

Researchers' perceptions of the trustworthiness, for reuse purposes, of government health data in Victoria, Australia: Implications for policy and practice

Health Information Management Journal
2025, Vol. 54(2) 139–149
© The Author(s) 2024



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/18333583241256049
journals.sagepub.com/home/himj



Merilyn Riley, *BAppSc(MRA), GDipEpidBiostats*¹ ,
Monique F Kilkenny, *BAppSc(MRA), GDipEpidBiostats, MPH, PhD*^{2,3} ,
Kerin Robinson, *BHA, BAppSc(MRA), MHP, PhD, CHIM, PhD*¹ ,
Sandra G Leggat, *PhD*^{1,4} 

Abstract

In 2022 the Australian Data Availability and Transparency Act (DATA) commenced, enabling accredited “data users” to access data from “accredited data service providers.” However, the DATA Scheme lacks guidance on “trustworthiness” of the data to be utilised for reuse purposes. **Objectives:** To determine: (i) Do researchers using government health datasets trust the data? (ii) What factors influence their perceptions of data trustworthiness? and (iii) What are the implications for government and data custodians? **Method:** Authors of published studies (2008–2020) that utilised Victorian government health datasets were surveyed via a case study approach. Twenty-eight trust constructs (identified via literature review) were grouped into data factors, management properties and provider factors. **Results:** Fifty experienced health researchers responded. Most (88%) believed that Victorian government health data were trustworthy. When *grouped*, data factors and management properties were more important than data provider factors in building trust. The most important *individual* trust constructs were: “compliant with ethical regulation” (100%) and “monitoring privacy and confidentiality” (98%). Constructs of least importance were knowledge of “participant consent” (56%) and “major focus of the data provider was research” (50%). **Conclusion:** Overall, the researchers trusted government health data, but data factors and data management properties were more important than data provider factors in building trust. **Implications:** Government should ensure the DATA Scheme incorporates mechanisms to validate those data utilised by accredited data users and data providers have sufficient quality (intrinsic and extrinsic) to meet the requirements of “trustworthiness,” and that evidentiary documentation is provided to support these “accredited data.”

Keywords (MeSH)

data sharing; health information management; data analysis; data quality

Supplementary keywords

Data trust; government; data curation; data reuse; secondary data

In the last two decades there has been a global increase in the sharing and reuse of data arising from policies for open data (Australian Government Department of Finance, 2022), advances in technology (Sielemann et al., 2020), and research funding and publication policies linked to data sharing (Sheehan, 2016). The potential advantages of data sharing and data reuse have been well-documented in the literature. These include (but are not limited to): potential increased cost-efficiencies through reduction in both duplication of and time spent on primary data collection (Curry, 2016; Whitlock et al., 2016); potential enrichment of research through new insights (Curry, 2016); increased transparency in data production and use (Roche et al.,

¹La Trobe University, Australia

²Monash University, Australia

³University of Melbourne, Australia

⁴James Cook University, Australia

Monique F Kilkenny, is now affiliated to Florey Institute of Neuroscience and Mental Health, Melbourne, Australia.

Accepted for publication May 4, 2024

Corresponding author:

Merilyn Riley, Department of Public Health, School of Psychology and Public Health, La Trobe University, Building Health Sciences 2, Kingsbury Drive, Bundoora, VIC 3083, Australia.
Email: merilyn.riley@latrobe.edu.au

2014); and an enhanced analytical capacity to inform decision-making (Verhulst, 2021).

Governments have been keen to adopt open government data (OGD) models (Umbaldi, 2013) with intent to enhance evidence-based decision making through responsible and secure sharing of data. There is widespread recognition among OGD advocates that “good information management is essential to building *trust* in the creation, collection and use of . . . Government information” (National Archives of Australia [NAA], n.d.).

Trust has been defined in many ways (Walterbusch et al., 2014) and in many different contexts (Huang, 2007; Sztompka, 2000). It is one of the most researched factors within the knowledge management and knowledge sharing domains (Asrar-ul-Haq and Anwar, 2016). Scholars have identified the complex and dynamic nature of trust in relation to data use and reuse (Sexton et al., 2017; Yoon, 2017; Yoon and Lee, 2019). For example, in their research into “factors of trust in data reuse” in the United States, Yoon and Lee (2019: 1248) identified several trust factors including “data, data producer, community and intermediary.” Steedman et al. (2020: 817, 829), in their research into trust in “data-driven practices,” postulated the multi-factorial nature of trust in data that either enhanced or undermined a user’s trust in situations which were “context dependent.” There is agreement among many scholars that data trust development is “situation specific” (Bauer and Freitag, 2016: 2).

The Australian Government promotes trust in its data production and data management abilities through a range of policies and organisational processes. In 2009, the Australian Bureau of Statistics (ABS) released its data quality framework enabling data users to self-assess the quality of government data and information outputs. In 2010, heads of Australian Government agencies agreed to principles for integrating government data for research purposes, with accredited agencies (“Integrating Authorities”) to be utilised for potentially “high risk” (sensitive) data integration (Greenall-Ota et al., 2021). In 2021, the Australian Government implemented the policy, *Building trust in the public record: Managing information and data for government and community* (NAA, n.d.). It aimed to foster community trust by creating transparency and accountability in organisational data and information management. However, provision of policies and procedures for effective data governance and data quality management does not presuppose their implementation. In 2020, the average information management maturity score for Australian Government Departments was 3.25/5 (NAA, 2021), indicating the need for considerable improvement in public sector organisations’ data and information management practices.

A recent component of the Australian Government’s digital data strategy was the enactment of the national Data Availability and Transparency Act 2022 (Cwth) (DATA Scheme). The main aim of the Act is to provide best practice guidelines to facilitate the sharing of Australian Government data safely and securely. The three defined “public interest” purposes for the sharing of government

data include: “government service delivery; informing government policy and programs; and research and development” (Office of the National Data Commissioner [ONDC], 2023a: 2). Data custodians can provide their data to “accredited data service providers,” from whom accredited “data users” can access data (ONDC, 2023a). This structure places the onus on research organisations and their employees for responsible and ethical use and management of data. Guidance is provided by the ONDC (2023b) on how to prepare and describe a dataset to ensure its system integrity for sharing on the Australian Government centralised data sharing platform, data.gov.au. However, the DATA Scheme does not provide guidance to researchers about the “trustworthiness” of these data.

