**Investigating the barriers and enablers to data sharing behaviours: A qualitative Registered Report**

**Emma L. Henderson 1\*,** **Afrodita Marcu 2, Lou Atkins 3, Emily K. Farran 1**

1 School of Psychology, University of Surrey, Guildford, Surrey, United Kingdom

2 School of Health Sciences, University of Surrey, Guildford, Surrey, United Kingdom

3 Centre for Behaviour Change, University College London, Gower Street, London, United Kingdom

\* Corresponding author:

E-mail: e.l.henderson@surrey.ac.uk. ORCID: 0000-0002-5396-2321

**Abstract**

Data sharing describes the process of making research data available for reuse. The availability of research data is the basis of transparent, effective research systems that democratise access to knowledge and advance discovery. Despite a broad recognition of the value of data sharing across the sector, many researchers are not yet engaging meaningfully with data sharing behaviours. Through a behavioural lens, this study aims to identify the barriers and enablers to data sharing experienced by researchers working at a UK university. Data will be collected using a theoretically-informed 26-item interview schedule (COM-B model: Capability, Opportunity, Motivation – Behaviour; TDF: Theoretical Domains Framework). [SAMPLE SIZE TO BE ADDED AT STAGE 2] participants, across a range of career levels and disciplines, will be recruited to take part in semi-structured interviews focused on data sharing behaviours and their influences. Transcripts will be analysed using thematic template analysis based on the COM-B constructs and TDF domains. [A BRIEF SUMMARY OF THE RESULTS AND IMPLICATIONS WILL BE ADDED AT STAGE 2].

**Keywords**: Behaviour change, open research, open science, COM-B, TDF, open data, data sharing, data stewardship

Access to research data supports a central tenet of Open Research, that “Access to scientific knowledge should be as open as possible” (UNESCO, 2021, p. 36). Data availability enables the verification of past findings and accelerates the discovery of new findings through reanalysis and evidence synthesis (Fecher et al., 2015; Hardwicke et al., 2018; J. N. Towse, Ellis, et al., 2020). Accordingly, data availability is the basis of transparent, effective research systems that create credible conclusions, democratise access to knowledge, and underpin equitable innovation (Concordat on Open Research Data, 2016; G7 Science and Technology Ministers’ Communique, 2023; UNESCO, 2021)**.** The research community is increasingly recognising the value of data sharing in these pursuits. Most significantly, the recent UNESCO Recommendation on Open Science positions Open Research and data as a global research priority that can improve the reliability of evidence needed for decision making and policy (UNESCO, 2021). However, despite the role of shared data in addressing global environmental, economic and social issues (UNESCO, 2021), many researchers are not yet engaging meaningfully with such behaviours (see Gabelica et al., 2022; Hardwicke et al., 2018; J. N. Towse, Ellis, et al., 2020). The present qualitative research uses a behaviour change framework to determine the barriers and enablers that researchers experience when (considering) engaging with data sharing behaviours, with a view to informing the design of future interventions.

Though formal data sharing has existed for over 100 years (Branney et al., 2019; Karhulahti, 2022; Sieber, 2015), it was the digital age and electronic access to data that created the conditions to facilitate widespread sharing. The broad recognition of the value of data sharing has occurred simultaneously across the sector, with funders, journals, societies, universities, and researchers all advocating for data sharing and creating top-down initiatives. In the UK, the largest national funding agency, the UKRI (formerly RCUK), has had their Common Principles on Data policy since 2011 (UKRI, personal communication, 06 June, 2023). Likewise, the country’s largest charity funder, Wellcome, launched their policy in 2007, the current iteration of which actively encourages data management and sharing costs to be included in grant applications (Wellcome, 2017). A diverse group of stakeholders, including funders and publishers, developed the FAIR Data Principles (Findable, Accessible, Interoperable, and Reusable); a set of guidelines for enhancing the reusability of data (Wilkinson et al., 2016). Many publishers have their own data sharing policies (e.g., PLOS, 2014), and it is also a key component of the TOP Guidelines (Transparency and Openness Promotion; Nosek et al., 2015); a tool to support the implementation of Open Research practices at journal level. Simultaneously, universities and other organisations have institutional-level policies and are providing support for data storage through managed external data archives (e.g., UK Data Service), institutional data repositories, or general-purpose services (e.g., Zenodo).

Though researchers are only part of this wider data sharing ecosystem (Borgman & Bourne, 2022), ultimately, it is individual researchers who are responsible for the act of data sharing (Bezuidenhout & Chakauya, 2018)**.** Research consistently shows that overall researchers view data sharing as positive and important (e.g., Cheah et al., 2015; Farran et al., 2020; Fleming et al., 2022; The State of Open Data, 2022; Soeharjono & Roche, 2021; Van den Eynden et al., 2016), and that lack of access to data is an impediment to research progress (Tenopir et al., 2011). Measures of engagement show progress, as illustrated by the global 2022 The State of Open Data survey, in which 35% of respondents reported being familiar with FAIR principles[[1]](#footnote-2), up from 28% the previous year and the highest percentage since the question was first asked in 2018. Yet despite this positive momentum implementation is often low (e.g., Farran et al., 2020; Fleming et al., 2022; Rowhani-Farid & Barnett, 2016) and may fall short of accepted standards. For example, when authors of papers with data availability statements (indicating that data are available upon request) were asked to share their data, 93% failed to reply or declined to share their data, and only 6.8% shared the requested data (Gabelica et al., 2022). This attitude-behaviour gap raises questions about the barriers preventing researchers from sharing their research data (Fecher et al., 2015).

In the present research, we use UNESCO’s definition of open research data as

… data that include, among others, digital and analogue data, both raw and processed, and the accompanying metadata, as well as numerical scores, textual records, images and sounds, protocols, analysis code and workflows that can be openly used, reused, retained and redistributed by anyone, subject to acknowledgement. Open research data are available in a timely and user-friendly, human- and machine-readable and actionable format, in accordance with principles of good data governance and stewardship, notably the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, supported by regular curation and maintenance. (UNESCO, 2021, p. 9).

This broad definition allows us to consider data as the evidence that underlies research publications, and therefore applies across a range of academic disciplines. In the present research we use the term ‘data sharing’[[2]](#footnote-3) to also include data that by necessity (i.e., for various legal, ethical, or commercial reasons) are not openly available, but that are made accessible to specific users according to defined access criteria.

**Benefits of Data Sharing**

There are many potential benefits of data sharing at individual researcher (C. Allen & Mehler, 2019; McKiernan et al., 2016), research community (e.g., Milham et al., 2018), and societal levels (e.g., Besançon et al., 2021). Importantly, access to data leads to a more equitable distribution of opportunities and promotes inclusion (The State of Open Data, 2022; UNESCO, 2021). Re-use of data facilitates greater efficiency, effectiveness, and innovation by using the same resources multiple times to create new knowledge (Burgelman et al., 2019; DuBois et al., 2018) rather than duplicating research efforts. Increased transparency and more focus on reproducibility enables verification of findings, and re-analysis when improved methods are developed. Specifically at the researcher level, sharing data enhances the visibility of research, and this can lead to a citation advantage (Piwowar et al., 2007; Piwowar & Vision, 2013) and more opportunities to collaborate (McKiernan et al., 2016). However, many of these potential benefits are distal, compared to the more proximal challenges posed by sharing data, and the ever-present pressure to publish frequently and thereby increase the chances of employment, promotion, and funding (Munafò et al., 2017).

**Concerns about Data Sharing**

Debates about data sharing commonly focus on qualitative human data (Karhulahti, 2022), and point particularly to concerns over epistemology, informed consent, and privacy (e.g., Parry & Mauthner, 2004). Issues of epistemology relate to the reflexive, subjective, and contextually-bound nature of qualitative research that suggests that reuse could lead to misinterpretation (e.g., Broom et al., 2009). The key concerns raised about informed consent are whether researchers are less willing to be candid about sensitive topics (MacLean et al., 2019), and whether participants truly understand the implications of consent (Parry & Mauthner, 2004). Relatedly, concerns have been raised about ensuring anonymisation of qualitative data, particularly for sensitive data or small, potentially re-identifiable communities (Broom et al., 2009; Parry & Mauthner, 2004). However, it is possible for these issues to be overcome with careful planning and sufficient resources (see, Bishop, 2005; Branney et al., 2019, 2023; DuBois et al., 2018; Karhulahti, 2022 for proposed solutions). Furthermore, the majority of participants consent to share their deidentified data (e.g., Mozersky et al., 2020), even for research on sensitive topics such as abortion (VandeVusse et al., 2022) and GP-patient conversations (Amelung et al., 2020; discussed in Whitaker, 2021), citing helping others as their primary motivation (VandeVusse et al., 2022).

**Data Sharing as Behaviour**

The term ‘data sharing’ encompasses a range of behaviours that occur across the research lifecycle, taking place before (e.g., preparing consent forms), during (e.g., recording exclusions), and after the research (e.g., depositing the data in a repository). Behaviours do not occur in isolation but in systems of behaviours that interact with and depend on one another (Michie et al., 2014). This inter-dependence means that if one data sharing behaviour does not occur, this may ultimately prevent data from being shared (see, Norris & O’Connor, 2019). For example, omitting information about future data sharing from participant consent forms or failing to secure suitable funding for data archiving, may preclude the data from being shared.

