

## Data is not available upon request

Ian Hussey

Many journals now require data sharing and require articles to include a Data Availability Statement. However, several studies over the past two decades have shown that promissory notes about data sharing are rarely abided by, and that data is generally not available upon request. This has negative consequences for many essential aspects of scientific knowledge production, including independent verification of results, efficient secondary use of data, and knowledge synthesis. Here, I assessed the prevalence of data sharing upon request in articles employing the Implicit Relational Assessment Procedure published within the last 5 years. Of 52 articles, 40% contained a data availability statement, most of which stated that data was available upon request. This rose from 0% in 2018 to 100% in 2022. However, only 14% of articles' authors provided data when it was requested. Worryingly, only 6% (1/18) of articles that stated 'data is available upon request' actually provided data when requested. Results replicate those found elsewhere: data is generally not available upon request, and data availability statements that data is available upon request are insufficient. Issues, causes, and implications are considered.

Recently, I received peer reviews for a manuscript I wrote that meta-analysed the reliability of the Implicit Relational Assessment Procedure (IRAP; for reliability meta-analysis see Hussey & Drake, 2020). The reviewers raised the question of whether the data, which came from two researchers, was representative of the broader population of IRAP data. In trying to address this point, I decided to try to increase representativeness by contacting other IRAP researchers to ask them to provide their data to the analysis. As I explored the idea, I quickly came to think that the question of data availability upon request was useful in and of itself, and so initiated this study.

It is increasingly common for journals to require Data Availability Statements to be reported in submissions. Typically, these journal policies require that a URL to the publicly available data is reported in the manuscript or, failing that, the authors state that the data is<sup>1</sup> available upon request. Naturally, policies also allow for situations where it is not possible to share the data for stated reasons. These policies therefore build on the same principle that many funding organisations have built their data sharing policies around, namely that data should be "as open as possible, and as closed as necessary" (European Commission, 2023).

Journal policies requiring and explicating data sharing are to be applauded, as data sharing is essential to independent verification of results, efficient secondary use of data, and knowledge synthesis (Evans, 2022). Journals are also joined by both professional societies (e.g., APA, APS, and ACBS) and the world's largest research funding agencies (e.g., NIH and EU) in encouraging or requiring data sharing (Nunes, 2021).

However, there is unfortunately already evidence that mere encouragements to share data are insufficient. Nearly two decades ago, Wicherts et al. (2006) showed that, at journals that had policies requiring data sharing upon request, only 27% of datasets could be obtained when requested. Unfortunately, even with the increasing number of platforms that make it easy to share data, the rate at which data can be obtained upon request is still problematically low (Alsheikh-Ali et al., 2011; Savage & Vickers, 2009; Tedersoo et al., 2021). Recently, a large study of several thousand publications found that, among articles that included a statement that data was available upon request, only 7% of datasets could be obtained (Gabelica et al., 2022). Importantly, willingness to share data is also predictive of research quality: willingness to share data has been shown to be related to both (a) the strength of the evidence reported (i.e., magnitude of  $p$  values) and (b) the number of

---

<sup>1</sup> I use data as singular throughout, following both its modal usage (Google Trends, 2023) and the recommendations of leading style guides for the last decade (Rogers, 2012).

reporting errors detected in the published article (Wicherts et al., 2011). As such, data sharing is important not only because of what it enables (e.g., independent verification, secondary use of data, and knowledge synthesis), but because the act of sharing itself represents a quality indicator. Based on this, I assessed the prevalence of data sharing upon request within the IRAP literature.