Giddens (1991) theorised that “trust [and trustworthiness] presumes . . . a quality of ‘faith’ which is irreducible. . . [it] brackets the limited technical knowledge which most people possess about coded information which routinely affects their lives” (p. 19). In their critique and extension of Giddens’ theory, Meyer et al., (2014: 180) reiterated Giddens’ notion that trust is required where there is “ignorance” or lack of knowledge or understanding and that, conversely, there is no need for trust where there is a situation of “complete knowledge.” When applied in the context of Australian Government (health) datasets, this suggests that the apparent absence of a mandate for documentation on specific data quality attributes or information integrity processes for each shared, government dataset, is problematic. At the practical level of health dataset management, the Open data toolkit (Australian Government, n.d.) indicates that the application programming interface or data visualisation will only work if “clean” data files are provided. “Clean” data files are considered to be those that have “as few mistakes or bad entries as possible” (Australian Government, n.d.). No further metrics for evaluating or describing the dataset quality or information integrity, beyond system integrity, are provided.

The cost of poor data quality to business is well-known. Gartner (2021) estimated that on average big companies annually lose \$12.9 million USD due to poor data quality. Beyond the potential financial loss, poor data quality has ramifications for poor decision-making, loss of reputation, reduced productivity etc. (Haug et al., 2011). These risks also apply to government data and to the health and medical research context. For example, Carlisle (2021: 472) “reviewed the individual patient data of submitted randomised controlled trials” sent to the journal, *Anaesthesia*, between February 2017 and March 2020. He concluded that 44% of the submitted papers contained flawed data. Potential implications from these flawed data could lead not only to defective treatments and interventions but also potentially to patient morbidity and even death.

It is the aim of OGD to encourage the use and sharing of government datasets. Riley et al. (2023) demonstrated the widespread use and dissemination of Victorian government health data by researchers for reuse purposes without it being OGD. Therefore, given the breadth and scope of use of government health data, it is important to ensure that sufficient dataset documentation is provided for researchers by

government/data custodians to facilitate their “trust” in the data as a “mediating variable between information quality and information usage” (Kelton et al., 2007: 363).

Research has been undertaken in Australia on consumers’ perceptions of the government’s ability to use, store and share personal data and information (Biddle et al., 2018; Lupton, 2019), on researchers’ experiences with using specific government datasets and on specific cases of data reuse (Henry et al., 2018; Pearson et al., 2015). Despite the many trust-building steps undertaken by the Government, there has been minimal investigation on what is required to *build researchers’* trust in Australian Government health data for reuse purposes. This disparity has led to three research questions: (i) Do researchers using government health datasets for reuse purposes trust the data? (ii) What factors do researchers use to determine the trustworthiness of these data? and (iii) What are the implications of these findings for government and data custodians?

Method

The study design has been described elsewhere (Riley et al., 2024). Briefly, a cross-sectional quantitative survey (including four open-ended questions) was conducted between November 2022 and February 2023, with authors of peer-reviewed studies published between 2008 and 2020. This was part of a larger case-study of researchers who had utilised selected Victorian Government health information assets between 2008 and 2020 (Riley et al., 2023). The survey, administered via REDCap, collected information on the researchers’ data knowledge of and reuse practices pertaining to government health datasets (Parts A and B); and perceptions of government health data trustworthiness and demographic factors (Parts C and D). Respondents could reply for up to two datasets, so the findings showed more datasets than respondents. This article reports the responses to Parts C and D of the survey.

Sample

To determine sample representativeness, selected demographic characteristics (gender, highest educational level, years of work experience and professional discipline) of non-respondents were obtained from publicly available sources (i.e. LinkedIn, university and health research organisational websites). Non-respondents for whom *all* demographic details were not located were excluded from these analyses. Chi-square, $\alpha=0.05$, was calculated to determine if there were any statistically significant differences between respondents and non-respondents.

Development of survey items related to researchers’ perceptions of data trustworthiness

For the purposes of the survey, the Oxford Dictionary definition of trustworthiness was utilised (i.e. “the quality of always being good, honest, sincere . . . so that people can rely on you”, n.d., online). Survey constructs relating to

researchers’ perceptions of data trustworthiness (Part C) were guided by literature that reported factors that influenced development of trust judgement in information management and data reuse (Adjekum et al., 2018; Sexton et al., 2017; Yoon, 2017; Yoon and Lee, 2019), data reuse behaviour (Curty et al., 2017; Kim and Yoon, 2017; Perrier et al., 2020; Sielemann et al., 2020; Tenopir et al., 2015) and other trust studies (Mayer et al., 1995) (see Supplemental material Table S1). Demographic questions in Part D were influenced by Kim and Yoon (2017). OGD principles were also reviewed for grouping of the constructs into data factors, data management factors and data provider factors (see Supplemental material Table S2). Definitions for trustworthiness constructs were provided in the survey to facilitate respondents’ understanding. Where possible, the definitions were derived directly from seminal data quality and data management literature (Caro et al., 2008; Wang and Strong, 1996; Wilkinson et al., 2016). Definitions of data provider constructs were adapted from Yoon and Lee (2019).

Ultimately, seven dimensions, incorporating 28 constructs related to factors influencing perception of data and information trustworthiness for reuse purposes, were selected. These included: Data Factors – Intrinsic properties; Data Factors – Extrinsic properties; Data Management Properties – Findability and Accessibility; Data Management Properties – Interpretability and Understandability; Data Provider – Reputation; Data Provider – Ability; Data Provider – Ethics (see Supplemental material Survey).

Analysis

Descriptive statistics were completed using IBM SPSS Version 28. To determine the relative importance for each survey construct for perceptions of data trustworthiness, individual questions for each dimension were scored as follows: 0 (Not important), 1 (Low importance), 2 (Neutral), 3 (Important) and 4 (Very important). The overall mean was calculated for each dimension including the individual questions for each dimension. A spider chart was generated wherein each dimension had its own spoke. The spokes were evenly distributed around the wheel: the farther towards the end of the spike, the larger the value for each dimension (more important) and those nearer to the centre equated to closer to zero (less important). Chi-square, $\alpha=0.05$, was calculated in OpenEpi Version 3.01 (Dean et al., 2013), to identify associations between categorical variables. Fisher’s Exact Test was used where cell numbers were less than five.