For the purposes of the present research, we are interested in individual researchers’ data sharing behaviours. Here we provide a synthesised list of the key behaviours that comprise an idealised data sharing process at the individual researcher level[[3]](#footnote-4). Not all behaviours listed are required in order to meet the overarching behaviour of data sharing (e.g., ethics is not required for all research); essential behaviours are depicted using asterisks. The planned research will validate this list of behaviours and add to it if appropriate.

1. **Seek out skills and resources**: Seeking out and engaging with educational resources, and/or participating in training to learn about what constitutes ‘data’, the benefits of sharing, how to share data within ethical, IP, and commercial constraints, and how to handle sensitive data. Reading and complying with university and funder mandates. Seeking practical, financial, or motivational support from peers, colleagues, ethics committees, pre-bid teams, funders, and other facilitators. For example, applying for funding to support data preparation and storage.
2. **Create a Data Management Plan**: Creating a data management plan that outlines what types of data will be collected and how researchers will handle the data during and after the study. The plan should address all stages of the research lifecycle from planning through to sharing. Data management plans are required for some funding applications.
3. **Obtain ethics**: Submitting an ethics application that includes plans to share data and details of how this will be done. For example, anticipating terms of access.
4. **Precursor behaviours:** Carrying out data sharing precursor behaviours throughout study design and the active project phase. For example, preparing participant information sheets and consent forms to gain consent from participants to share their data, or acquiring agreement from other stakeholders to share the project data. Then during the active project phase, collecting and analysing data with reuse in mind.
5. **Prepare and manage data\***: Preparing data for sharing by following relevant standards (e.g., FAIR) and disciplinary norms to ensure that data will be findable, accessible, interoperable, and reusable. This behaviour includes storing, naming, and versioning the data in a format that can be shared, and creating documentation and metadata. For personal sensitive data this would include anonymising it (i.e., removing identifying information to protect participants identities), or for commercial data/IP protection this might include aggregation.
6. **Deposit data\***: Depositing the data and metadata in a repository and providing reuse guidance by adding a license. For sensitive data shielding may be required in the form of access control, that is, specifying the conditions under which the data can be accessed. The data may be placed under a reasonable embargo, for example to delay the release of the data to coincide with a publication, end of project, or to protect first use rights.

Ultimately, the aim of data sharing is to facilitate reusability and subsequent new knowledge. In order to enhance the value and reusability of data (A. S. Towse et al., 2021), it should comply with the FAIR Data Principles (Wilkinson et al., 2016). Therefore, the core data sharing steps – preparing (5) and depositing data (6) – should be carried out with reuse in mind. Ensuring that data are stored in a suitable permanent repository, with rich metadata, clearly labelled and described to ensure it can be independently understood, in a future-proof and ideally non-proprietary format, with a global persistent identifier and an appropriate, preferably open, license (e.g., CC BY). Without these provisions, data have limited reusability (J. N. Towse, Ellis, et al., 2020).

**Behaviour Change**

Whether researchers decide to adopt data sharing behaviours is a behavioural question (Bishop, 2020; Norris & O’Connor, 2019; Osborne & Norris, 2022), and behaviour change theory has the potential to help understand and improve adoption and maintenance of such behaviours (Norris & O’Connor, 2019). The present research has been developed using the COM-B model (Capability, Opportunity, Motivation – Behaviour) from the Behaviour Change Wheel (BCW, see Figure 1; Michie et al., 2011, 2014): The BCW is a layered framework designed to guide the development of theory-based behaviour change from analysis to intervention design (Michie et al., 2014). We selected this framework because it can be applied to behaviour across different fields and contexts, and was developed based on overcoming the limitations of 19 multidisciplinary frameworks (Michie et al., 2011). It has recently been applied in the domain of Open Research to develop interventions to increase the uptake of preregistration among researchers (Osborne & Norris, 2022) and to investigate the barriers and enablers to implementing the TOP Guidelines (Naaman et al., 2023).

**Figure 1**

*The Behaviour Change Wheel from Michie et al.* (2014)

Chart, sunburst chart

Description automatically generated

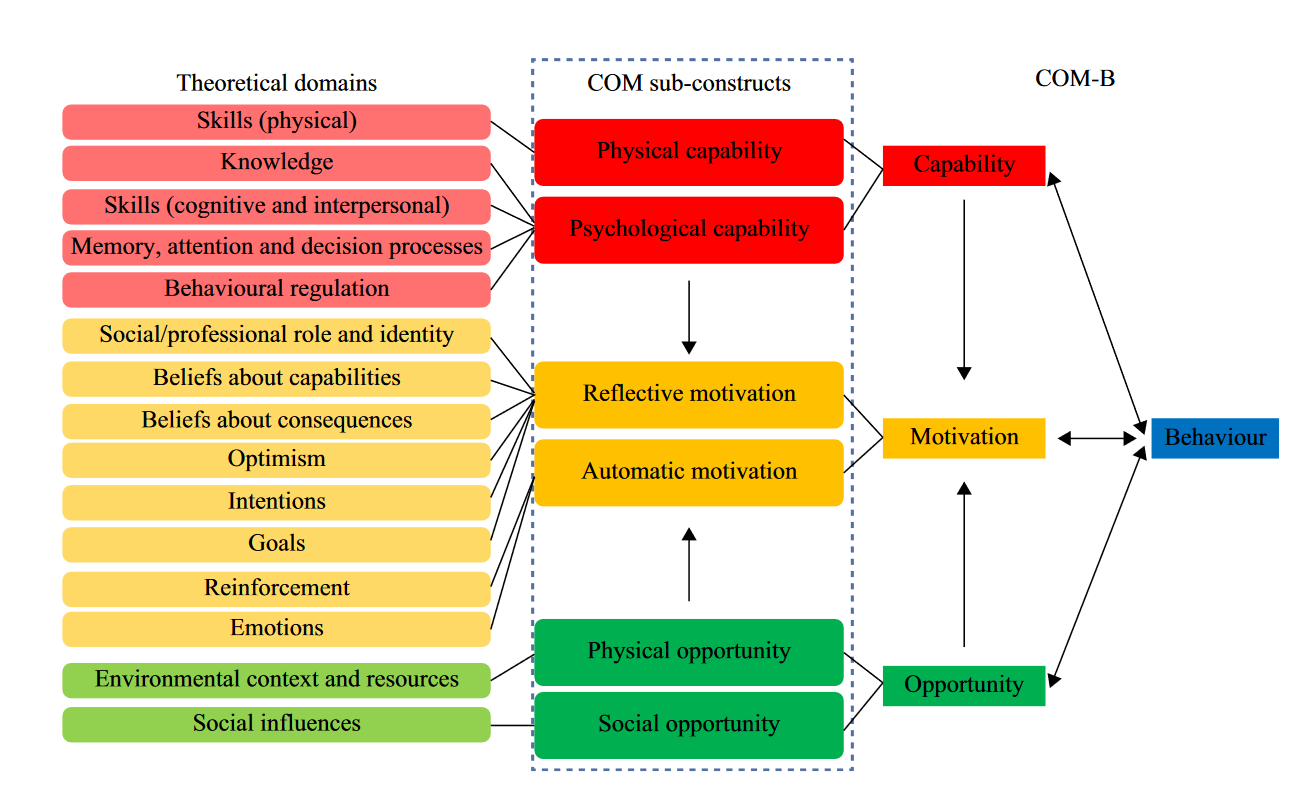
*Note.* The green ring shows influences on behaviour, while the red ring shows intervention types, and the grey ring represents policy options.

The COM-B model is at the centre of the BCW (the green ring) and is used to perform a behavioural diagnosis. This process involves identifying a target behaviour, investigating individual, socio-cultural and environmental influences (i.e., barriers that decrease the likelihood of the behaviour occurring and enablers that increase the likelihood), and assessing what needs to change in terms of capability, opportunity, and motivation. These three components are part of an interacting system and must be present in sufficient amount for the behaviour to occur: ‘capability’ is the individual’s physical and psychological ability to enact a behaviour, ‘opportunity’ refers to the physical and social environment that enables behaviour, and ‘motivation’ constitutes the reflective (i.e., rational choice) and automatic (i.e., feelings, habits) mechanisms that activate or inhibit behaviour (Michie et al., 2011, 2014). To change behaviour, one or more of the components must change in order to reconfigure the system. The choice of behaviour change intervention should be evidence-based, and informed by the factors that influence current behaviour, to develop something that might be most effective in the specific setting (Hulscher & Prins, 2017).

In addition to COM-B, the Theoretical Domains Framework (TDF; Atkins et al., 2017; Cane et al., 2012) is used in the current study for the development of the interview schedule and analysis. This validated integrative theoretical framework (Cane et al., 2012) comprises 14 domains (knowledge; skills; memory, attention and decision processes; behavioural regulation; social/professional role and identity; beliefs about capabilities; optimism; beliefs about consequences; intentions; goals; reinforcement; emotions; environmental context and resources; and social influences (Cane et al., 2012)) which map to the three COM-B components (see Figure 2) and can provide a granular understanding of behaviour (Michie et al., 2014).

