 the confidentiality of their data and 79.2% were concerned about the ethics of data
58
59
60

1
2
3 342 sharing. Similarly, 77.9% indicated that they would be more willing to share data with
4
5 343 someone that they trusted. *'Where it was not considered ethically appropriate to do*
6
7 344 *so, or I was compelled to do so by the institution without a valid explanation of why; I*
8
9 345 *would be deterred from sharing'* (Participant 50).

10
11
12
13 346 Lastly, reflective motivation was assessed. About one-half were concerned that their
14
15 347 research would be stolen (53.2%), 79.2% were concerned it might be misinterpreted
16
17 348 or misused, and 75.3% wanted to protect their intellectual property. *'People, when*
18
19 349 *pushed to the extreme for research and publications are happy to steal other's ideas*
20
21 350 *and would be quite capable of stealing their data'* (Participant 80).

22 351 **Identifying intervention functions and target behaviours**

23
24
25
26
27 352 Following application of the BCW, intervention functions were identified to address
28
29 353 barriers and facilitators (Table 2). It is possible to see how one intervention function
30
31 354 may serve to influence more than one behaviour. For example, incorporating
32
33 355 modelling techniques (i.e. demonstration of the behaviour) would address five of the
34
35 356 seven behaviours. Intervention functions that were not aligned to the target
36
37 357 behaviours, include incentivisation and coercion.

38
39
40
41
42 358 Considering that behaviours do not occur in isolation of one another, the research
43
44 359 team critically reviewed each behaviour and where appropriate grouped co-
45
46 360 dependent behaviours. These were: (1) understanding the data sharing process
47
48 361 (knowledge) and knowing how, where and who to share with (physical skills); and (2)
49
50 362 protecting the confidentiality of data and safeguarding intellectual property from
51
52 363 being stolen or data being misinterpreted or misused (both beliefs about
53
54 364 consequences). This resulted in eight key behaviours that influence data sharing and
55
56 365 should be addressed through targeted intervention strategies (see Box 2).

57
58
59
60

1
2
3 366 **Box 2.** Target behaviours to promote data sharing in health research.
4
5

- 6 1. The researcher understands and is competent to conduct data sharing
7 practices.
 - 8 2. The researcher prioritises their time and resources to prepare data for
9 sharing.
 - 10 3. The researcher shares data to meet funding, institutional and/or journal
11 requirements
 - 12 4. The researcher shares data for public and/or patient benefit.
 - 13 5. The researcher receives credit and increased visibility of their research
14 following data sharing.
 - 15 6. The researcher understands data sharing responsibilities.
 - 16 7. The researcher understands the way in which the confidentiality and
17 safety of data are protected and how to share data ethically.
 - 18 8. The researcher shares data with trusted and credible people.
- 19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 367
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

368 Discussion

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

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

1
2
3 393 ecology, medicine and social sciences do not (Tenopir et al., 2011; Zuiderwijk and
4
5 394 Spiers, 2019).

6
7
8 395 In developing the next steps of our data sharing strategy, we recognise that
9
10 396 information generated or collected from health research may contain large volumes
11
12 397 of highly sensitive data, have explicit privacy and/or security considerations, or a
13
14 398 degree of commercialisation. In order to protect privacy, confidentiality and respect
15
16 399 the terms under which participants consented to take part in the original study, data
17
18 400 needs to be planned, collected and stored in such a way that is appropriate. In some
19
20 401 cases, it may not be possible to share or reuse data. This is explored within the
21
22 402 larger cohort of this study, where the majority (65.5%) of researchers obtain specific
23
24 403 consent for their studies and collect data in a non-identifiable format (69.0%), limiting
25
26 404 it for use in the immediate project and not permitting data linkage or sharing for
27
28 405 future projects (Krahe et al., 2019a). Considering this, it is now more important than
29
30 406 ever to educate researchers about planning their research to enable sharing and
31
32 407 reuse which involves the incorporation of considered and appropriate consent
33
34 408 (Cheah et al., 2015; O'Keefe and Connolly, 2010). This is one example of a
35
36 409 contextual challenge that presents within the health research discipline that must be
37
38 410 reflected within the strategy; perhaps learning from other disciplines where sharing is
39
40 411 more commonplace.

41
42
43 412 Applying the COM-B model raised our awareness of the significant barriers to
44
45 413 perceived capability, opportunity and motivation to data sharing, but also provided
46
47 414 guidance on the most appropriate methods needed to elicit behaviour change. We
48
49 415 believe that this behavioural approach to intervention design will allow more informed
50
51 416 decisions about which “active ingredients” to include in the intervention strategy and
52
53 417 will assist in selecting the right mix of strategies to drive behaviour change. The
54
55
56
57
58
59
60

1
2
3 418 capability, opportunity and motivations of individual researchers to share data are
4
5 419 strongly influenced by contextual factors (i.e. institutional policies and regulations on
6
7 420 sharing data or the degree to which sharing data is encouraged by supervisors and
8
9 421 colleagues) and hence cannot be isolated from institutional context.