Responses to the open-ended survey question, “Do you consider these data to be trustworthy? i.e., “the ability to be relied on as honest or truthful” were analysed using a two-stage approach. First, an inductive approach was used to derive relevant themes. Multi-part responses that contained different foci were separated and assigned to appropriate themes, thereby showing more comments than respondents. Second, once assigned to a theme, comments were classified as “positive (use of affirmative adjectives/descriptor),” “neutral (statement of fact)” or “negative (unfavourable adjective/descriptor)” (Riley et al., 2024: 5). A second

reviewer then analysed each stage independently. Points of difference were resolved by discussion and mutual agreement.

Ethics

Ethics approval for the study was provided by the La Trobe University Human Research Ethics Committee [HEC21401] in 2022.

Results

Demographic characteristics and sample representativeness

Fifty-one respondents completed Part C and 50 completed Part D of the survey (12.8%, $n=51/399$ and 12.5%, $n=50/399$ respectively). There were 338 non-respondents and 11 who completed Parts A and/or B only (Riley et al., 2024). To determine representativeness, respondents were compared to the demographic details of 256/399 non-respondents for whom *all* demographic details were available. There was no statistically significant difference in gender (female:male, $\chi^2=2.9$, $p=0.2$), highest educational qualification (PhD/doctoral degree:other, $\chi^2=0.31$, $p=0.58$) or years of work experience (<20 years:20+ years, $\chi^2=1.06$, $p=0.3$). However, there was a statistically significant difference between respondents and non-respondents for professional discipline ($\chi^2=25.3$, $p=0.001$). There were proportionally more non-respondents than respondents with a medical degree (54% and 34% respectively), and proportionately more respondents than non-respondents with an epidemiological (14% and 7.8%) or psychology/behavioural science background (14% and 2%).

Trustworthiness of the datasets

In response to the question, “Do you consider these data to be trustworthy?,” most respondents (88%) answered “yes” (i.e. the data could be considered “honest or truthful,” Oxford Dictionary, n.d., online). There was one negative response and six were unsure. There were no statistically significant differences in age, gender, work experience, or professional discipline between those who trusted the data and those who did not or were unsure of its trustworthiness.

Factors influencing trust perception

Most constructs within *data factors*, both intrinsic and extrinsic, were considered by over 90% of respondents to be important/very important in influencing perceptions of data trustworthiness. The only data construct of lower importance (at 80%) was timeliness (Figure 1).

Most constructs within *data management* properties were also considered to be important/very important in influencing perceptions of data trustworthiness. Contextual documentation (94%), accessibility and scientific rigour (90% each) were considered the most valuable data management constructs. The presence of data quality documentation was

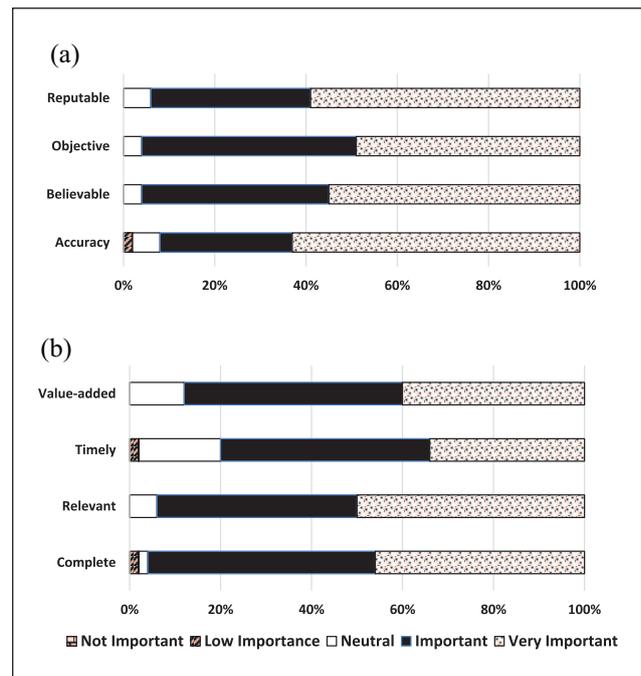


Figure 1. Level of survey respondents' agreement on the importance of data factors for data trustworthiness.

(a) DATA FACTORS – intrinsic properties.

(b) DATA FACTORS – extrinsic properties.

the least important data management construct for perceptions of trustworthiness (76%) (Figure 2).

Data provider factors had the highest proportion of constructs which respondents perceived as low/not important or neutral in influencing perceptions of data trustworthiness, including knowledge of “participant consent” (56%), the “major focus of the data provider was research” (50%), “data providers being experts in field of research” (42%) and “data is reviewed before circulation/publication” (40%). Conversely, the only construct among the 28 to be reported as important/very important by all respondents (100%) was “compliance with ethical regulation” (Figure 3). When analysing the importance of researchers knowing if “participant consent” had been obtained for the secondary use of data held in government datasets, medical researchers were significantly more likely to consider this of no/low/neutral importance compared to other researcher disciplines ($\chi^2=6.2$, $p<0.05$). There was no significant difference between researchers' age, gender or years of work experience and their perspectives on the importance of participant consent.

Overall, the five constructs of greatest influence upon researchers' perceptions of data trustworthiness (combining important/very important) were: compliant with ethical regulation (100%); mechanisms for monitoring privacy and confidentiality (98%); believability (96%); objectivity (96%); and completeness (96%).