**Figure 2**

*The Theoretical Domains Framework (TDF) mapped to the subconstructs of capability, opportunity, and motivation from COM-B (reproduced from Chater et al.* (2022)*)*



**Barriers and Enablers to Data Sharing**

Despite important reasons to share data, including individual career-based reasons (C. Allen & Mehler, 2019; Markowetz, 2015; McKiernan et al., 2016), many researchers do not share their data because of perceived costs (Abele-Brehm et al., 2019; Miyakawa, 2020) and lack of incentives (Adimoelja & Athreya, 2022; Chawinga & Zinn, 2019). With data sharing becoming an increasing priority across the sector, the determinants of researchers’ attitudes and behaviours to data sharing have received some scholarly interest. Existing research, spanning various disciplines and geographical areas, has largely focused on real and perceived barriers, and has used survey formats. Below we discuss current evidence categorised according to the three COM-B components.

**Barriers**

**Opportunity**: Lack of resources is regularly reported as a barrier to data sharing (Fecher et al., 2015). For example, in a survey of over 13,000 scientists conducted in 2009-2010, insufficient time and funding were the most frequently named barriers to data sharing, cited by 55% and 40% of respondents respectively (Tenopir et al., 2011). The fact that time is a frequently highlighted barrier (Astell et al., 2018; Chawinga & Zinn, 2019; Cheah et al., 2015; Farran et al., 2020; Houtkoop et al., 2018; Van den Eynden et al., 2016) is unsurprising as it is well acknowledged that academics have increasingly untenable workloads (Hostler, 2023; Long et al., 2020). Data sharing has the potential to increase research efficiency in the medium to long term at a systems level, but in the short term and at the individual level such behaviours increase workload and require more time and effort compared to “closed” research (Gomes et al., 2022; Hostler, 2023). Other opportunity-related barriers relate to physical resources: In low to middle income countries, lack of specialised data management expertise (Cheah et al., 2015) and infrastructure issues such as lack of current hardware, software and suitable internet access (Bezuidenhout & Chakauya, 2018) also pose a challenge.

**Capability**: Acknowledged barriers also include lack of knowledge and skills (Chawinga & Zinn, 2019), resulting in researchers not feeling fully equipped to complete data sharing tasks (Tenopir et al., 2015). Participants report that they have not learned how to share data (Houtkoop et al., 2018) and lack knowledge about how to share data in a useful way (Astell et al., 2018). The variety of available repositories and the lack of integration between them also poses a challenge in terms of selecting the most suitable storage (Astell et al., 2018). Researchers report a lack of knowledge about copyright, licensing (Astell et al., 2018; Farran et al., 2020), ethics and confidentiality issues that can affect data sharing (Gownaris et al., 2022).

**Motivation**: In a survey of 600 psychologists asked about 15 barriers, data sharing being uncommon in their field was selected as the most relevant reason for not sharing data (Houtkoop et al., 2018). Other studies show that researchers might not share data due to fear of the implications. For example, the possibility of compromising confidentiality and harming research participants if they can be identified, particularly for sensitive data or stigmatised communities (Cheah et al., 2015). Researchers are also concerned that their research reputation could be harmed (Cheah et al., 2015) if they are scooped (Bezuidenhout & Chakauya, 2018; Soeharjono & Roche, 2021) or if others who have insufficient information and context to understand the data misinterpret or misuse it (Bezuidenhout & Chakauya, 2018; Gomes et al., 2022; Sayogo & Pardo, 2013; Soeharjono & Roche, 2021; Tenopir et al., 2015; Van den Eynden et al., 2016), or even find errors in the data (Gomes et al., 2022). Furthermore, previous research has found that lack of credit and appropriate attribution when others re-use data is a barrier (Cheah et al., 2015; Farran et al., 2020; Gownaris et al., 2022).

**Enablers**

**Opportunity**: To be able to share data, researchers require opportunities including suitable infrastructure, technical, legal, financial and time allocation support from institutions and funders (European Commission, 2017). For example, the availability of a data repository has a significant influence on STEM researchers sharing data (Kim & Zhang, 2015), and Wellcome funded researchers cited funding to cover the costs of data preparation as their biggest motivator (Van den Eynden et al., 2016). Researchers who work solely on research, and do not have time consuming teaching obligations, are more likely to share their data (Tenopir et al., 2011). Similarly, researchers were more likely to share their data if minimal effort was required (Wallis et al., 2013). Opportunity also includes social opportunities such as institutions providing a positive research culture in which data sharing is recognised and rewarded (Huang et al., 2012).

**Capability**: Researchers must have the necessary skills to carry out the various sub-behaviours that comprise data sharing. This includes not just knowledge and skills about how to share data, but also planning during study design phases. Reanalysis of data from Tenopir et al. (2011) found that having data management skills increased the likelihood of data sharing (Sayogo & Pardo, 2013).

**Motivation**: Researchers who perceive career benefits to data sharing are more likely to have positive attitudes towards it and engage in more data sharing behaviours (Kim & Zhang, 2015). Direct personal benefits such as data sharing being looked upon favourably in funding and promotion decisions, and enhanced reputation are also motivating factors (Van den Eynden et al., 2016). In the aforementioned survey of psychologists, mandates to share data from funders or institutions were ranked top of the conditions most likely to encourage data sharing (Houtkoop et al., 2018). Increased impact, visibility and opportunities for collaboration are cited as incentives to share data (Farran et al., 2020; The State of Data, 2022; Van den Eynden & Bishop, 2014). When their data are re-used, researchers consider acknowledgement or citation to be essential (Sayogo & Pardo, 2013; The State of Data, 2022; Tenopir et al., 2015). But researchers also recognise broader incentives of public benefit and transparency and re-use (Farran et al., 2020).

**Research Questions**

The majority of research on factors influencing researchers’ data sharing behaviours is based on survey data and focuses on barriers; a more comprehensive and nuanced understanding is missing. For example, in survey responses, we cannot disentangle the often-cited barrier ‘lack of time’ from a lack of motivation to prioritise data sharing because it is not incentivised. Like other behaviours, data sharing is not stable within an individual, and may vary across time (Corker, 2018; Norris & O’Connor, 2019) based on internal factors such as motivation and habit, and external factors including resources and project priorities (Kwasnicka et al., 2016; Norris & O’Connor, 2019). Therefore, for those researchers who are currently engaging or have engaged with data sharing behaviours, we are interested in understanding what facilitated these behaviours, and what needs to change in the system to ensure maintenance and adoption by others.

Given the centrality of shared data in accelerating knowledge and solving global social issues (UNESCO, 2021), more thorough insight into the barriers and enablers to data sharing is important. Such an understanding can help facilitate the future development of effective behaviour change interventions. From this perspective, we are particularly interested in participants from one university because the insights from this study will be used by the university to develop future interventions to encourage data sharing. The overall aim of this study is to draw on the COM-B model and TDF to explore the factors that help and hinder researchers in sharing their research data. To do so, we conduct qualitative interviews analysed using thematic template analysis. Our research question is:

**RQ1**: What are the barriers and enablers to data sharing experienced by researchers?

The results will be presented in written format and synthesised visually in the form of a behavioural map that plots data sharing behaviours and their dependencies within the broader university system, and shows relationships between actors, behaviours, and influences (barriers and enablers).

**Method**

## Note: all OSF links are ‘view only’ provided for the purposes of Stage 1 peer review. At Stage 2, the links will be updated, and references added for the data ##

**Design**

The study consists of semi-structured qualitative interviews with researchers carried out during [enter date at Stage 2]. An interview design was selected to allow an in-depth exploration of the topic that extends beyond the strictures of quantitative surveys, and enables participants to talk about their individual experiences, and the barriers and enablers that are particularly pertinent for them. Interviews help to ensure that voices across different disciplines and career levels are given equal opportunity to be heard, and a semi-structured approach allows for prompts to help obtain further details. Furthermore, because Open Research-related terminology differs between disciplines, a one-to-one approach will minimise misunderstandings that might occur in a focus group setting or via a survey.

We supplemented the methodological details below by completing the consolidated criteria for reporting qualitative research (COREQ; Tong et al., 2007); a 32-item checklist for reporting key aspects of qualitative research (see “Materials & Procedures” component on the OSF <https://osf.io/w3sfq/?view_only=53487da8f8af4eb79a69784de9bc5c62>). We note that the COREQ is controversial, with criticisms including the inability to replicate the development of COREQ (Buus & Perron, 2020), a focus on data saturation (Braun & Clarke, 2021c) and on the interviewer gender over other relevant characteristics (Clarke, 2022). In the present study, the COREQ checklist did not guide our decisions, but provides a quick summary of the research. Additionally, we supplement interviewer characteristics by providing positionality statements (see “Positionality” component on the OSF https://osf.io/d4sjk/?view\_only=d2ada9f1d54141c28d3dd3714c86ea46).

The research received a favourable opinion from the first author's university's Research Ethics Committee (ref: FHMS 22-23 072 EGA).