10
11
12
13 422 The next step in this study is to select intervention functions aimed at overcoming the
14
15 423 barriers and enhancing the facilitators identified. Research team members with
16
17 424 expertise in data sharing best practices, intervention design and behaviour change,
18
19 425 will consider which intervention functions are most appropriate or have the best
20
21 426 potential chance of success, and how they would best be delivered. This process will
22
23 427 be guided by the APEASE criteria (Affordability, Practicability, Effectiveness/Cost-
24
25 428 Effectiveness, Acceptability, Safety/Side Effects, Equity) to make context-based
26
27 429 decisions (Michie et al., 2014). Interventions will be designed that aim to reduce the
28
29 430 barriers and increase the likelihood (facilitators) that researchers will partake in data
30
31 431 sharing practices. While several interventions could be applied, a pragmatic
32
33 432 approach will be taken, in combination with further stakeholder consultation prior to
34
35 433 piloting. The effectiveness of the selected interventions will be tested using mixed
36
37 434 methods to evaluate which components are the most effective in changing
38
39 435 behaviour.

46 436 ***Implications of our findings***

47
48 437 The advancement of science thrives on the sharing and accessibility of data, but
49
50 438 despite this, it is yet to become common practice in research. A major strength of
51
52 439 this study is that a systematic approach was applied which strengthens the
53
54 440 theoretical foundations towards the development of data sharing interventions in the
55
56 441 research setting. It contextualises two important factors of research best practice and
57
58
59
60

1
2
3 442 behaviour change, and while the process was time intensive, it provides a foundation
4
5 443 for the next stage of intervention design, implementation and evaluation.
6
7

8 444 ***Limitations of this study***
9

10 445 In this study a possible limitation is the potential lack of generalisability of the results,
11
12 446 which in part is due to the limited survey completion rate. It is feasible that factors
13
14 447 perceived as important for this cohort could be different in other settings with varying
15
16 448 attitudes, capabilities and/or research culture. While our outcomes are likely to be
17
18 449 representative, or at least comparable to similar health research settings, it is
19
20 450 important to consider the contextual intricacies that influence behaviour. For
21
22 451 example, our cohort is predominantly (57.0%) representative of researchers with 10
23
24 452 or more years of experience; where compared to researchers with less experience
25
26 453 (i.e. 1-5 years or 5-10 years), may likely hold different attitudes and practices based
27
28 454 upon their exposure to and agility towards data sharing. Central to the recent Wiley
29
30 455 Survey is growing researcher demand for open access and open research, where
31
32 456 older, more experienced researchers require different forms of open research
33
34 457 centred around guidelines for the sharing of data and improved measurement of the
35
36 458 impact of research (Roscoe, 2020). We recommend using this study as a guide to
37
38 459 understand data sharing behaviours but advise developing localised intervention
39
40 460 strategies for different contexts. Another limitation is the iterative nature of the
41
42 461 process, which is often navigated in a bi-directional state due to ongoing consultation
43
44 462 and further consideration or refinement. Despite this, it is the first study to provide a
45
46 463 behavioural analysis of the data sharing practices of health researchers and the
47
48 464 rigorous method used merits consideration.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **465 Conclusion**
4

5 466 This is to our knowledge, the first study that has assessed and quantified barriers
6
7 467 and facilitators to data sharing in health research. Using a theoretically informed
8
9
10 468 methodology, we have identified that while the barriers to improving data sharing
11
12 469 behaviours are considerable, opportunities for improvement do exist. Interventions
13
14 470 that address specific behaviours are important towards motivating participation in
15
16 471 data sharing and open researcher practices. It is hoped that the process we have
17
18 472 described may be used by others working in research, health information/informatics,
19
20 473 library and information services, and/or those interested in the adoption and
21
22
23 474 acceleration of research practices, capabilities and solutions.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **475 Author's contributions**
4

5 476 All authors meet the criteria for authorship and contributed to the conceptualisation
6
7 477 and design of this study. Michelle Krahe (MK) was responsible for the acquisition,
8
9 478 analysis of data and drafted the article. Malcolm Wolski (MW), Sharon Mickan (SM)
10
11 479 and Julie Toohey (JT) aided in the interpretation of the data. Paul Scuffham (PS) and
12
13 480 Sheena Reilly (SR) provided critical editing, and final approval.
14
15
16
17

18 **481 Declaration of conflicting interests**
19

20 482 The author(s) declare no potential conflicts of interest with respect to the research,
21
22 483 authorship, and/or publication of this article.
23
24
25

26 **484 Funding**
27

28 485 The author(s) received no financial support for the research, authorship, and/or
29
30 486 publication of this article.
31
32
33

34 **487 Data availability statement**
35

36 488 The data that support the findings of this study are openly available in Figshare at:
37
38 489 <https://doi.org/10.6084/m9.figshare.11521890.v1>
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **490 References**
4