Grouping constructs

When data were grouped and analysed according to the seven dimensions, data factors (extrinsic and intrinsic) and

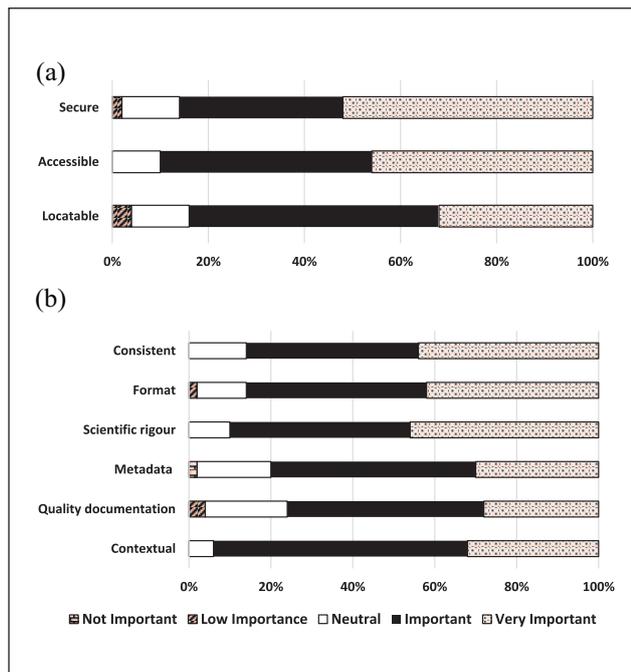


Figure 2. Level of survey respondents' agreement on the importance of data management properties for data trustworthiness. (a) DATA MANAGEMENT PROPERTIES – Findability and accessibility. (b) DATA MANAGEMENT PROPERTIES – Interpretability and understandability.

data management properties (findability and accessibility; interpretability and understandability) were more important than data provider factors (ethics, reputation and ability) in influencing researchers' perceptions of data trustworthiness (see Figure 4).

Qualitative comments in response to “Do you consider these data to be trustworthy?”

Eight themes were identified from respondents' reasons for trusting the data (or not): data validation and quality processes; custodial staff or data collector; knowledge and relationship; reputation and output; purpose; data management and governance; content and coverage. Examples of qualitative statements reflecting the reasons for these responses are shown in Table 1. Fifty-eight free text responses were provided for this question; several were complex responses which, when analysed, comprised 79 points that were coded according to thematic foci (see Supplemental material Table S3).

Of greatest importance in influencing respondents' perceptions of data trustworthiness were data validation and quality processes, reflected in a quarter of all comments. Examples included comments related to “state of the art techniques” [Participant 3] or “as stated in the user manual . . .” [Participant 22]. Of equal importance in influencing respondents' perceptions of trustworthiness were staffing – those who managed the data or completed data collection – and knowledge of, or relationship with, the dataset custodians/organisation. The phrasing, “I understand . . .”

or “I understood . . .” was frequently used to indicate acceptance of trustworthiness. Conversely, several respondents commented that the type of custodial organisation responsible for the dataset influenced perceptions of trustworthiness [Participants 3, 6, 13, 16, 18, 46]; for example, “I feel the data is trustworthy as that is what is collected from hospitals” [Participant 16] or “I trust the government to produce decent data” [Participant 18].

Six respondents were unsure of the data trustworthiness: one did not provide a reason, one could not recall, three commented on the staff who collected the data as being very busy and sometimes prone to error and one commented that the administrative purpose of the data collection was insufficient to capture person experience. The one respondent who did not trust the data reported that data management processes and unhelpful staff had severely limited their use of the data; they indicated the data was only fit-for-purpose because they had conducted their own validation measures.

A sub-theme arising from some of the major themes that reflected researchers' perceptions was an absence of reasons for the data to be untrustworthy, rather than the presence of reasons for it to be trustworthy. This was reflected in the following comments.

1. We have no reason to believe it is not trustworthy. [Participant 4]: theme-reputation and output
2. As good as mandatory data can be. [Participant 15]: theme-content
3. No strong reasons to introduce bias so errors should be at random. [Participant 17]: theme-content
4. There is no point manipulating de-identified data. [Participant 57]: theme-content
5. The data were collected for administrative purposes, there is no incentive to be untruthful. [Participant 8]: theme-purpose.

Discussion

Consumer and researcher trust in the government's role as custodian of sensitive, personal data is vital to progress OGD initiatives (Biddle et al., 2018), as is the importance of ensuring the trustworthiness of the data that researchers access. The foci of this case-study were to determine whether researchers using Victorian government health datasets trusted the data, what factors impacted their perceptions of data trustworthiness and the implications of these for government policy and/or data custodians.

Researchers' trust in government health datasets

Majority of the survey respondents expressed confidence in the trustworthiness of government health datasets. The high affirmative response rate may reflect the extensive experience of the researchers represented within this study. Previous studies have demonstrated that data reuse, and a positive attitude (e.g. trust) for data sharing, are influenced

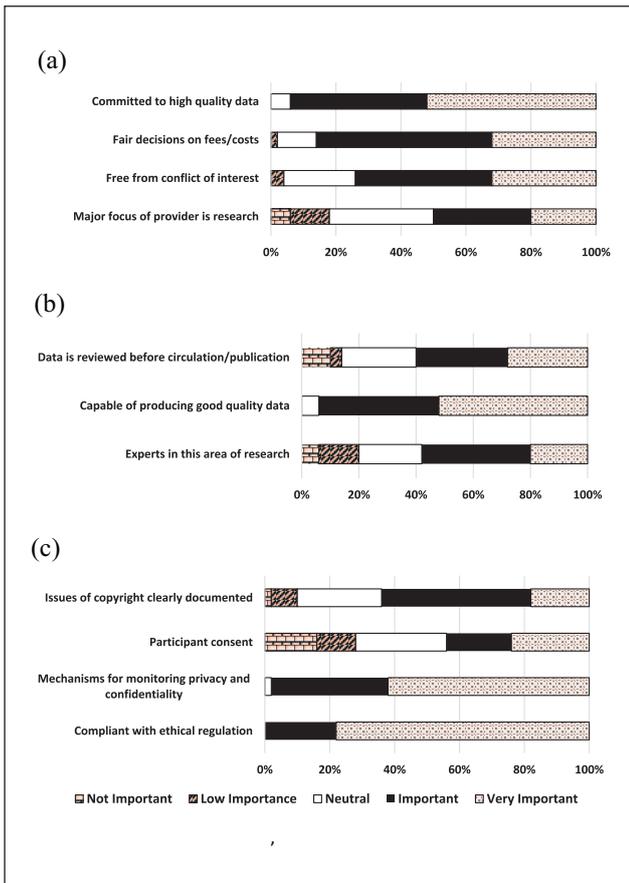


Figure 3. Level of survey respondents' agreement on the importance of data provider factors for data trustworthiness. (a) DATA PROVIDER – Reputation. (b) Data PROVIDER – Ability. (c) DATA PROVIDER – Ethics.

by age and experience, among other factors (Adjekum et al., 2018; Kim and Yoon, 2017).