**Recruitment and Participants**

Purposive sampling will be used to recruit research active staff and PhD students working at a university in the South of England. We deliberately recruit only researchers who are aware of or practice Open Research to ensure that participants can talk about their experiences of barriers and enablers to data sharing. Inclusion criteria: Researchers who produce potentially shareable data in their research or work in a team that does so, and self-report as one or more of the following: (a) They have shared data once or more, (b) They have experience using one or more of the following Open Research practices: open software/code, preregistration or Registered Reports, preprints, open monographs, open educational resources, or (c) They are aware of two or more of the aforementioned Open Research practices and have considered data sharing, but have not yet engaged with it.

Statistical generalisability is not the goal of qualitative research, rather we aim to provide rich knowledge that reveals the breadth of participant experiences (Smith, 2018). To maximise diversity within our target group, we will recruit participants to include a range of the following characteristics: career stages, genders, disciplines, and experience with data sharing (the latter being as per the inclusion criteria above). As a minimum, we will ensure that our final sample includes one female and one male participant from each of the four career stages (see Table 1), one participant from each of the three broad research discipline (STEM, Social Sciences, and Humanities), two participants from ethnic groups other than White British, and two participants who have not shared data.

The first round of recruitment was conducted prior to submitting the Stage 1 Registered Report because an apt opportunity occurred for people to express interest in taking part in the study: Initially potential participants were identified based on their contribution to a prior survey, led by the UK Reproducibility Network (UKRN), that ran in early 2023 and investigated attitudes towards and experience in Open Research. Upon completing the UKRN survey, if the potential participant was interested, they were directed to a short, separate sign-up survey where they were asked “How important do you believe Open Research is to your field?” and “Thinking about one of your recent research projects, did you/do you plan to make your research data open (i.e., information you collect, observe, generate or create as part of your research)?”. Twenty people indicated their interest in being interviewed (one of whom does not work with data and is therefore not eligible). If necessary, to ensure diversity on the characteristics mentioned above, we will recruit additional participants by advertising the study internally at the university via email (see “Materials & Procedures” component on the OSF <https://osf.io/w3sfq/?view_only=53487da8f8af4eb79a69784de9bc5c62>). This round of recruitment will be conducted post in-principle acceptance of the Stage 1 Registered Report. All potential participants (including those that have already shown interest) will complete a short screening survey to assess them against the inclusion criteria, and to collect demographic information relating to our characteristics of interest: career stage, gender, discipline, plus additional demographics: age and ethnicity (see “Materials & Procedures” component on the OSF <https://osf.io/w3sfq/?view_only=53487da8f8af4eb79a69784de9bc5c62>). Answers will be assessed against the inclusion criteria. If all criteria are met, participants will be invited for interview. To ensure that pseudonyms are allocated respectfully, the survey asks participants to provide their own pseudonym (R. E. S. Allen & Wiles, 2016).

Personal data from the recruitment and screening survey will be password protected and stored in a separate folder to the pseudonymised participant interviews.

**Sample Size Justification**

A priori, we set a minimum sample size and a maximum stopping rule. As described in the ‘data analysis’ section below, our use of template analysis sits on the spectrum between codebook and reflexive thematic analysis, and therefore data saturation is theoretically incoherent (see, Braun & Clarke, 2021c). Data saturation, originally coined by Glaser and Strauss (1967) as a specific element of grounded theory, is broadly defined as the point at which no additional themes, information or insights are yielded from the data (Braun & Clarke, 2021c). However, reflexive thematic analysis does not have a saturation point because it assumes that meaning occurs during the interaction between the data and the researcher, and therefore new insights are always possible (see Braun & Clarke, 2021c for a more nuanced discussion of data saturation in relation to thematic analysis). An alternative option suggested by Braun and Clarke (2021c), is information power (Malterud et al., 2015). We do not believe that there is an ideal single approach to determine qualitative sample sizes a priori (see Sim et al., 2018 for discussions against a priori estimates), but a combination of methods can provide a reasonable estimate. Therefore, we use three approaches as the rationale for our sample size: (a) information power; (b) previous research; and (c) resource constraints.

***Information Power***

Information power proposes that the more relevant information a sample holds, the fewer participants are required (Malterud et al., 2015). Five dimensions affect information power: 1) study aim – information power increases with a narrower research question and decreases with a broader question; 2) sample specificity – a sample comprising participants with characteristics and knowledge highly relevant to the research has high information power; 3) established theory – applying established theories increases information power; 4) quality of dialogue – if the data are rich fewer participants are required; 5) analysis strategy – single case or cross-case analysis decreases information power (Malterud et al., 2015). In summary, studies with focused research question(s), participants specific to the study aim, rich data that are supported by theory and analysed using in-depth exploration of narratives, have higher information power and require smaller samples (Malterud et al., 2015).

In this study we have dense sample specificity because participants are purposively recruited based on their knowledge and/or experience of data sharing, the semi-structured interview format should promote good quality of dialogue, we use established theory to design and interpret the study, and we do not use single case or cross-case analysis. However, our research question (RQ1) is neither broad nor narrow because although the topic - data sharing - is narrow, we ask it in the context of researchers across disciplines and career stages. Overall, information power considerations suggest a “smaller” sample size.

***Previous Research on Qualitative Sample Sizes***

Braun and Clarke (2013) typically recommend a sample size of 10 – 20 for a medium thematic analysis project using interviews. Notwithstanding our comments above about data saturation, we note that a recent systematic review of qualitative sample sizes found that on average 12-13 interviews reached saturation (Hennink & Kaiser, 2022), confirming previous work that also reported saturation at 12 interviews (Guest et al., 2006).

***Pragmatic Resource Constraints***

We also consider pragmatic constraints related to funding (limited internal funding) and time (the first author’s temporary contract and the time pressure that researchers, our participants, are under). Due to these resource constraints, we set the maximum number of interviews to 20.

***Sample Size***

Our aim is to capture the depth and nuances of the topic in relation to the research questions, while avoiding research waste in terms of funding and participant time. Based on the above three considerations, we set an anticipated lower sample size of 12 and an upper sample size of 20. The final sample size will be decided in-situ via discussion with the research team who will consider “the adequacy (richness, complexity) of the data for addressing the research question” (Braun & Clarke, 2021c, p. 211). An initial discussion (led by ELH) will occur after 12 interviews are completed. If further interviews are required, the research team will discuss sample size again after every two additional interviews until we reach the upper limit of 20. If we reach 20 interviews and have concerns about the adequacy of the data, we will note these as a limitation of the study.

Participant demographics of our final sample are presented in Table 1 [to be completed at Stage 2]. In total we interviewed XX participants aged XX – XX years. XX participants have previously shared their data, and XX had not. [a further summary of participants demographics may be added at Stage 2 if it is a useful addition to Table 1.]

**Table 1**

*Participant Demographics* [to be completed at Stage 2]

|  |  |  |
| --- | --- | --- |
| Characteristic | N | % |
| Gender |  |  |
| Female |  |  |
| Male |  |  |
| Non-binary |  |  |
| Prefer to self-describe |  |  |
| Prefer not to disclose |  |  |
| Ethnicity |  |  |
| [included groups to be completed at Stage 2] |  |  |
| Career stage |  |  |
| Junior (e.g., PhD candidate, Research Assistant) |  |  |
| Early (e.g., Research Associate or Fellow, Lecturer) |  |  |
| Mid (e.g., Senior Lecturer or Researcher, Reader) |  |  |
| Established (e.g., Professor, Principal Fellows) |  |  |
| Broad research discipline |  |  |
| STEM |  |  |
| Social Sciences |  |  |
| Humanities |  |  |
| Research methods used |  |  |
| Quantitative |  |  |
| Qualitative |  |  |
| Mixed |  |  |
| Number of Open Research practices used\* |  |  |
| 0 |  |  |
| 1 - 2 |  |  |
| 3 - 4 |  |  |
| 5 |  |  |

*Note*: \*The Open Research practices we consider relevant are open software/code, preregistration or Registered Reports, preprints, open monographs, open educational resources.

**Materials**

A 26-item interview schedule will be used to identify the barriers and enablers to data sharing behaviour (see Table 2). Interview questions were created informed by the COM-B model (Michie et al., 2011, 2014) and TDF (Atkins et al., 2017; Cane et al., 2012) and developed to extend previous work that suggests that opportunity-related factors such as time and resources, capability-related factors such as knowledge and skills, and motivation-related factors such as incentives, are barriers and enablers to data sharing. The schedule covers all COM-B constructs and TDF domains apart from ‘physical capability’ because we assume that if researchers are physically capable of conducting research, they are also capable of sharing data. The interview schedule was piloted in May 2023, with a participant who is familiar with Open Research practices. The questions were subsequently modified to ensure clarity. For details of how the interview will be introduced and closed, see “Materials & Procedures” component on the OSF <https://osf.io/w3sfq/?view_only=53487da8f8af4eb79a69784de9bc5c62>.

**Procedure**

One-to-one semi-structured interviews will be conducted by the first author (ELH) online via Teams. Participants will be provided with the information sheet and consent form (see “Materials & Procedures” component on the OSF) via email a minimum of 3 days before the interview. Participants will be advised that they may withdraw their data at any point and up to one month after interview completion without providing a reason. The information sheet explains that pseudonymised transcriptions of the interviews will be made openly available.