- 5
6 491 Ali-Khan SE, Harris LW and Gold ER (2017) Motivating participation in open science
7
8 492 by examining researcher incentives. *Elife* 6(e29319): 1-12.
9
10 493 Alter GC and Vardigan M (2015) Addressing global data sharing challenges. *Journal*
11
12 494 *of Empirical Research on Human Research Ethics* 10(3): 317-323.
13
14 495 Anderson JR and Schonfeld TL (2009) Data-sharing dilemmas: Allowing
15
16 496 pharmaceutical company access to research data. *Ethics & Human Research*
17
18 497 31(3): 17-19.
19
20 498 Antman E (2014) Data sharing in research: Benefits and risks for clinicians. *British*
21
22 499 *Medical Journal* 348: g237.
23
24 500 Arzberger P, Schroeder P, Beaulieu A, et al. (2004) Science and government: An
25
26 501 international framework to promote access to data. *Science* 303(5665): 1777-
27
28 502 1778.
29
30 503 Atkins L, Francis J, Islam R, et al. (2017) A guide to using the Theoretical Domains
31
32 504 Framework of behaviour change to investigate implementation problems.
33
34 505 *Implementation Science* 12(77): 1-18.
35
36 506 Australian Research Council (2017) ARC open access policy. Canberra, Australia:
37
38 507 Australian Government.
39
40 508 Bele S, Cassidy C, Curran J, et al. (2019) Barriers and enablers to implementing a
41
42 509 virtual tertiary-regional Telemedicine Rounding and Consultation (TRAC)
43
44 510 model of inpatient pediatric care using the Theoretical Domains Framework
45
46 511 (TDF) approach: A study protocol. *BMC Health Services Research* 19(1): 29.
47
48 512 Bezuidenhout L (2019) To share or not to share: Incentivizing data sharing in life
49
50 513 science communities. *Developing World Bioethics* 19(1): 18-24.
51
52
53
54
55
56
57
58
59
60

- 1
2
3 514 Campbell EG and Bendavid E (2003) Data-sharing and data-withholding in genetics
4
5 515 and the life sciences: Results of a national survey of technology transfer
6
7 516 officers. *Journal of Health Care Law and Policy* 6(2): 241-255.
8
9
10 517 Cane J, O'Connor D and Michie S (2012) Validation of the theoretical domains
11
12 518 framework for use in behaviour change and implementation research.
13
14 519 *Implementation Science* 7(37): 1-17.
15
16
17 520 Cheah PY, Tangseefa D, Somsaman A, et al. (2015) Perceived benefits, harms, and
18
19 521 views about how to share data responsibly: A qualitative study of experiences
20
21 522 with and attitudes toward data sharing among research staff and community
22
23 523 representatives in Thailand. *Journal of Empirical Research on Human*
24
25 524 *Research Ethics* 10(3): 278-289.
26
27
28 525 Curtis K, Van C, Lam M, et al. (2017) Implementation evaluation and refinement of
29
30 526 an intervention to improve blunt chest injury management: A mixed-methods
31
32 527 study. *Journal of Clinical Nursing* 26(23-24): 4506-4518.
33
34
35 528 Eccles M, Grimshaw J, Walker A, et al. (2005) Changing the behavior of healthcare
36
37 529 professionals: The use of theory in promoting the uptake of research findings.
38
39 530 *Journal of Clinical Epidemiology* 58(2): 107-112.
40
41
42 531 Fecher B, Friesike S and Hebing M (2015) What drives academic data sharing?
43
44 532 *PLoS One* 10(2): e0118053.
45
46
47 533 Federer LM, Lu YL, Joubert DJ, et al. (2015) Biomedical data sharing and reuse:
48
49 534 Attitudes and practices of clinical and scientific research staff. *PLoS One*
50
51 535 10(6): e0129506.
52
53
54 536 Ferguson L (2014) How and why researchers share data (and they don't). In: The
55
56 537 Wiley Network. Available at: www.wiley.com/network/researchers/licensing-
57
58
59
60

- 1
2
3 538 [and-open-access/how-and-why-researchers-share-data-and-why-they-dont](#)
4
5 539 (accessed 10 October 2019).
6
7
8 540 Flannery C, McHugh S, Anaba AE, et al. (2018) Enablers and barriers to physical
9
10 541 activity in overweight and obese pregnant women: An analysis informed by
11
12 542 the theoretical domains framework and COM-B model. *BMC Pregnancy and*
13
14 543 *Childbirth* 18(178).
15
16
17 544 Hauessler C (2011) Information-sharing in academic and the industry: A comparative
18
19 545 study. *Research Policy* 40(1): 105-122.
20
21
22 546 Hickson S, Poulton KA, Connor M, et al. (2016) Modifying researchers' data
23
24 547 management practices: A behavioural framework for library practitioners.
25
26 548 *International Federation of Library Association Journal* 42(4): 253-265.
27
28
29 549 Kalkman S, Mostert M, Gerlinger C, et al. (2019) Responsible data sharing in
30
31 550 international health research: A systematic review of principles and norms.
32
33 551 *BMC Medical Ethics* 20(21): 1-13.
34
35
36 552 Kidwell MC, Lazarevic LB, Baranski E, et al. (2016) Badges to acknowledge open
37
38 553 practices: A simple, low-cost, effective method for increasing transparency.
39
40 554 *PLoS Biology* 14(5): e1002456.
41
42
43 555 Kim HS (2012) Climate change, science and community. *Public Understanding of*
44
45 556 *Science* 21(3): 268-285.
46
47
48 557 Kok G, Gottlieb NH, Peters GJ, et al. (2016) A taxonomy of behaviour change
49
50 558 methods: An intervention mapping approach. *Health Psychology Review*
51
52 559 10(3): 297-312.
53
54
55 560 Krahe MA, Toohey J, Wolski M, et al. (2019a) Research data management in
56
57 561 practice: Results from a cross-sectional survey of health and medical
58
59 562 researchers from an academic institution in Australia. *Health Information*
60