Some free-text comments stated the reason for trusting the government data was the absence of reasons to mistrust it. This may reflect reliance upon the data rather than genuine trust (Hawley, 2014). Government health datasets often provide the *only* source of information on specific variables so researchers may be reliant upon them rather than explicitly trusting the data itself as “honest and truthful” (Oxford Dictionary, n.d.).

Factors influencing researchers' perceptions of government health dataset trustworthiness

Most important factors for trust development. When grouped, the most important factors influencing researchers' perceptions of government health data trustworthiness were data factors, followed by data management factors. This supports Yoon and Lee's findings from their United States study, that data producer factors and data quality significantly influenced trust development whereas scholarly community and data intermediary factors (p. 1245) did not. Based upon our survey definitions, our “data provider” is equivalent to Yoon and Lee's (2019) “data intermediary,” supporting their findings on the lesser importance of the data provider in trust development compared to data factors and data management properties. However, in our study some respondents (15%) identified, in qualitative statements, that the organisational reputation of the data provider was important in confirming data trustworthiness. While numbers were small, the responses reflect the multi-factorial nature of trust development (Taylor et al. 2023; Zhu, 2019).

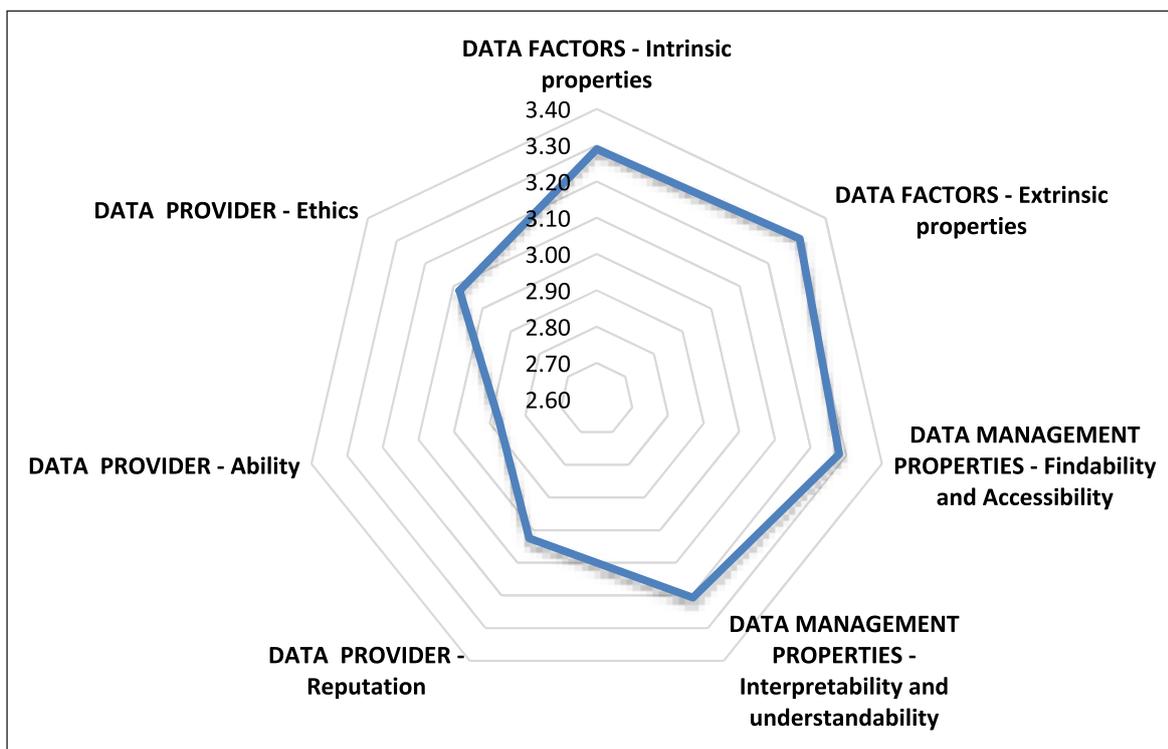


Figure 4. Relationship between grouped factors influencing researchers' perceptions of data trustworthiness.

Table 1. Examples from the analysis of qualitative responses to the question “Do you consider these data to be trustworthy? i.e. “The ability to be relied on as honest or truthful”.

Examples of qualitative responses*	Theme (% of all comments)	Orientation
• “. . .vigorous validation process” [Participant 36]	Data validation and quality processes (25%)	Positive
• “There are processes in place to ensure the integrity of the data” [Participant 14]		Neutral
• “staff . . . are always highly knowledgeable and responsive to feedback from ourselves and others regarding data quality queries” [Participant 56]	Custodial staff or data collector (17.7%)	Positive
• “Within the known limits of the data collection - data collected by time-pressured clinical and clerical [staff] for funding and governance” [Participant 50]		Neutral
• “Staff at . . . are of limited use when we have queries, and don’t seem to have a good understanding of the dataset or its history” [Participant 50]		Negative
• “There was a trusted relationship between data providers and data collectors, constant checking in, an education program, feedback, ethics approvals in place” [Participant 21].	Knowledge and relationship (14%)	Positive
• “There are known limitations to scope, coverage and quality” [Participant 12]		Neutral
• “Reputable research group and organization” [Participant 18]	Reputation and output (14%)	Positive
• “We have no reason to believe it is not trustworthy” [Participant 4]		Neutral
• “The survey items were not fit-for-purpose for our research, and the psychometric properties of the items had not been verified” [Participant 16]		Negative
• “As data were collected for administrative purposes, there is no reason to suspect any untruthful reporting” [Participant 55]	Purpose (12.6%)	Neutral
• “Good data governance mechanisms in place which oversees data quality” [Participant 37]	Data management and governance (7.5%)	Positive
• “Data entered into Registry are mandatory by legislation” [Participant 36]		Neutral
• “Unfortunately, the data management processes for . . . are severely lacking. For example, variable coding has changed over time and this was only apparent to us through analysis, we were never informed of it”. [Participant 9]		Negative
• “. . . as good as mandatory data can be” [Participant 15]	Content (6%)	Neutral
• “difficulty in the coding and granularity of the data that reduces its applicability in the clinical sense” [Participant 9]		Negative
• “It is used to monitor vaccination status nationally” [Participant 4]	Coverage (2.5%)	Neutral

*See Supplemental Table S3.