Interviews will last approximately one hour during which both participants and the interviewer will have their cameras on. At the start of the interview the researcher will explain the purpose of the research, mention that participants may ask for a break or withdraw at any time, and remind them that the interview is being recorded. Questions will be asked in the same fixed order for all participants (Table 2). If during the interviews it becomes apparent that participants understand a question differently from how it was

**Table 2**

*Interview Schedule Informed by COM-B and the TDF*

|  |  |  |
| --- | --- | --- |
| **COM-B construct** | **TDF domain** | **Interview Questions** |
|  |  | *Different disciplines use different terms to describe data, for example, some people conceptualise it as the evidence underpinning their conclusions. What does the term ‘data’ mean to you?* |
|  |  | *What kinds of data do you usually work with or produce in your research?* |
|  |  | *How would you describe your attitude towards data sharing? [prompt if necessary] Do you generally feel positive or negative? [follow up] Is it something you consider to be important?* |
|  |  | *To what extent is data sharing something you normally do with your research data? [follow up] Are you or have you been responsible for sharing the data yourself?* |
|  |  | *When did you start (thinking about) sharing your data and why? [prompt if necessary] For example, is it something you feel/don’t feel is important?* |
|  |  | *Have the reasons you share/don’t share your data changed?* |
| Reflective motivation | Social or professional role & identity | *Do you consider yourself to be an open researcher? [prompt if necessary] What does that term mean to you? [prompt if necessary] Does the term ‘open researcher’ have any positive or negative connotations?* |
| Reflective motivation | Social or professional role & identity | *Do you feel that sharing your data is part of your role? [follow up] Does it ever change, for example, based on the team or project you’re working on?* |
| Psychological capability | Knowledge | *The University and funders have policies on research data management, do you know what is expected from you in terms of data sharing by the University and funders? [follow up if yes/maybe] Can you summarise those expectations for me? [current behaviour] And how do your practices compare to those expectations - do you generally meet the requirements? [if behaviours differ] Can you tell me about why your practices differ?* |
|  |  | *Now I’m going to describe the various steps a researcher might follow when they’re sharing data. While I’m talking, think about whether there is anything missing from my list: For someone to share data really well, first you’d need to know why and how to share your data, you might need to seek or use practical or financial support, create a data management plan, include data sharing in your ethics application, and then you need to apply data sharing practices throughout planning, study design and data collection (for example, mentioning data sharing in your consent forms), and finally manage the data by anonymising it and saving it in a shareable file format, and then actually share the data in a repository according to disciplinary standards. Did I miss anything in terms of the behaviours that make up data sharing?* |
| Psychological capability | Knowledge | *Was there anything on that list that’s unfamiliar to you?* |
|  |  | *Do you follow all those steps?* |
| Psychological capability | Memory, attention & decision/Behavioural regulation | *When in the research process do/would you typically start to think about data sharing?* |
| Psychological capability/  Reflective motivation | Cognitive & interpersonal skills/ Belief about capabilities | *Do you feel that you have the necessary skills to complete all the data sharing practices I described? [follow up] Do you feel confident in those skills? [follow up] Are there any parts you find particularly difficult? [follow up] Have you ever received any training in data sharing?* |
| Automatic motivation | Emotion | *Are there any stages of data sharing that particularly worry you, or anything more broadly that concerns you about data sharing?* |
| Reflective motivation | Optimism | *To what extent do you think you can overcome these barriers or challenges?* |
| Reflective motivation | Beliefs about capabilities | *What support do you need to overcome those challenges? What do you think needs to change?* |
| Physical opportunity | Environmental context & resources | *How does your environment at work influence your data sharing behaviours? [follow up] Do you have sufficient resources in terms of things like time and infrastructure (e.g., a repository for your data), but also organisational culture, in order to complete all those data sharing practices? [follow up if time is mentioned] Where does/would the time come from that you spend on data sharing activities?* |
| Physical opportunity | Environmental context & resources | *How would your work environment, resources or environmental culture need to change to help you engage more with data sharing practices?* |
| Social opportunity | Social influences | *To what extent do the people you work with help or hinder you with data sharing? [prompt if necessary] What type of support is that?* |
| Social opportunity | Social influences | *How would the people around you need to change to help you engage more with data sharing practices?* |
| Physical opportunity | Environmental context & resources | *To what extent is data sharing prioritised and valued by your department/school, the university, and funders?* |
| Automatic motivation | Reinforcement | *Do you think that there are any incentives or benefits for you personally to share data? [follow up] Has data sharing ever affected your career?* |
| Reflective motivation | Goals/Intentions | *Where in your order of priorities does sharing your data come? [follow up] Would you like to change that? [follow up] What would need to change to make that shift?* |
| Reflective motivation | Beliefs about consequences | *Have you ever thought about who might use your data? [follow up] Do you know whether any of your data has been accessed or reused? [follow up] Would you like to know?* |
| Reflective motivation | Beliefs about consequences | *What do you think would happen if you were to always share your data? Would there be positive or negative implications for you, the wider research community, or society?* |

intended to be understood, we will amend the question and transparently report this. Upon completion of the interview, participants will be thanked and provided with a debrief (see “Materials & Procedures” component on the OSF <https://osf.io/w3sfq/?view_only=53487da8f8af4eb79a69784de9bc5c62>). They will be offered the opportunity to review their pseudonymised transcripts for the purpose of highlighting any parts that they do not wish to share. Any such parts will be reported transparently as ‘redacted’. Participants will receive a £50 Amazon voucher via email in return for participation.

**Data Analysis**

Interviews will be video recorded and transcribed using Otter.ai and Word’s automated audio transcription, and stored in the university's research folder. Transcriptions will be checked by the first author against the recordings to ensure verbatim accuracy of all verbal utterances, and that punctuation is used to preserve the original meaning (Braun & Clarke, 2006). The recordings will be deleted at the point that the Stage 2 manuscript is accepted.

***Pseudonymisation***

Pseudonymisation will be carried out by the first author and follow the UK Data Service’s guidance for qualitative data (UK Data Service, n.d.). We will use the following steps:

1. Where possible we will not collect disclosive data. For example, we will not ask for names of people, departments, universities, or companies.
2. We will use the UK Data Service’s Text Anonymisation Helper Tool (UK Data Service, n.d.) that runs MS Word macros to help find any disclosive information.
3. Pseudonymisation will occur once transcription is complete. The original, unedited version of the transcription will be kept for use within the research team.
4. We will replace any identifying information, rather than blanking it out. Replacements will be clearly indicated using square brackets. For example, “My colleague Indiana Jones” would be edited to “My colleague [name]”. We will keep a pseudonymisation log of any edits and an identifying key[[4]](#footnote-5), to be stored separately from the pseudonymised transcripts.

***Thematic Template Analysis***

Broadly, the purpose of thematic analysis is to develop themes in the dataset in relation to the research question (Braun & Clarke, n.d.). There are three main types of thematic analysis (Braun & Clarke, 2019, 2023) that appear on a spectrum from ‘coding reliability’ that prioritises coding accuracy, to ‘codebook’ where the coding structure is developed based on both the data and a priori theory, through to ‘reflexive’ that emphasises “the inescapable subjectivity of data interpretation” (Braun & Clarke, 2021a, p.37). Template analysis is a flexible form of thematic analysis that uses a hierarchical coding structure (Brooks et al., 2015). This style of thematic analysis was selected because it allows the theoretical underpinnings of the research, in this case COM-B and TDF, to be used to develop a priori themes but these themes remain flexible. The coding template is further developed based on a subset of the data, and then refined and advanced as it is applied to the full dataset (Brooks et al., 2015).

Where template analysis sits on the spectrum of thematic analysis depends on researchers’ epistemological position. As this research aims to explore what factors influence researchers' data sharing behaviours, we will adopt a critical realist ontology assuming that a meaningful reality exists but that our experience of it is subjective and socially influenced (Braun & Clarke, 2013). The analysis will also be underpinned by a contextualist epistemology: Contextualism aims to understand truth, but views knowledge as contextually located and influenced by the researcher’s position, and therefore truth is bound to the context in which data are collected and analysed (Madill et al. 2000). This position is consistent with a data-focused approach to thematic analysis that acknowledges the active role of the researcher (Brooks et al., 2015). From this philosophical position, template analysis sits on the spectrum between codebook and reflexive thematic analysis, and on the continuum between deductive (initial themes are established before coding) and inductive (themes are developed and refined through engagement with the data) thematic analysis.

The pseudonymised transcribed data will be coded using the software package NVivo [version added at Stage 2]. The template analysis will follow Brook et al.’s (2015) and King et al.’s (2018) guiding framework. Initially, template analysis is typically carried out on a subset of the data. The subset should capture the variety of experiences covered in the full dataset, therefore the precise number cannot be determined in advance but we anticipate that it will be a non-random sample of around five interviews. We use King et al.’s (2018)’s understanding of ‘codes’ as comments linked to extracts of text, indicating that they are relevant to the research question. Codes develop into themes; and ‘coding’ is the process of assigning codes and themes to the text.