- 1
2
3 563 *Management Journal*. Epub ahead of print 13 March 2019. DOI:
4
5 564 10.1177/1833358319831318: 1833358319831318.
6
7
8 565 Krahe MA, Toohey J, Wolski M, et al. (2019b) Supplemental Material: Research data
9
10 566 management in practice: Results from a cross-sectional survey of health and
11
12 567 medical researchers from an academic institution in Australia. *SAGE*
13
14 568 *Journals*,. doi.org/10.25384/SAGE.7830878.v1.
15
16
17 569 Kuipers T and van der Hoeven J. (2009) Insight into digital preservation of research
18
19 570 output in Europe: Survey report. Report for the European Commission,
20
21 571 Brussels, Belgium, December.
22
23
24 572 Levin N and Leonelli S (2017) How does one "open" science? Questions of value in
25
26 573 biological research. *Science, Technology, & Human Values* 42(2): 280-305.
27
28
29 574 Levin N, Leonelli S, Weckowska D, et al. (2016) How do scientists define openness?
30
31 575 Exploring the relationship between open science policies and research
32
33 576 practice. *Bulletin of Science, Technology & Society* 36(2): 128-141.
34
35
36 577 Michie S, Atkins L and West R (2014) *The behaviour change wheel: A guide to*
37
38 578 *designing interventions*. Great Britain: Silverback Publishing.
39
40 579 Michie S and Johnston M (2012) Theories and techniques of behaviour change:
41
42 580 Developing a cumulative science of behaviour change. *Health Psychology*
43
44 581 *Review* 6(1): 1-6.
45
46
47 582 Michie S, Johnston M, Abraham C, et al. (2005) Making psychological theory useful
48
49 583 for implementing evidence based practice: A consensus approach. *Quality*
50
51 584 *and Safety in Health Care* 14(1): 26-33.
52
53
54 585 Michie S, van Stralen MM and West R (2011) The behaviour change wheel: A new
55
56 586 method for characterising and designing behaviour change interventions.
57
58 587 *Implementation Science* 6(42): 1-11.
59
60

- 1
2
3 588 National Health and Medical Research Council (2018) Open access policy.
4
5 589 Canberra, Australia: Australian Government.
6
7
8 590 O'Keefe CM and Connolly CJ (2010) Privacy and the use of health data for research.
9
10 591 *Medical Journal of Australia* 193(9): 537-541.
11
12 592 Patel D (2016) Research data management: A conceptual framework. *Library*
13
14 593 *Review* 65(4-5): 226-241.
15
16
17 594 Piwowar HA (2011) Who shares? Who doesn't? Factors associated with openly
18
19 595 archiving raw research data. *PLoS One* 6(7): e18657.
20
21 596 Piwowar HA, Day RS and Fridsma DB (2007) Sharing detailed research data is
22
23 597 associated with increased citation rate. *PLoS One* 2(3): e308.
24
25
26 598 Rawahi SHA, Asimakopoulou K and Newton JT (2018) Factors related to reducing
27
28 599 free sugar intake among white ethnic adults in the UK: A qualitative study.
29
30 600 *BDJ Open* 4(17024): 1-6.
31
32
33 601 Roscoe J (2020) Building new societies: Insights and predictions from the 5th Wiley
34
35 602 Society Member Survey. *Learned Publishing* 33(1): 29-36.
36
37
38 603 Rowhani-Farid A, Allen M and Barnett AG (2017) What incentives increase data
39
40 604 sharing in health and medical research? A systematic review. *Research*
41
42 605 *Integrity and Peer Review* 2(4): 1-10.
43
44
45 606 Sandelowski M (2000) Combining qualitative and quantitative sampling, data
46
47 607 collection, and analysis techniques in mixed-method studies. *Research in*
48
49 608 *Nursing and Health* 23(3): 246-255.
50
51
52 609 Savage CJ and Vickers AJ (2009) Empirical study of data sharing by authors
53
54 610 publishing in PLoS journals. *PLoS One* 4(9): e7078.
55
56
57
58
59
60

- 1
2
3 611 Spallek H, Weinberg SM, Manz M, et al. (2019) Perceptions and attitudes toward
4
5 612 data sharing among dental researchers. *JDR Clinical and Translational*
6
7 613 *Research* 4(1): 68-75.
8
9
10 614 Tenopir C, Allard S, Douglass K, et al. (2011) Data sharing by scientists: Practices
11
12 615 and perceptions. *PLoS One* 6(6): e21101.
13
14 616 Tenopir C, Dalton ED, Allard S, et al. (2015) Changes in data sharing and data reuse
15
16 617 practices and perceptions among scientists worldwide. *PLoS One* 10(8):
17
18 618 e0134826.
19
20
21 619 Zuiderwijk A and Spiers H (2019) Sharing and re-using open data: A case study of
22
23 620 motivations in astrophysics. *International Journal of Information Management*
24
25 621 49: 228-241.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

622 **Tables and figures**

623

624 **Table 1.** Characteristics of respondents (n = 77).

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

625 ^a researcher working for the Institute, not one specific area626 ^b excluding postgraduate students (n = 65)

627 **Table 2.** Behavioural analysis of data sharing among health researchers using the TDF and BCW.