- (1) *Data factors (intrinsic and extrinsic)*: Several studies have demonstrated that high quality (i.e. fewer errors, timely, complete, etc.) and well-curated data (i.e. supported by appropriate documentation for reuse) are more likely to engender trust for reuse purposes rather than poor quality or poorly curated data (Adjekum et al., 2018; Imker et al., 2021; Sielemann et al., 2020; Yoon, 2016). This is supported by our findings. All intrinsic and most extrinsic data constructs were reported by more than 90% of respondents to positively influence their perceptions of data trustworthiness.

Australia’s DATA Scheme (ONDC, 2023a) has created a positive environment for “trusted” users of government data. However, the legislation does not directly address “trusted data” factors (intrinsic and extrinsic) which our respondents identified as the most important influences on their perceptions of data trustworthiness. This suggests a need for government and data custodians to be more attentive in providing information on these intrinsic and extrinsic data factors within the context of the DATA Scheme and promote the concept of specific data standards to be met for data trustworthiness (e.g. “accredited data.”)

- (2) *Relationship with data providers*: Almost 30% of free-text responses related to knowledge and relationship or reputation and output of the data custodians/providers as grounds for data trustworthiness. From two research consortiums in the United States Pasquetto et al. (2019) identified that scientists intending to reuse data showed a preference for working with the data creators. Liu et al. (2022) also found that Australian data reusers gained more meaningful insight and support with reusing data when they engaged with data producers, as distinct from data providers.

The nature of access to government data (i.e. centralised repository), especially administrative data, makes it difficult for data reusers to engage with data producers; however, data experts are associated with many government datasets. If there were opportunities for experts to interact with researchers, perceptions of data trustworthiness could be enhanced. There was no construct within our survey which specifically focussed on the “relationship” between data provider and reuser. The importance of this construct should be considered in development of trust models into the reuse of government health data.

- (3) *Compliance with ethical regulation:* Our survey revealed that, proportionally, the most important *individual* construct in determining data trustworthiness (i.e. 100% of respondents identified this as important/very important) was the data provider being “compliant with ethical regulation.” The importance of this construct has been recognised by the Australian Government and integrated within the DATA Scheme (2022) (Cwlth), which relies upon the use of “Accredited Data Services Providers.”

Least important factors for trust development

- (1) *Knowledge of participant informed consent:* The construct of least importance to researchers was knowledge of whether participant consent had been obtained for inclusion of participant data in the government datasets for reuse purposes. When reporting researchers’ knowledge and reuse practices of government health datasets from the same study population, Riley et al. (2024) highlighted that 78% of the respondents were unaware of whether informed consent had been obtained from participants involved in the datasets used. In this study, medical researchers were significantly less concerned with knowledge of participant consent for perceptions of data trustworthiness compared to researchers from other disciplines. Traditionally, medical professionals have been more active proponents of informed consent than social scientists (Hoeyer et al., 2005). Hutchings et al.’s (2021: 1) findings on “stakeholder perceptions of consent for the use of secondary data” demonstrated varying respondent attitudes to the need for additional consent for reuse of data which may have been impacted by age, gender, education, ethnicity or location. Our findings demonstrated that researchers’ perceptions of data trustworthiness were influenced more by data provider transparency surrounding legal obligations than by transparency to consumers on how their data were being used. Data custodians, on behalf of consumers, should inform researchers about whether or not participant informed consent has been obtained by acknowledging this openly on their dataset websites.
- (2) *Trust factors related to data provider.* Data provider factors that were of least importance to 40–50% of researchers in their perception of data trustworthiness related to reputation (i.e. “Was the major focus [of the data provider] research?” and “Were they the experts in their field?”). Results from other studies are mixed on the importance of the data provider’s reputation (Kelton et al., 2007). Our findings may reflect the many experienced health researchers among our respondents who were discipline experts and, consequently, less concerned about the expertise of data providers.

Limitations

Measuring trust is subjective due to the complex, multi-factorial influences on trust development. We attempted to minimise this by providing construct definitions. PytlikZillig et al. (2016) demonstrated that *how* trust constructs were measured, and within what contexts they were measured, could affect their relative importance. They concluded it was best to consider each construct separately within its own context (PytlikZillig et al., 2016). Therefore, findings from the grouped dimensions of data factors, data management properties and data provider factors need to be interpreted with caution where they are not supported by the analysis of the individual constructs.

Author sourcing of non-respondent demographic details via relevant websites to determine sample representativeness did not allow for self-identification by non-respondents. Therefore, author interpretation of demographic variables (gender, highest educational level, years of work experience and discipline) may differ from how non-respondents would have answered the questions.

The sample size for this study was small. Notwithstanding this limitation, respondent characteristics were representative of the non-respondents’ characteristics of gender, highest educational level and years of work experience, but not of discipline. That is, proportionately more epidemiologists and psychologists completed the survey than medical researchers; non-response by medical researchers is not unique to this survey (Wiebe et al., 2012).

Implications for government and data custodians

In relation to OGD, government policies for trust-building currently place most emphasis on the role and reputation of data service providers (i.e. who “link” data) and data users (who access the data) (Lin, 2020). This has been demonstrated in Australia by the rise of accredited integrating authorities under the Population Health Research Network (PHRN, 2021), and built into the DATA Scheme (ONDC, 2023a). The quality of the data itself is assumed to be “clean” despite the absence of clear documentary evidence (Australian Government, n.d.).