Stages 1 – 4 below will be carried out independently by the first author (ELH), who is familiar with COM-B and the TDF, and the fourth author (EKF), whose familiarity with the models is limited to this study, and who therefore brings a different perspective to the analysis. Throughout the process, coding will be discussed with all authors for the purpose of developing a richer understanding of the data. The final coding template will be agreed by all authors.

1. **Familiarisation with the data**: Familiarisation is a key step because template analysis requires that extracts of text are interpreted in the context of their meaning within the participant’s complete account. The coders will become immersed in the data by listening to the interview recordings and reading the transcripts while looking for meaning and patterns. Informal notes will be made, for example, noting quirks and connections in the data and broadly what is going on in the data.
2. **Preliminary coding**: Preliminary coding will be carried out based on what appears interesting in the data in relation to RQ1. We will use a coding template of initial themes (version 1) developed a priori based on the COM-B constructs and TDF domains (see “Materials & Procedures” component on the OSF <https://osf.io/w3sfq/?view_only=53487da8f8af4eb79a69784de9bc5c62>). A priori themes may be removed or modified if they are ineffective at characterising the data.
3. **Clustering**: Once preliminary coding is completed, we will group the codes and a priori themes to begin to identify meaningful clusters. This process will involve sorting, collating, and combining similar codes into clusters of meaning to capture significant patterns in the dataset. The hierarchical structure of the themes will be developed, with narrower themes being nested under those that capture a broader perspective.
4. **Developing the coding template**: Having identified clusters, themes and their relationships, the coding template will be developed. Seeing as we started with an initial theoretically driven coding template, this will be version 2 of the template.
5. **Apply and modify the coding template**: The first author will develop the coding template by applying it to all remaining interviews and considering whether the themes capture the meaning of all interviews. Where themes do not capture the new data, the template will be modified. This phase might involve inserting new themes or moving them between clusters, defining, or discarding themes. A new version will not be produced after applying it to each interview; instead we will code three interviews before producing a new version. All versions of the coding template will be transparently reported along with a log of which interviews were coded to which version of the template (see “Materials & Procedures” component on the OSF <https://osf.io/w3sfq/?view_only=53487da8f8af4eb79a69784de9bc5c62>). The coding template will be modified until it captures all data relevant to RQ1. If changes between versions of the coding template are substantial, we will go back and recode the previously coded interviews (this decision will be taken in situ by the first author). Otherwise, we will recode interviews with the final coding template.
6. **Finalise the coding template**: We will consider the coding template to be ‘final’ when it, a) captures all data relevant to our research question, and b) is clear enough to facilitate interpretation of the findings. Once the template is finalised, it will be applied to the full dataset.
7. **Writing up**: Findings will be presented theme-by-theme. At the start of each thematic section, the theme and sub-themes will be presented in a table. The write up will focus on reporting the (sub)themes most relevant to the research question. Vivid examples of themes that capture their core meaning will be used to illustrate each of those themes.

***Credibility Strategies***

As described above, coding will be led by the first author (ELH), version 2 of the coding template will be developed by the first author and last author (EKF) whose familiarity with COM-B and the TDF are limited to this study, and the research team will input at various stages. We will not use consensus coding or inter-rater reliability because these methods are inconsistent with the philosophical assumptions that underly more reflexive or codebook types of thematic analysis (Braun & Clarke, 2021b). For example, inter-rater reliability assumes that there is a single accurate reality that should be coded in the data, where reflexive thematic analysis holds that the researcher is an active participant in meaning making, and that codes are derived via a situated interaction between the researcher and the data (Braun et al., 2019; Braun & Clarke, 2013). The researcher’s subjectivity is embraced as a resource that “sculpts the knowledge produced, rather than a must-be-contained threat to credibility” (Braun & Clarke, 2021b, p. 334). Explicating a researcher’s motives, background, and perspectives via a positionality statement allows the reader to consider the researcher’s influence on data collection and analysis, thereby increasing transparency and rigour (Steltenpohl et al., 2023). The authors have provided pre-study positionality statements (see “Positionality” component on the OSF <https://osf.io/d4sjk/?view_only=d2ada9f1d54141c28d3dd3714c86ea46>). A second positionality statement, in which the first author will reflect on how their assumptions and position might have shaped the coding process, will be completed once the data have been analysed and written up. In addition, to establish the rigor and dependability of the work, we will share the raw transcripts (subject to participant consent) and transparently document the research process by reporting all versions of the coding template (see Steltenpohl et al., 2023).

**Data availability**

The pseudonymised transcripts and study materials will be made available in the OSF repository <https://osf.io/> [to be completed at Stage 2]. All versions of the coding template will also be made available. To help maximise adherence to FAIR principles, the data will also be archived with the UK Data Service. For further details on how we ensure adherence to FAIR principles see the “Data” component on the OSF https://osf.io/ejcp5/?view\_only=5a52bec09805486b9af2e5a588263855.

**Results**

[to be completed at Stage 2]

**Discussion**

[to be completed at Stage 2]

**Acknowledgements**

We thank Emily McDougal for her valuable feedback on the draft interview schedule, and Alice Motes, Michelle Willows, and Robert Darby for their input on the list of data sharing behaviours.

**Conflict of interest**

ELH is a member of the Registered Reports Steering Committee supported by the Center for Open Science and works within the Open Research team at the University where this research is being conducted.

AM - None declared.

LA - None declared.

EKF is the UK Reproducibility Network (UKRN) Institutional Lead (a role that includes promoting Open Research) for the Institution where this research is being conducted .

**Ethics**

The research received a favourable opinion from the first author's university's Research Ethics Committee (ref: FHMS 22-23 072 EGA).

**Funding Statement**

ELH and funding for participant payments are supported by a Higher Education Innovation Funding (HEIF) grant to EKF, awarded via their institution.

**CRediT Statement** [to be reviewed and possibly revised at Stage 2]

**ELH**: Conceptualization (lead), Methodology (lead), Resources, Investigation, Data Curation, Formal Analysis, Visualisation, Writing – original draft, Writing – review & editing (lead), Project administration. **AM**: Conceptualization (supporting), Methodology (supporting), Writing – review & editing (supporting). **LA**: Methodology (supporting), Writing – review & editing (supporting), Validation. **EKF**: Conceptualization (supporting), Writing – review & editing (supporting), Formal Analysis, Funding acquisition.

**References**

Abele-Brehm, A. E., Gollwitzer, M., Steinberg, U., & Schönbrodt, F. D. (2019). Attitudes Toward Open Science and Public Data Sharing: A Survey Among Members of the German Psychological Society. *Social Psychology*, *50*(4), 252–260. https://doi.org/10.1027/1864-9335/a000384

Adimoelja, A., & Athreya, A. (2022). Reducing Barriers to Open Science by Standardizing Practices and Realigning Incentives. *Journal of Science Policy & Governance*, *21*(02). https://doi.org/10.38126/JSPG210201

Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLOS Biology*, *17*(5), e3000246. https://doi.org/10.1371/journal.pbio.3000246

Allen, R. E. S., & Wiles, J. L. (2016). A rose by any other name: Participants choosing research pseudonyms. *Qualitative Research in Psychology*, *13*(2), 149–165. https://doi.org/10.1080/14780887.2015.1133746

Amelung, D., Whitaker, K. L., Lennard, D., Ogden, M., Sheringham, J., Zhou, Y., Walter, F. M., Singh, H., Vincent, C., & Black, G. (2020). Influence of doctor-patient conversations on behaviours of patients presenting to primary care with new or persistent symptoms: A video observation study. *BMJ Quality & Safety*, *29*(3), 198–208. https://doi.org/10.1136/bmjqs-2019-009485

Astell, M., Hrynaszkiewicz, I., Allin, K., Penny, D., Mithu Lucraft, Baynes, G., & Springer Nature Admin. (2018). *Practical challenges for researchers in data sharing—Springer Nature survey data (anonymised)* (p. 655626 Bytes) [dataset]. figshare. https://doi.org/10.6084/M9.FIGSHARE.5971387

Atkins, L., Francis, J., Islam, R., O’Connor, D., Patey, A., Ivers, N., Foy, R., Duncan, E. M., Colquhoun, H., Grimshaw, J. M., Lawton, R., & Michie, S. (2017). A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implementation Science*, *12*(1), 77. https://doi.org/10.1186/s13012-017-0605-9

Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., Billy, E., Deforet, M., & Leyrat, C. (2021). Open science saves lives: Lessons from the COVID-19 pandemic. *BMC Medical Research Methodology*, *21*(1), 117. https://doi.org/10.1186/s12874-021-01304-y

Bezuidenhout, L., & Chakauya, E. (2018). Hidden concerns of sharing research data by low/middle-income country scientists. *Global Bioethics*, *29*(1), 39–54. https://doi.org/10.1080/11287462.2018.1441780

Bishop, L. (2005). Protecting Respondents and Enabling Data Sharing: Reply to Parry and Mauthner. *Sociology*, *39*(2), 333–336. https://doi.org/10.1177/0038038505050542

Bloom, T., Ganley, E., & Winker, M. (2014). Data Access for the Open Access Literature: PLOS’s Data Policy. *PLoS Biology*, *12*(2), e1001797. https://doi.org/10.1371/journal.pbio.1001797

Branney, P. E., Brooks, J., Kilby, L., Newman, K., Norris, E., Pownall, M., Talbot, C. V., Treharne, G. J., & Whitaker, C. M. (2023). Three steps to open science for qualitative research in psychology. *Social and Personality Psychology Compass*, *17*(4), e12728. https://doi.org/10.1111/spc3.12728

Branney, P. E., Reid, K., Frost, N., Coan, S., Mathieson, A., & Woolhouse, M. (2019). A context-consent meta-framework for designing open (qualitative) data studies. *Qualitative Research in Psychology*, *16*(3), 483–502. https://doi.org/10.1080/14780887.2019.1605477

Braun, V., & Clarke, V. (n.d.). *Https://www.thematicanalysis.net/*. Thematic Analysis. Retrieved 18 May 2023, from https://www.thematicanalysis.net/

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. SAGE.