TDF domain ^a	COM-B model	Perceived barrier / facilitator to data sharing among health researchers	Intervention function ^b
Knowledge	Psychological Capability	Barrier: Lack of understanding how to share data	Education and training
Physical Skills	Physical Capability	Barrier: Lack of experience (i.e. how, where and who)	Training and enablement
Environmental Context and Resources	Physical Opportunity	Barrier: Resources (i.e. time) are inadequate to prepare data for sharing Facilitator: Appropriate institutional, funding or publication incentives/requirements are needed	Training, restriction, environmental restructuring and enablement
Social Influences	Social Opportunity	Facilitator: Understanding the public or patient benefits of data sharing	Restriction, environmental restructuring, modelling and enablement

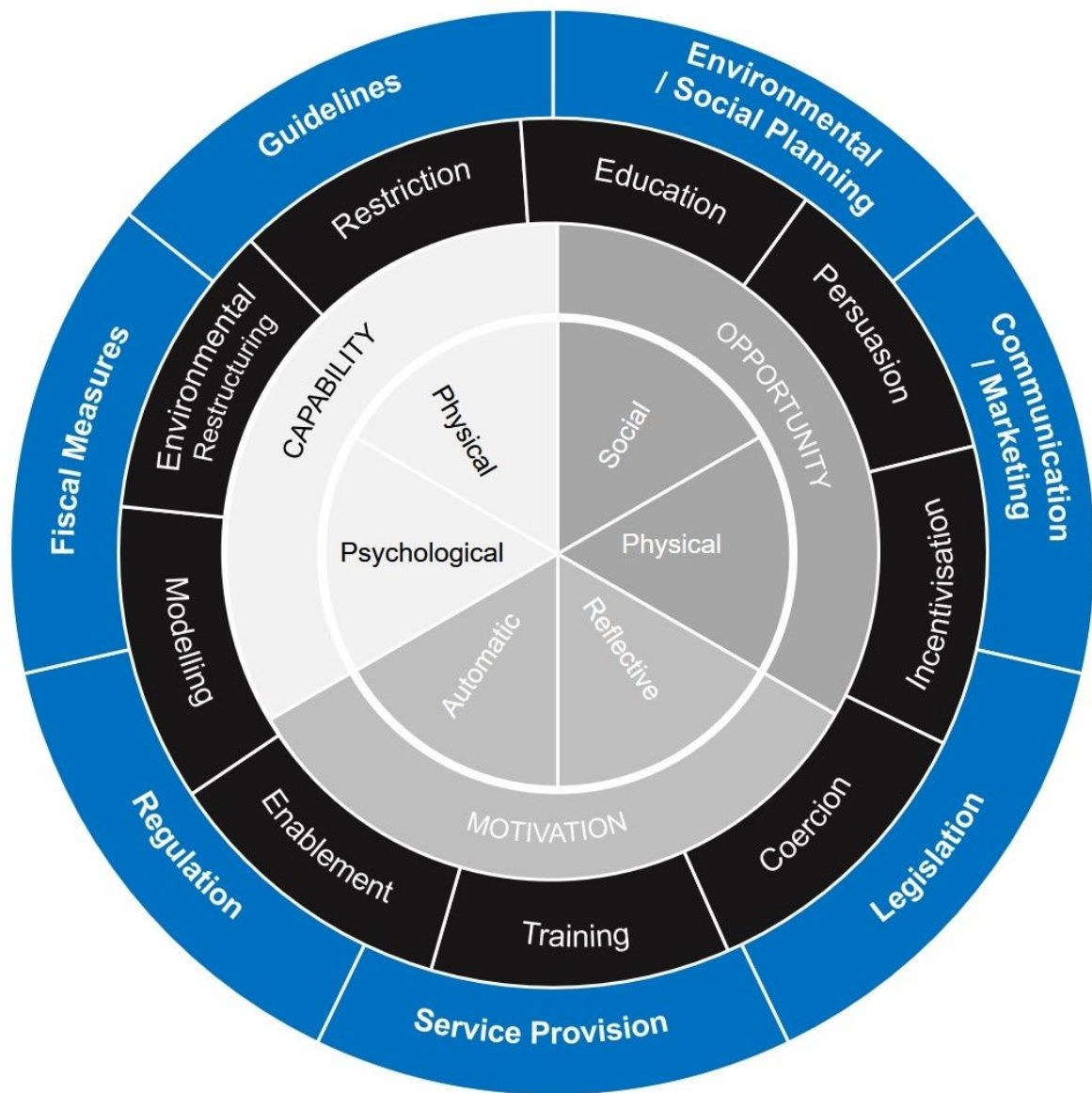
Professional	Automatic	Facilitator: Improving visibility of research and ensuring credit (for use) is received	Education, persuasion and modelling
Role and Identity	Motivation		
Beliefs About Capabilities		Barrier: Uncertainty about who is responsible for data sharing	Education, persuasion, modelling and enablement
	Reflective	Barrier: Protecting data confidentiality in the absence of adequate sharing guidelines	
Beliefs About Consequences	Motivation	Barrier: Safeguarding intellectual property from being stolen, or data being misinterpreted or misused	Education, persuasion and modelling
		Facilitator: Trust in those requesting or using the data	