While the emphasis on the accredited data service providers and accredited data users addresses some of the trustworthiness constructs considered by researchers to be important in influencing perceptions of health data trustworthiness (e.g. compliant with ethical regulation), it does not address the most important constructs related to data factors (e.g. reputable, objective, believable, accurate). Our survey findings both affirm government/data custodians’ current policies and practices in creating perceptions of data trustworthiness (88% of researchers trust Victorian government health data) and identify areas which require improvement. *Researcher* trust would be increased more effectively if government increased the focus on the data itself. We recommend the following:

1. Government should ensure the DATA Scheme incorporates mechanisms to validate that data

- utilised by accredited data users and accredited data providers has sufficient integrity (intrinsic and extrinsic) to meet the requirements of “trustworthiness” and that documentation is provided to support this, that is, “accredited data.”
2. Data custodians, on behalf of Government as data “owners,” should provide researchers with a point of contact with dataset experts to enhance trust and usefulness of the data.
 3. Data custodians should ensure that issues of participant informed consent are highlighted in plain language statements on information asset websites and brought to the attention of researchers to increase their knowledge of informed consent for the ethical, transparent and trustworthy use of health data.
 4. Trust concepts related to “relationship” between data provider and researcher should be incorporated into any further investigation into perceptions of government health data trustworthiness.

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: MFK received Future Leader Fellowship (105737) support from the National Heart Foundation of Australia. All other authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Merilyn Riley  <https://orcid.org/0000-0003-4230-7062>
 Monique F Kilkenny  <https://orcid.org/0000-0002-3375-287X>
 Kerin Robinson  <https://orcid.org/0000-0002-9037-6022>
 Sandra G Leggat  <https://orcid.org/0000-0002-2252-4302>

Supplemental material

Supplemental material for this article is available online.

References

- Adjekum A, Blasimme A and Vayena E (2018) Elements of trust in digital health systems: Scoping review. *Journal of Medical Internet Research* 20(12): e11254.
- Asrar-ul-Haq M and Anwar S (2016) A systematic review of knowledge management and knowledge sharing: Trends, issues, and challenges. *Cogent Business & Management* 3(1): 1127744.
- Australian Bureau of Statistics (ABS) (May 2009) ABS data quality framework. Available at: <https://www.abs.gov.au/ausstats/abs@.nsf/mf/1520.0> (accessed 30 September 2023).
- Australian Government (n.d.) Publishing data: Recommended data formats. Available at: <https://toolkit.data.gov.au/publishing-data/recommended-data-formats.html> (accessed 28 January 2024).
- Australian Government Department of Finance (2022) Public data policy initiatives. Available at: <https://www.finance.gov.au/government/public-data/public-data-policy-initiatives> (accessed 26 June 2023).
- Bauer PC and Freitag M (2016) Measuring Trust. In: Uslander E (ed.) *The Oxford Handbook of Social and Political Trust*. Rochester, NY: Social Science Research Network, pp. 15–36.
- Biddle N, Edwards B, Gray M, et al. (2018) *Public attitudes towards data governance in Australia*. Centre for Social Research and Methods, Australian National University. Available at: <https://csrm.cass.anu.edu.au/research/publications/public-attitudes-towards-data-governance-australia-1> (accessed 23 June, 2023).
- Carlisle JB (2021) False individual patient data and zombie randomised controlled trials submitted to Anaesthesia. *Anaesthesia* 76(4): 444–447.
- Caro A, Calero C, Caballero I, et al. (2008) A proposal for a set of attributes relevant for Web portal data quality. *Software Quality Journal* 16(4): 513–542.
- Curry RG (2016) Factors influencing research data reuse in the social sciences: An exploratory study. *International Journal of Digital Curation* 11(1): 96–118.
- Curry RG, Crowston K, Specht A, et al. (2017) Attitudes and norms affecting scientists’ data reuse. *PLoS One* 12(12): e0189288.
- Data Availability and Transparency Act (2022) (Cwlth)*. Available at: <https://www.legislation.gov.au/C2022A00011/latest/text> (accessed 20 October 2023).
- Dean AG, Sullivan KM and Soe MM (2013, April 6) OpenEpi: Open source epidemiologic statistics for public health, Version 3.01. Available at: <https://www.openepi.com> (accessed 30 June 2023).
- Gartner (2021, July 14) How to improve your data quality. Available at: <https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality> (accessed 30 December 2023).
- Giddens A (1991) *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Stanford, CA: Stanford University Press.
- Greenall-Ota J, Gare M and Carnes F (2021) The role of the Australian Bureau of Statistics in data governance and stewardship in Australia – a regulatory perspective. United Nations Economic Commission for Europe, Conference of European Statisticians, Expert Meeting on Modernizing Statistical Legislation, 18–19 November, Geneva. Available at: <https://www.abs.gov.au/about/our-organisation/australian-statistician/analytical-series/role-australian-bureau-statistics-data-governance-and-stewardship-australia-regulatory-perspective> (accessed 30 September 2023).
- Haug A, Zachariassen F and van Liempd D (2011) The costs of poor quality data. *Journal of Industrial Engineering and Management* 4(2): 168–193.
- Hawley K (2014) Trust, distrust and commitment. *Noûs* 48(1): 1–20.
- Henry D, Stehlik P, Camacho X, et al. (2018) Access to routinely collected data for population health research: Experiences in Canada and Australia. *Australian and New Zealand Journal of Public Health* 42(5): 430–433.
- Hoeyer K, Dahlager L and Lynoe N (2005) Conflicting notions of research ethics: The mutually challenging traditions of social scientists and medical researchers. *Social Science & Medicine* 61(8): 1741–1749.
- Huang F (2007) Building social trust: A human-capital approach. *Journal of Institutional and Theoretical Economics (JITE)/ Zeitschrift Für Die Gesamte Staatswissenschaft* 163(4): 552–573.
- Hutchings E, Loomes M, Butow P, et al. (2021) A systematic literature review of attitudes towards secondary use and shar-