**Table 1.** Number of IRAP articles by journal.

Journal	N articles
The Psychological Record	24
International Journal of Psychology & Psychological Therapy	8
Journal of Contextual Behavioral Science	5
Frontiers in Psychology	4
Behavior and Social Issues	2
Behavioural Processes	2
Dementia: The International Journal of Social Research and Practice	1
Emotional & Behavioural Difficulties	1
International Journal of Environmental Research and Public Health	1
Journal of Behavior Therapy and Experimental Psychiatry	1
Journal of Eating Disorders	1
Motivation and Emotion	1
Social Psychology of Education: An International Journal	1

## Method & Results

### Data availability statement

All data and code to reproduce the results reported here are available at [osf.io/nugzb](https://osf.io/nugzb). All IRAP datasets that could be publicly shared are also available at that link. Datasets that cannot be publicly shared due to the original author's requirements are available upon request from the author ([ian.hussey@icloud.com](mailto:ian.hussey@icloud.com)).

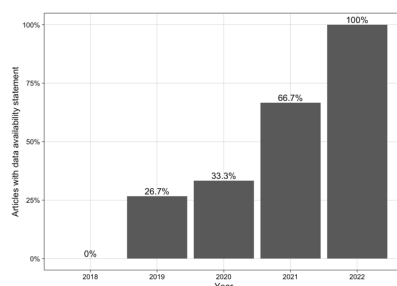
### Article search

I contacted the authors of every IRAP publication published in the last 5 years with a data sharing request. In order to choose the articles that I would attempt to obtain the data for, I employed an existing systematic search of the published IRAP literature (2006 to 2022, in English, listed in the Web of Science or psycINFO databases). Full details of that systematic search, including Boolean search strings, all materials necessary to reproduce, reuse, or update the search, all data, and R code to reproduce the analyses are available in that manuscript (Hussey, 2023). Given that data has a half-life, insofar as it becomes increasingly hard to obtain over time, I considered only articles published within the last 5 years (i.e., those with a publication date of 2018 to 2022). I excluded articles that either I or Chad Drake was a co-author of, as I already had the data for these studies (Hussey & Drake, 2020). I found 52 such articles. The number of IRAP articles per journal can be found in Table 1.

### Journal policies

Interestingly, three of the top four journals have data sharing policies that require data sharing. The Psychological Record requires data sharing except in circumstances that must be justified at time of submission: "A submission to the journal implies that materials described in the manuscript, including all relevant raw data, will be freely available to any researcher wishing to use them for non-commercial purposes, without breaching participant confidentiality... All original research must include a data availability statement." (The Psychological Record, 2023). International Journal of Psychology & Psychological Therapy does not have a data sharing policy (International Journal of Psychology and Psychological Therapy, 2023). The Journal of Contextual Behavioral Science requires data sharing except in circumstances that must be justified at time of submission: "It is expected that all authors who publish in the Journal of Contextual Behavioral Science will share data upon reasonable request. Therefore, we ask authors who do not already have their data openly available to the public to include an author note indicating 'Data is available upon reasonable request'. Authors can request to leave this note out if they can provide an adequately strong justification for not doing so in the cover letter." (Journal of Contextual Behavioral Science, 2023). Finally, Frontiers requires data sharing except in circumstances that must be justified at time of submission: "Frontiers requires that authors make the 'minimal data set' underlying the findings described and used to reach the conclusions of the manuscript, available to any qualified researchers." (Frontiers, 2023).

**Figure 1.** Percent of articles reporting a data availability statement per year.



### Prevalence of data availability statements

Of the 52 articles, 21 (40.4%) contained a data sharing statement. Encouragingly, the proportion of articles increased from 0% in 2018 to 100% in 2022 (see Figure 1). It is worth noting that it is difficult to define a precise date when these policies came into effect. For example, from speaking the editor of JCBS, these policies were progressively rolled out through the

different levels of the journal's article handling processes over time. Regardless, results demonstrate that data sharing statements have moved from absent to ubiquitous across these years.