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

629 ^a Definition of TDF domains adapted from Atkins et al. (2017): Knowledge: An awareness or existence of something; Physical Skills:
630 An ability or proficiency acquired through practice; Environmental Context and Resources: A person's situation or environment;
631 Social Influences: Process that can change their thoughts, feelings, or behaviours; Beliefs About Capabilities: Acceptance of the
632 truth, reality, or validity about an ability, talent or facility that can be used constructively; Beliefs About Consequences: Acceptance
633 of the truth, reality, or validity about outcomes of a behaviour in a given situation. TDF domains not covered: optimism;
634 reinforcement; intentions; goals; memory, attention and decision processes; emotion, behavioural regulation. ^b Definition of

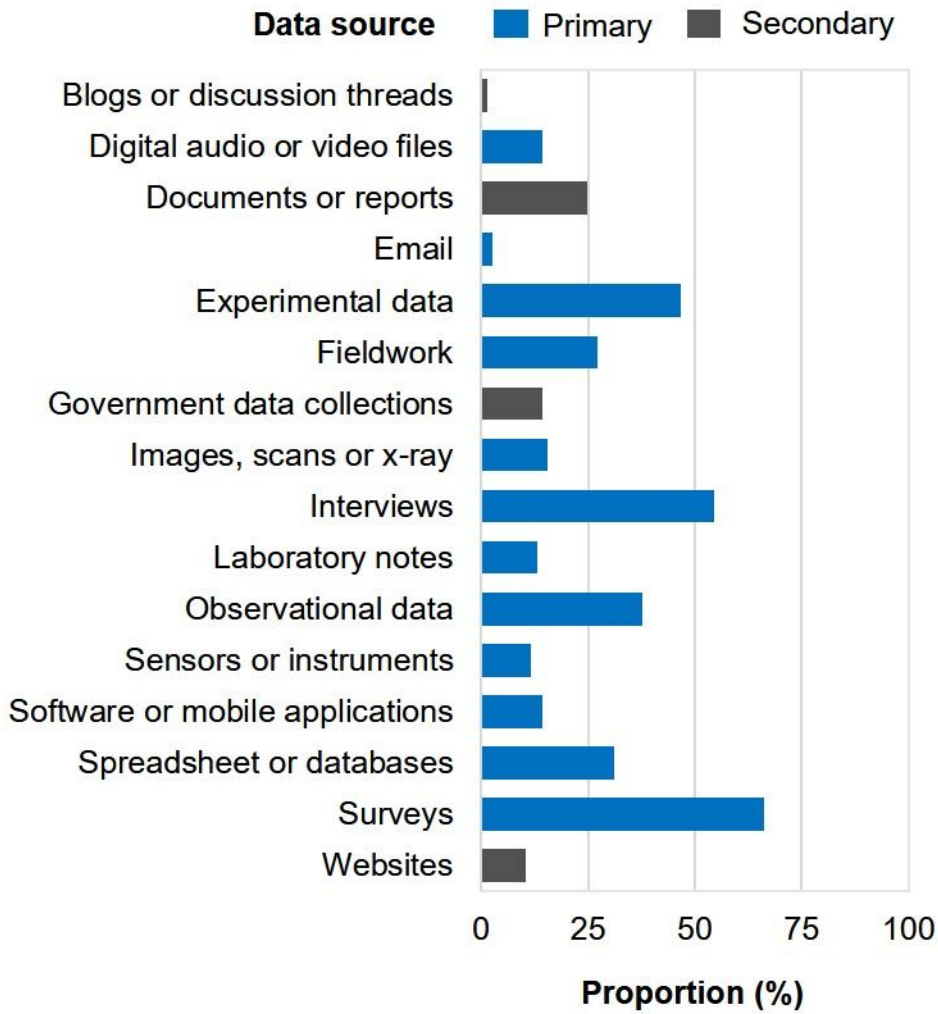
1
2
3 635 intervention functions (adapted from Michie et al. (2011)): Education: Increase knowledge or understanding; Persuasion: Using
4
5 636 communication to induce positive or negative feelings or stimulate action; Training: Imparting skills; Restriction: Using rules to
6
7 637 reduce the opportunity to engage in the target behaviour; Environmental Restructuring: Changing the physical or social context;
8
9 638 Modelling: Providing an example for people to aspire to or imitate; Enablement: Increasing means/reducing barriers to increase
10
11 639 capability or opportunity.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