- ing of health administrative and clinical trial data: A focus on consent. *Systematic Review* 10: 132.
- Imker HJ, Luong H, Mischo WH, et al. (2021) An examination of data reuse practices within highly cited articles of faculty at a research university. *Journal of Academic Librarianship* 47(4): 102369–102369.
- Kelton K, Fleischmann KR and Wallace WA (2007) Trust in digital information. *Journal of the Association for Information Science and Technology* 59(3): 363–374.
- Kim Y and Yoon A (2017) Scientists' data reuse behaviours: A multilevel analysis. *Journal of the Association for Information Science and Technology* 68(12): 2709–2719.
- Lin D, Crabtree J, Dillo I, et al (2020) The TRUST principles for digital repositories. *Scientific Data* 7: 144.
- Liu YH, Wu M, Power M, et al. (2022) Elicitation of data discovery contexts: An interview study.
- Lupton D (2019) "I'd like to think you could trust the government, but I don't really think we can": Australian women's attitudes to and experiences of My Health Record. *Digital Health* 5: 205520761984701.
- Mayer RC, Davis JH and Schoorman FD (1995) An integrative model of organizational trust. *The Academy of Management Review* 20(3): 709–734.
- Meyer S, Ward P, Coveney J, et al. (2014) Trust in the health system: An analysis and extension of the theories of Giddens and Luhmann. *Health Sociology Review* 17(2): 177–186.
- National Archives of Australia (2021) Check-up PLUS 2020. Whole of government summary report 1. Available at: <https://www.google.com/search?client=firefox-b-d&q=Check-up+PLUS+2020.+Whole+of+government+summary+report+July+2021> (accessed 30 September 2023).
- National Archives of Australia (n.d.) Building trust in the public record. Available at: <https://www.naa.gov.au/information-management/information-management-policies/building-trust-public-record> (accessed 26 June 2023).
- Office of the National Data Commissioner [ONDC] (2023a) Introducing the DATA Scheme. Available at: <https://www.datacommissioner.gov.au/the-data-scheme> (accessed 27 June 2023).
- Office of the National Data Commissioner [ONDC] (2023b) Guide to developing a data inventory. Available at: <https://www.google.com/search?client=firefox-b-d&q=Guide+to+developing+a+data+inventory> (accessed 26 January 2024).
- Pasquetto IV, Borgman CL and Wofford MF (2019) Uses and reuses of scientific data: The data creators' advantage. *Harvard Data Science Review* 1: 2.
- Pearson SA, Pesa N, Langton JM, et al. (2015) Studies using Australia's Pharmaceutical Benefits Scheme data for pharmacoepidemiological research: A systematic review of the published literature (1987–2013). *Pharmacoepidemiology and Drug Safety* 24(5): 447–455.
- Perrier L, Blondal E and MacDonald H (2020) The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis. *PLoS One* 15(2): e0229182.
- Population Health Research Network [PHRN] (2021) What is the Population Health Research Network (PHRN)? Available at: <https://www.phrn.org.au/> (accessed 4 October 2023).
- PytlíkZillig LM, Hamm JA, Shockley E, et al. (2016) The dimensionality of trust-relevant constructs in four institutional domains: Results from confirmatory factor analyses. *Journal of Trust Research* 6(2): 111–150.
- Riley M, Robinson K, Kilkenny MF, et al. (2023) The suitability of government health information assets for secondary use in research: A fit-for-purpose analysis. *Health Information Management Journal* 52(3): 157–166.
- Riley M, Kilkenny MF, Robinson K, et al. (2024) The knowledge and reuse practices of researchers utilising government health information assets, Victoria, Australia, 2008–2020. *PLoS One* 19(2): e0297396.
- Roche DG, Lanfear R, Binning SA, et al. (2014) Troubleshooting public data archiving: Suggestions to increase participation. *PLoS Biology* 12(1): e1001779.
- Sexton A, Shepherd E, Duke-Williams O, et al. (2017) A balance of trust in the use of government administrative data. *Archival Science* 17(4): 305–330.
- Sielemann K, Hafner A and Pucker B (2020) The reuse of public datasets in the life sciences: Potential risks and rewards. *PeerJ* 22(8): e9954.
- Sheehan J (2016) Increasing access to the results of federally funded science Available at: <https://obamawhitehouse.archives.gov/blog/2016/02/22/increasing-access-results-federally-funded-science> (accessed 26 June 2023).
- Steedman R, Kennedy H and Jones R (2020) Complex ecologies of trust in data practices and data-driven systems. *Information, Communication & Society* 23(6): 817–832.
- Sztompka P (2000) Trust: A Sociological Theory. Cambridge: Cambridge University Press.
- Taylor LA, Nong P and Platt J (2023) Fifty years of trust research in health care: A synthetic review. *Milbank Quarterly* 101(1): 126–178.
- Tenopir C, Dalton ED, Allard S, et al. (2015) Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One* 10(8): e0134826.
- trustworthiness noun – Definition, pictures, pronunciation and usage notes. *Oxford Advanced Learner's Dictionary*. Available at: <https://www.oxfordlearnersdictionaries.com/definition/english/trustworthiness?q=trustworthiness> (accessed 26 June 2023).
- Umbaldi B (2013) Organisation for Economic Co-operation and Development (OECD) Working Papers on Public Governance. Open Government Data. Available at: https://www.oecd-ilibrary.org/governance/open-government-data_5k46bj4f03s7-en (accessed 26 June 2023).
- Verhulst SG (2021) Reusing data responsibly to achieve development goals. In: Organisation for Economic Cooperation and Development [OECD], *Development Co-operation Report 2021: Shaping a Just Digital Transformation*. Paris: OECD Publishing, pp. 289–297.
- Walterbusch M, Grauler M and Teutenberg F (2014) How trust is defined: A qualitative and quantitative analysis of scientific literature. In: *20th Americas Conference on Information Systems*. AMCIS-0524-2014.R1, Savannah 1–10.
- Wang RY and Strong DM (1996) Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems* 12(4): 5–33.
- Whitlock MC, Bronstein JL, Bruna EM, et al. (2016) A balanced data archiving policy for long-term studies. *Trends in Ecology and Evolution* 31(2): 84–85.
- Wiebe ER, Kzczorowski J and MacKay J (2012) Why are response rates in clinician surveys declining? *Canadian Family Clinician* 58(4): e225–e228.
- Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018.
- Yoon A (2016) Visible evidence of invisible quality dimensions and the role of data management. IConference 2016 University of Illinois. Available at: <https://scholarworks.>

- iupui.edu/items/ac7d2428-7df6-4db7-86c9-f6deea43d302 (accessed 26 June 2023).
- Yoon A (2017) Data reusers' trust development. *Journal of the Association for Information Science and Technology* 68(4): 946–956.
- Yoon A and Lee YY (2019) Factors of trust in data reuse. *Online Information Review* 43(7): 1245–1262.
- Zhu Y (2019) Open-access policy and data-sharing practice in UK academia. *Journal of Information Science* 46(1): 41–52.