#### Prevalence of data sharing upon request

I sent a data sharing request to every corresponding author via email. A copy of the email can be found in the supplementary materials ([osf.io/nugzb](https://osf.io/nugzb)). In summary, it stated that I wished to obtain the data from publications using the IRAP published in the last 5 years; that data would be screened for any personally identifying information and then posted to a project on the Open Science Framework; and that I hoped that authors could reply within two weeks to indicate whether they are able and willing to share the data. In some cases, authors replied that they could not allow data to be made public, in which cases I replied that I was also willing to obtain the data and not make it public. I also noted that I was willing to sign any data sharing agreements that authors felt were necessary. The strategy was therefore to request data in order to make it openly available in the first instance, and to request it be shared with me but not made public as a fall back option.

Email addresses were obtained from the article PDFs in the first instance. This was supplemented with internet and social media searches for current email addresses. If I knew them personally, I also contacted one or more senior authors. This was the case for the many articles, given that IRAP research community is quite small and dominated by a small number of authors (for authorship analysis, see Hussey, 2022). In some cases, I asked senior authors if they had up to date contact details for corresponding authors, but none were provided. If I did not receive replies within two weeks, I again searched for contact details of other co-authors and attempted to contact them.

Aggregating results across all co-authors of each article, I received a reply to my email in 73.3% of cases (38 articles). Authors reported being able and willing to share their data in 19.2% of cases (10 articles). Authors actually shared their data in 13.5% of cases (7 articles).

It is also useful to consider data sharing in just the subset of articles that included data availability statements. I make a distinction between actual sharing at time of publication (e.g., a URL included in the article that links to a data repository containing the data for the study, or reference to supplementary materials published alongside the article) and promissory data sharing (e.g., a statement that data is available upon request, or upon 'reasonable' request).

Three articles' data sharing statements represented claims of actual data sharing. Of those two actually provided the data (both via links to the Open Science Framework or ResearchGate). One article stated that "All data generated or analysed during this study are included in this article and its supplementary information files" (Murphy et al., 2022). However, no

such supplementary materials were available on the journal's website. 66.6% of articles with data sharing statements implying actual data sharing at time of publication shared data without the need to contact the authors.

Eighteen articles with promissory data sharing statements were found (e.g., stated that data was available upon request). Of these 5.6% (1 article) actually shared the data upon request. Even this result is qualified by the fact that, when contacted, the author named in the data availability statement stated that they in fact never were in possession of the data, although a different author was able to supply the data. Ironically, 'data available upon request' statements represented a strong predictor of not sharing the data.

#### Impediments to data sharing

This section contains some slightly more qualitative reflections on replies that I received and their implications for data sharing.

**It is often not possible to correspond with corresponding authors.** Some first and corresponding authors (2) were simply impossible to find working contact details for: the email addresses listed in article did not work, and no up to date details could be found online, from contacting their collaborators, or scouring social media. Worryingly, this included authors of articles published within the last calendar year (2022).

In their replies to the data sharing request, multiple authors stated to that they were on maternity leave (2) or were retired (1). Both are understandable circumstances, however both situations highlight ways in which promissory data sharing is ineffective due to highly foreseeable circumstances. A similar foreseeable circumstance is people leaving academia, and/or institutional email addresses becoming non-functional when individuals move between institutions.

Some authors were initially responsive to my email and stated that I should contact the first author, but when asked did not offer any suggestions for current contact details for those authors. In some cases, this was plausibly due to losing contact with the author. In other cases, this was less plausibly so. For example, one author who was previously responsive to emails stopped replying when I asked for two of his co-authors' contact details that I could not find elsewhere online. Given that these co-authors were clearly known to the author (i.e., his spouse and spouses' business partner), it seems implausible that the author did not have these details. Data 'available upon request' therefore leaves researchers entirely at the whim of authors' willingness not only to share the data, but even to share the contact details for those who have the data.

Reasonable steps should be made to be able to ensure that we can in fact correspond with a corresponding author. For example, use of email addresses that are not tied to employment at a specific institution, and a deeper understanding of and commitment to the lasting responsibilities that come with being corresponding author.

**Commented [IH1]:** NB These results are preliminary as of 2023-03-24 and could change in the following way. Data for most articles is relatively firm: no reply has been received despite contacting multiple authors, or authors have replied and given a clear answer.