For Peer Review



640 ■ Sources of behaviour ■ Intervention functions ■ Policy categories

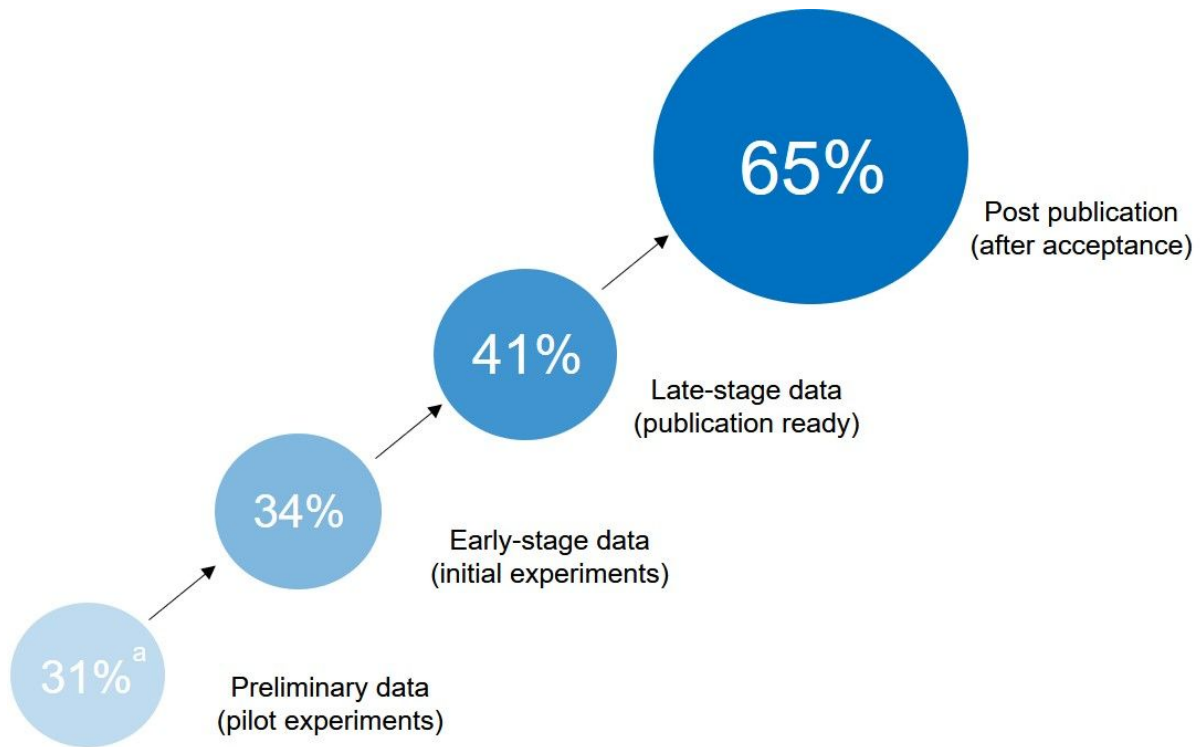
641 **Figure 1.** The Behaviour Change Wheel. Adapted from Michie et al. (2011)



642

643 **Figure 2.** Places where researchers typically source their data.

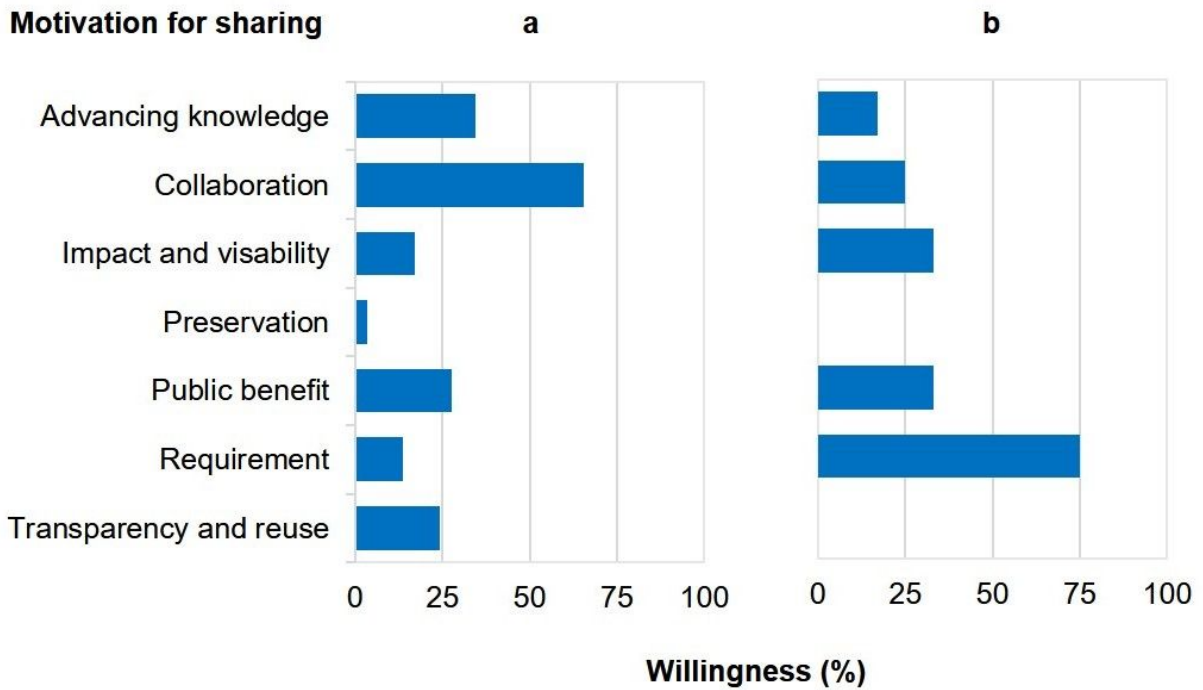
644 Note: Respondents could select multiple responses.



645

646 **Figure 3.** The proportion (%) of researchers who share their data external to the
647 research project, as distributed by research data timepoints.