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Braun, V., & Clarke, V. (2021a). Can I use TA? Should I use TA? Should I *not* use TA? Comparing reflexive thematic analysis and other pattern‐based qualitative analytic approaches. *Counselling and Psychotherapy Research*, *21*(1), 37–47. https://doi.org/10.1002/capr.12360

Braun, V., & Clarke, V. (2021b). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, *18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Braun, V., & Clarke, V. (2021c). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, *13*(2), 201–216. https://doi.org/10.1080/2159676X.2019.1704846

Braun, V., & Clarke, V. (2023). Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a *knowing* researcher. *International Journal of Transgender Health*, *24*(1), 1–6. https://doi.org/10.1080/26895269.2022.2129597

Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic Analysis. In P. Liamputtong (Ed.), *Handbook of Research Methods in Health Social Sciences* (pp. 843–860). Springer Singapore. https://doi.org/10.1007/978-981-10-5251-4\_103

Brooks, J., McCluskey, S., Turley, E., & King, N. (2015). The Utility of Template Analysis in Qualitative Psychology Research. *Qualitative Research in Psychology*, *12*(2), 202–222. https://doi.org/10.1080/14780887.2014.955224

Broom, A., Cheshire, L., & Emmison, M. (2009). Qualitative Researchers’ Understandings of Their Practice and the Implications for Data Archiving and Sharing. *Sociology*, *43*(6), 1163–1180. https://doi.org/10.1177/0038038509345704

Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., & Schouppe, M. (2019). Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. *Frontiers in Big Data*, *2*, 43. https://doi.org/10.3389/fdata.2019.00043

Buus, N., & Perron, A. (2020). The quality of quality criteria: Replicating the development of the Consolidated Criteria for Reporting Qualitative Research (COREQ). *International Journal of Nursing Studies*, *102*, 103452. https://doi.org/10.1016/j.ijnurstu.2019.103452

Cane, J., O’Connor, D., & Michie, S. (2012). Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implementation Science*, *7*(1), 37. https://doi.org/10.1186/1748-5908-7-37

Chater, A. M., Family, H., Abraao, L. M., Burnett, E., Castro-Sanchez, E., Du Toit, B., Gallagher, R., Gotterson, F., Manias, E., McEwen, J., Moralez De Figueiredo, R., Nathan, M., Ness, V., Olans, R., Padoveze, M. C., & Courtenay, M. (2022). Influences on nurses’ engagement in antimicrobial stewardship behaviours: A multi-country survey using the Theoretical Domains Framework. *Journal of Hospital Infection*, *129*, 171–180. https://doi.org/10.1016/j.jhin.2022.07.010

Chawinga, W. D., & Zinn, S. (2019). Global perspectives of research data sharing: A systematic literature review. *Library & Information Science Research*, *41*(2), 109–122. https://doi.org/10.1016/j.lisr.2019.04.004

Cheah, P. Y., Tangseefa, D., Somsaman, A., Chunsuttiwat, T., Nosten, F., Day, N. P. J., Bull, S., & Parker, M. (2015). Perceived Benefits, Harms, and Views About How to Share Data Responsibly: A Qualitative Study of Experiences With and Attitudes Toward Data Sharing Among Research Staff and Community Representatives in Thailand. *Journal of Empirical Research on Human Research Ethics*, *10*(3), 278–289. https://doi.org/10.1177/1556264615592388

Clarke, V. (2022). *‘1/ A thread on why I find the COREQ 32-item checklist reporting criteria for qualitative research so problematic. You can read the paper outlining the consolidated criteria for reporting qualitative research here: Https://t.co/sOA5rFlwN2’ / X*. Twitter. https://twitter.com/drvicclarke/status/1497213812545671170?s=20

Concordat on Open Research Data. (2016). *Concordat on Open Research Data*.

Corker, K. (2018, September 12). Open Science is a Behavior. *Open Science Is a Behavior*. https://www.cos.io/blog/open-science-is-a-behavior

DuBois, J. M., Strait, M., & Walsh, H. (2018). Is it time to share qualitative research data? *Qualitative Psychology*, *5*(3), 380–393. https://doi.org/10.1037/qup0000076

European Commission. (2017). *Providing researchers with the skills and competencies they need to practise Open Science.* Publications Office. https://data.europa.eu/doi/10.2777/121253

Farran, E. K., Jerrom, L., Hamilton, A., & Daoutis, C. (2020). *Open Research Practice Questionnaire Report 2020; University of Surrey*. PsyArXiv. https://doi.org/10.31234/osf.io/5nv4s

Fecher, B., Friesike, S., & Hebing, M. (2015). What Drives Academic Data Sharing? *PLOS ONE*. https://doi.org/DOI:10.1371/journal.pone.0118053

Fleming, J. I., Wilson, S. E., Espinas, D., van Dijk, W., & Cook, B. G. (2022). *Special Education Researchers’ Knowledge, Attitudes, and Reported Use of Open Science Practices* [Preprint]. EdArXiv. https://doi.org/10.35542/osf.io/tyc36

G7 Science and Technology Ministers’ Communique. (2023). *G7 Science and Technology Ministers’ Communique*. https://www8.cao.go.jp/cstp/kokusaiteki/g7\_2023/230513\_g7\_communique.pdf

Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: A mixed-methods study. *Journal of Clinical Epidemiology*, *150*, 33–41. https://doi.org/10.1016/j.jclinepi.2022.05.019

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine.

Gomes, D. G. E., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., Turba, R., Martinez, P. A., Moreau, D., Bertram, M. G., Smout, C. A., & Gaynor, K. M. (2022). *Why don’t we share data and code? Perceived barriers and benefits to public archiving practices*. 11. https://doi.org/doi.org/10.1098/rspb.2022.1113

Gownaris, N. J., Vermeir, K., Bittner, M.-I., Gunawardena, L., Kaur-Ghumaan, S., Lepenies, R., Ntsefong, G. N., & Zakari, I. S. (2022). Barriers to Full Participation in the Open Science Life Cycle among Early Career Researchers. *Data Science Journal*, *21*(1), 2. https://doi.org/10.5334/dsj-2022-002

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, C., Kidwell, M. C., Mohr, A. H., Clayton, E., Yoon, E. J., Henry, M., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*. https://doi.org/doi.org/10.1098/rsos.180448

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, *292*, 114523. https://doi.org/10.1016/j.socscimed.2021.114523

Hostler, T. (2023). *The Invisible Workload of Open Research* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/8xd65

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., & Wagenmakers, E.-J. (2018). Data Sharing in Psychology: A Survey on Barriers and Preconditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 70–85. https://doi.org/10.1177/2515245917751886

Huang, X., Hawkins, B. A., Lei, F., Miller, G. L., Favret, C., Zhang, R., & Qiao, G. (2012). Willing or unwilling to share primary biodiversity data: Results and implications of an international survey: Biodiversity data sharing and archiving. *Conservation Letters*, *5*(5), 399–406. https://doi.org/10.1111/j.1755-263X.2012.00259.x

Hulscher, M. E. J. L., & Prins, J. M. (2017). Antibiotic stewardship: Does it work in hospital practice? A review of the evidence base. *Clinical Microbiology and Infection*, *23*(11), 799–805. https://doi.org/10.1016/j.cmi.2017.07.017

Karhulahti, V. (2022). Reasons for qualitative psychologists to share human data. *British Journal of Social Psychology*, bjso.12573. https://doi.org/10.1111/bjso.12573

Kim, Y., & Zhang, P. (2015). Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Library & Information Science Research*, *37*(3), 189–200. https://doi.org/10.1016/j.lisr.2015.04.006

King, N., Brooks, J., & Tabari, S. (2018). Template Analysis in Business and Management Research. In M. Ciesielska & D. Jemielniak (Eds.), *Qualitative Methodologies in Organization Studies* (pp. 179–206). Springer International Publishing. https://doi.org/10.1007/978-3-319-65442-3

Kwasnicka, D., Dombrowski, S. U., White, M., & Sniehotta, F. (2016). Theoretical explanations for maintenance of behaviour change: A systematic review of behaviour theories. *Health Psychology Review*, *10*(3), 277–296. https://doi.org/10.1080/17437199.2016.1151372

Long, D. W., Barnes, A. P. L., Northcote, P. M., & Williams, P. T. (2020). Accounting Academic Workloads: Balancing Workload Creep to Avoid Depreciation in the Higher Education Sector. *Education, Society and Human Studies*, *1*(2), p55. https://doi.org/10.22158/eshs.v1n2p55

MacLean, L. M., Posner, E., Thomson, S., & Wood, E. J. (2019). Research Ethics and Human Subjects: A Reflexive Openness Approach. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3332887

Malterud, K., Siersma, V. D., & Guassora, A. D. (2015). Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qualitative Health Research*. https://doi.org/10.1177/1049732315617444

Markowetz, F. (2015). Five selfish reasons to work reproducibly. *Genome Biology*, *16*(1), 274. https://doi.org/10.1186/s13059-015-0850-7

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., & Yarkoni, T. (2016). How open science helps researchers succeed. *ELife*, *5*, e16800. https://doi.org/10.7554/eLife.16800

Michie, S., Atkins, L., & West, R. (2014). The behaviour change wheel: A guide to designing interventions. 2014. *Great Britain: Silverback Publishing*.