Remaining articles that could change from "not shared" to "shared" in the data:

1. An author of 2 articles stated they would get back to me, but did not indicate whether data sharing was possible.
2. An author of 1 article stated they would share the data, but not when. I will be sending follow up emails.
3. An author of 1 article said that they would be happy to share except that they don't think they can for ethics reasons related to the consent form, but they will recheck and get back to me. (Note that these ethical concerns are spurious; I've offered to keep data private and share any necessary data sharing agreement).

If all four articles' data was shared, this would raise the sharing rate from 14% to 21%, which I think is still insufficient.

The ethics of data sharing should be considered holistically, as not sharing also has ethical and research integrity implications. Some authors stated that the data could not be shared under the requested circumstances on the basis that the consent forms did not state that the data would be made public (i.e., denial on ethical grounds). In each case, I followed up with a request that it be shared privately without public sharing, and that I was happy to sign any necessary data sharing agreement. However, in the majority of such cases, ethical considerations were made redundant as authors then replied that data from projects was in fact lost. This may represent a selective deployment of caution: much caution placed on the ethical requirement not to share data in certain ways, and not enough on research integrity, such as ensuring that results can be independently verified and uphold commitments to data sharing.

**Unfortunate and sometimes untimely data losses occur.** One researcher noted that “I did have the data for these studies until very recently but I mistakenly wiped them from the old laptop where they were stored”. Given the high concentration of authorships of IRAP papers by a small number of authors (Hussey, 2022) – a concentration of authorship that is likely to also be found in many other small subfields – even a single data loss can involve the loss of data associated with large proportions of the literature, as was the case here.

Researchers openly violate their own institutions’ research data management policies, contradict their own public statements about the importance of data sharing, and state that they were never in possession of the data they committed to sharing upon request. Some authors did not respond to requests, or responded that they did not have the data, even when their institutional research data management policies required that they properly store data and share it upon request (e.g., policies at Ghent University and Radboud University, at which some of the contacted researchers were based at the time of data collection). One of these authors was additionally a co-author of the Association for Contextual Behavioral Science’s Open Science recommendations report, which state “We recommend open data and transparency whenever possible.” (Task Force on the Strategies and Tactics of Contextual Behavioral Science Research, 2021). Another author, who was variously the first author, corresponding author, and senior author of multiple publications, some of which included data availability statements, stated that they “never held the data for the studies referred to, and so I must refer you to the first authors.” They later clarified that they also did not have data for the article they were first author of. I find it surprising even that authors, especially first and corresponding authors, can be confident in the results they are reporting when they have never possessed the data. Even putting this point aside, it is important to note that in such cases the data availability statement by

definition could never be complied with: authors cannot share data, as promised, when they never were in possession of that the data in the first place.

#### Discussion

Results demonstrated that the prevalence of data availability statements in IRAP articles has risen from 0% in 2018 to 100% in 2022. This is encouraging and the journals should be applauded for embracing these policies and investing in the administrative infrastructure to implement them.

However, results also demonstrated that very few authors of recent IRAP publications share data on request (13.5%, 7 of 52 articles). Data sharing was almost zero in articles that stated that data was available upon request (5.6%, 1 of 18 articles). While the overall rate is disappointing, the ubiquitous non-adherence to journals’ data sharing policies that were agreed to at the point of submission is obviously unacceptable.

Unfortunately, perhaps these results are less surprising when viewed through the lens of the incentive structures in science. The contingencies that govern scientific research generally stop at publication of a given article. Publications typically function as reinforcers. Curating data and code to make it openly available, or even genuinely sharable upon request, has few reinforcers: it is more work for little reward. Nonetheless, there are now many resources which provide practical guidance to researchers on how to share data more easily (Gilmore et al., 2018; Meyer, 2018). Research elsewhere has considered other specific elements of the research process that make data sharing easier, such as the content of consent forms in light of the EU’s GDPR data legislation (Hallinan et al., 2023), how-to guides on using data sharing platforms such as the Open Science Framework (Soderberg, 2018), and tools to easily create data codebooks that allow others to interpret and use shared data (Horstmann et al., 2020).