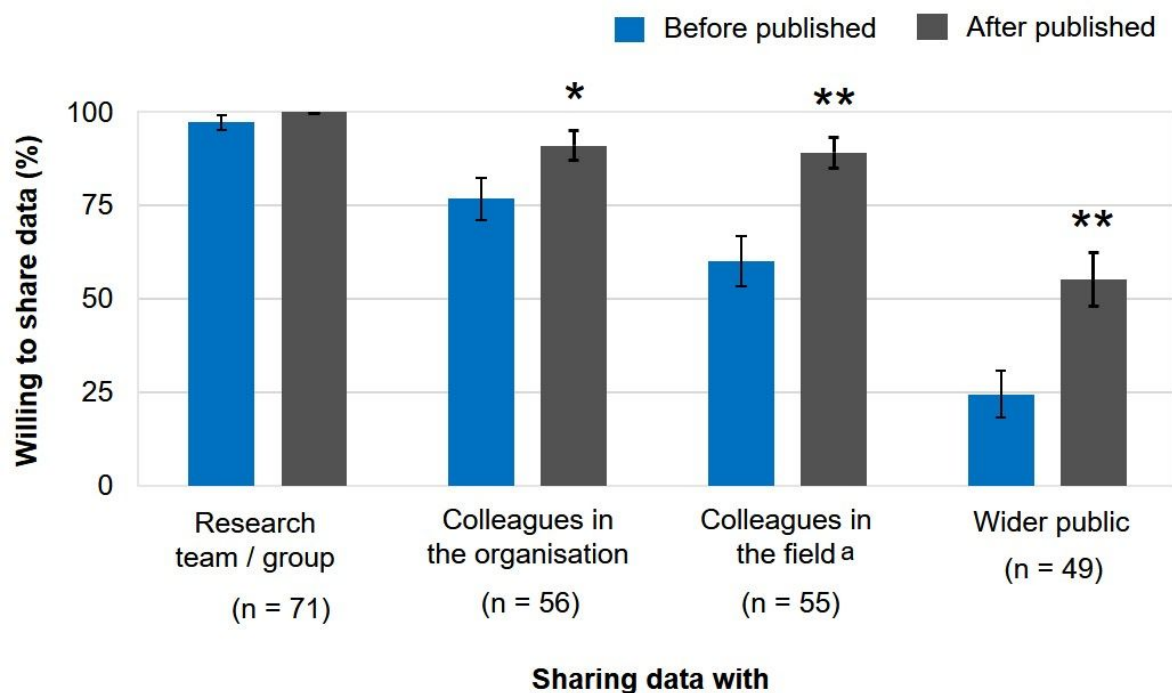
648 ^aData are calculated for researchers who had ever shared data external to the
649 research project (n = 29).



650

651 **Figure 4.** What motivates researchers to share their data with: (a) external

652 individuals to the research project and, (b) the wider public.



653

654 **Figure 5.** Researcher willingness (mean \pm SEM) to share research data.

655 ^aIncludes internal and external colleagues; n = represents the denominator excluding
 656 “neutral”; *correlation is significant at the 0.05 level (2-tailed); **correlation is
 657 significant at the 0.001 level (2-tailed).

Supplemental Material

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

Survey Question	Response Options
Q1. Have you ever shared research data outside of your project team?	Yes / No / Unsure
^Q1.1. If yes, what kind of data was it?	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
^Q1.2. If yes, why did you choose to share your research data?	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

Q2. Have you ever shared research data publicly?	Yes / No / Unsure
^Q2.1. If yes, where did you make your data publicly available?	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
^Q2.2. If yes, why did you choose to share your research data?	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
Q3. <u>Before</u> you publish your research, how willing are you to share your data with...? a) Members of your research team/group b) Colleagues at the university c) Colleagues in your field	5-point Likert scale: definitely willing; somewhat willing; neither willing nor unwilling; somewhat willing; not at all

d) The public at large	willing (dichotomised into two categories: willing; unwilling)
<p>Q4. <u>After</u> you publish your research, how willing are you to share your data with...?</p> <p>a) Members of your research team/group</p> <p>b) Colleagues at the university</p> <p>c) Colleagues in your field</p> <p>d) The public at large</p>	<p>5-point Likert scale: definitely willing; somewhat willing; neither willing nor unwilling; somewhat willing; not at all willing (dichotomised into two categories: willing; unwilling)</p>
<p>Q5. How much do you agree with the following statements...?</p> <p>a) I do not know where to share my data or who to share it with</p> <p>b) I do not know how to share my data</p> <p>c) I don't have the time or resources to prepare my data for sharing</p> <p>d) I don't know if it is my responsibility to share the data</p> <p>e) I am concerned about given proper credit for the use of the data</p> <p>f) I am concerned that my research will be stolen</p> <p>g) I am concerned about the misinterpretation or misuse of the data</p>	<p>5-point Likert scale: strongly agree; somewhat agree, neither agree or disagree; somewhat agree; strongly agree</p>

<p>h) I am concerned whether it is unethical to share the data</p> <p>i) I want to protect my intellectual property</p> <p>j) I want to protect the confidentiality of the data</p>	
<p>Q6. Would you be motivated to share your research data if...?</p> <p>a) It was a funding requirement</p> <p>b) It was an institutional requirement</p> <p>c) It was a journal requirement</p> <p>d) There was a public/patient benefit</p> <p>e) The impact and visibility of my research increased</p> <p>f) I was guaranteed proper credit for its use</p> <p>g) I had absolute trust in the person requesting the data</p> <p>h) I had a better understanding of the data sharing process.</p>	<p>Yes / No / Unsure</p>

^Multiple responses to question could be selected by respondent