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, *6*(1), 42. https://doi.org/10.1186/1748-5908-6-42

Milham, M. P., Ai, L., Koo, B., Xu, T., Amiez, C., Balezeau, F., Baxter, M. G., Blezer, E. L. A., Brochier, T., Chen, A., Croxson, P. L., Damatac, C. G., Dehaene, S., Everling, S., Fair, D. A., Fleysher, L., Freiwald, W., Froudist-Walsh, S., Griffiths, T. D., … Schroeder, C. E. (2018). An Open Resource for Non-human Primate Imaging. *Neuron*, *100*(1), 61-74.e2. https://doi.org/10.1016/j.neuron.2018.08.039

Miyakawa, T. (2020). No raw data, no science: Another possible source of the reproducibility crisis. *Molecular Brain*, *13*(1), 24, s13041-020-0552–2. https://doi.org/10.1186/s13041-020-0552-2

Mozersky, J., Parsons, M., Walsh, H., Baldwin, K., McIntosh, T., & DuBois, J. M. (2020). Research Participant Views regarding Qualitative Data Sharing. *Ethics & Human Research*, *42*(2), 13–27. https://doi.org/10.1002/eahr.500044

Naaman, K., Grant, S., Kianersi, S., Supplee, L., Henschel, B., & Mayo-Wilson, E. (2023). Exploring enablers and barriers to implementing the Transparency and Openness Promotion Guidelines: A theory-based survey of journal editors. *Royal Society Open Science*, *10*(2), 221093. https://doi.org/10.1098/rsos.221093

Norris, E., & O’Connor, D. B. (2019). Science as behaviour: Using a behaviour change approach to increase uptake of open science. *Psychology & Health*, *34*(12), 1397–1406. https://doi.org/10.1080/08870446.2019.1679373

Osborne, C., & Norris, E. (2022). Pre-registration as behaviour: Developing an evidence-based intervention specification to increase pre-registration uptake by researchers using the Behaviour Change Wheel. *Cogent Psychology*, *9*(1), 2066304. https://doi.org/10.1080/23311908.2022.2066304

Parry, O., & Mauthner, N. S. (2004). Whose Data are They Anyway?: Practical, Legal and Ethical Issues in Archiving Qualitative Research Data. *Sociology*, *38*(1), 139–152. https://doi.org/10.1177/0038038504039366

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE*, *2*(3), e308. https://doi.org/10.1371/journal.pone.0000308

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175. https://doi.org/10.7717/peerj.175

Research, N., & Goodey, G. (2022). *State of Open Data Survey 2022 additional resources* (Version 2) [dataset]. figshare. https://doi.org/10.6084/m9.figshare.21295422.v2

Rowhani-Farid, A., & Barnett, A. G. (2016). Has open data arrived at the *British Medical Journal (BMJ)* ? An observational study. *BMJ Open*, *6*(10), e011784. https://doi.org/10.1136/bmjopen-2016-011784

Sayogo, D. S., & Pardo, T. A. (2013). Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*, *30*, S19–S31. https://doi.org/10.1016/j.giq.2012.06.011

Science, Digital, Goodey, Gregory, Hahnel, Mark, Zhou, Yuanchun, Jiang, Lulu, Chandramouliswaran, Ishwar, Hafez, Amy, Paine, Taunton, Gregurick, Susan, Simango, Samuel, Palma Peña, Juan Miguel, Murray, Holly, Cannon, Matt, Grant, Rebecca, McKellar, Kate, & Day, Laura. (2022). *The State of Open Data 2022* (p. 6824477 Bytes). Digital Science. https://doi.org/10.6084/M9.FIGSHARE.21276984.V5

Sieber, J. (2015). Data Sharing in Historical Perspective. *Journal of Empirical Research on Human Research Ethics*, *10*(3), 215–216. https://doi.org/10.1177/1556264615594607

Sim, J., Saunders, B., Waterfield, J., & Kingstone, T. (2018). Can sample size in qualitative research be determined a priori? *International Journal of Social Research Methodology*, *21*(5), 619–634. https://doi.org/10.1080/13645579.2018.1454643

Smith, B. (2018). Generalizability in qualitative research: Misunderstandings, opportunities and recommendations for the sport and exercise sciences. *Qualitative Research in Sport, Exercise and Health*, *10*(1), 137–149. https://doi.org/10.1080/2159676X.2017.1393221

Soeharjono, S., & Roche, D. G. (2021). Reported Individual Costs and Benefits of Sharing Open Data among Canadian Academic Faculty in Ecology and Evolution. *BioScience*, *71*(7), 750–756. https://doi.org/10.1093/biosci/biab024

Steltenpohl, C. N., Lustick, H., Meyer, M. S., Lee, L. E., Stegenga, S. M., Standiford Reyes, L., & Renbarger, R. L. (2023). Rethinking Transparency and Rigor from a Qualitative Open Science Perspective. *Journal of Trial and Error*. https://doi.org/10.36850/mr7

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, *6*(6), e21101. https://doi.org/10.1371/journal.pone.0021101

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE*, *10*(8), e0134826. https://doi.org/10.1371/journal.pone.0134826

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, *19*(6), 349–357. https://doi.org/10.1093/intqhc/mzm042

Towse, A. S., Ellis, D. A., & Towse, J. N. (2021). Making data meaningful: Guidelines for good quality open data. *The Journal of Social Psychology*, *161*(4), 395–402. https://doi.org/10.1080/00224545.2021.1938811

Towse, J. N., Ellis, D. A., & Towse, A. S. (2020). Opening Pandora’s Box: Peeking inside Psychology’s data sharing practices, and seven recommendations for change. *Behavior Research Methods*, *53*(4), 1455–1468. https://doi.org/10.3758/s13428-020-01486-1

Towse, J. N., Rumsey, S., Owen, N., Langford, P., Jaquiery, M., & Bolibaugh, C. (2020). *Data Sharing: A Primer from UKRN* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/wp4zu

UK Data Service. (n.d.). *Anonymising qualitative data*. https://ukdataservice.ac.uk/learning-hub/research-data-management/anonymisation/anonymising-qualitative-data/

UNESCO. (2021). *UNESCO Recommendation on Open Science*. UNESCO. https://doi.org/10.54677/MNMH8546

Van den Eynden, V., & Bishop, L. (2014). *Sowing the seed: Incentives and motivations for sharing research data, a researcher’s perspective*. Knowledge Exchange. https://www.knowledge-exchange.info/event/sowing-the-seed

Van den Eynden, V., Knight, G., Vlad, A., Radler, B., Tenopir, C., Leon, D., Manista, F., Whitworth, J., & Corti, L. (2016). *Survey of Wellcome researchers and their attitudes to open research*. 1843500 Bytes. https://doi.org/10.6084/M9.FIGSHARE.4055448

VandeVusse, A., Mueller, J., & Karcher, S. (2022). Qualitative Data Sharing: Participant Understanding, Motivation, and Consent. *Qualitative Health Research*, *32*(1), 182–191. https://doi.org/10.1177/10497323211054058

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS ONE*, *8*(7), e67332. https://doi.org/10.1371/journal.pone.0067332

Wellcome. (2017). *Wellcome data, software and materials management and sharing policy*. https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy

Whitaker, K. (2021, March 1). *Maximising the impact of patient-doctor video observation data*. University of Surrey. https://www.surrey.ac.uk/news/maximising-impact-patient-doctor-video-observation-data

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

1. The percentage was calculated using the raw survey data shared with the report (Research & Goodey, 2022). [↑](#footnote-ref-2)
2. The terms ‘data sharing’ and ‘data stewardship’ have been used somewhat interchangeably. In the present research we use the term data sharing because our starting point is the sharing of data and then expanding out to what is required to share data well (see our six-point list of behaviours below). This differs from (though overlaps with) data stewardship which comes from a data governance framework. Data stewardship captures a broad set of activities, one of which is data sharing, but focuses on long term management, storage, and preservation, and may be carried out by a data steward. [↑](#footnote-ref-3)
3. The list was informed by J. N. Towse, Rumsey, et al. (2020) and evidence cited in the introduction, and has been reviewed by three data sharing experts. [↑](#footnote-ref-4)
4. We plan to pseudonymise the data and keep an identifying key to allow the possibility of future follow up interviews with the participants. [↑](#footnote-ref-5)