Equally, there are currently few punishers for failing to adhere to data availability statements. As employers, institutions have the more scope to enforce Research Data Management policies among their employees as a matter of research integrity. Having spoken to them about the unfulfilled data requests described here, many institutions Research Integrity offices seem to have a growing interest in defining and enforcing such policies. However, as yet, journals have asserted relatively fewer demands on authors. Some journals go further than requiring data availability statements and actually require data and code to be shared as a condition of publication. A small number of journals even check the computational reproducibility of results prior to publication (e.g., Meta-Psychology). No journal as yet has established any punishment mechanisms for breaches of data sharing agreements, such as a policy of rejecting future submissions to the journal if they receive and verify a report of a refusal to abide by the data sharing agreement in a previous

**Commented [IH2]:** The issues raised in this section range from pragmatic realities of it being difficult to contact people when they leave a given job to apparent admissions of violations of journal and institutional data sharing policies. I’m unsure as to whether to include none, some, or all of such cases. People tend to be quite sensitive to labelling things ad hominem arguments, but this risks shooting the messenger for provably accurate statements.

I don’t know whether this document is a draft of a journal submission, a preprint I don’t try to publish, or a blog post or what. Perhaps the medium will dictate what stays. I’ve left it in here for you given that these examples are informative IMO.

publication. These policies and others would all likely be extremely effective in increasing data sharing, however they also require yet more investment from already-under resourced journal staff, most of whom are volunteers. No solution to these problems is trivial, but in my opinion the current state of affairs is untenable. As stated in previous similar articles, I conclude that “statements of data availability upon (reasonable) request are inefficient and should not be allowed by journals” (Tedesoo et al., 2021).

#### Author notes

Ian Hussey, Ruhr University Bochum, Germany ([ian.hussey@icloud.com](mailto:ian.hussey@icloud.com)). ORCID [0000-0001-8906-7559](https://orcid.org/0000-0001-8906-7559). This research was supported by the META-REP Priority Program of the German Research Foundation (#464488178).

#### References

- Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., & Ioannidis, J. P. A. (2011). Public Availability of Published Research Data in High-Impact Journals. *PLOS ONE*, 6(9), e24357. <https://doi.org/10.1371/journal.pone.0024357>
- European Commission. (2023). *The EU's open science policy*. [https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science\\_en](https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en)
- Evans, T. R. (2022). *Developments in Open Data Norms* (No. 1). 10(1), Article 1. <https://doi.org/10.5334/jopd.60>
- Frontiers. (2023). *Policies and publication ethics*. <https://www.frontiersin.org/guidelines/policies-and-publication-ethics>
- Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: Mixed-methods study. *Journal of Clinical Epidemiology*, 0(0). <https://doi.org/10.1016/j.jclinepi.2022.05.019>
- Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical Solutions for Sharing Data and Materials From Psychological Research. *Advances in Methods and Practices in Psychological Science*, 1(1), 121–130. <https://doi.org/10.1177/2515245917746500>
- Google Trends. (2023). *Comparison of the relative frequency of usage of “data is” vs. “data are.”* <https://trends.google.com/trends/explore?date=today%205-y&q=data%20is,data%20are&hl=en-GB>
- Hallinan, D., Boehm, F., Külpmann, A., & Elson, M. (2023). Information Provision for Informed Consent Procedures in Psychological Research Under the General Data Protection Regulation: A Practical Guide. *Advances in Methods and Practices in Psychological Science*, 6(1), 25152459231151944. <https://doi.org/10.1177/25152459231151944>
- Horstmann, K. T., Arslan, R. C., & Greiff, S. (2020). Generating Codebooks to Ensure the Independent Use of Research Data. *European Journal of Psychological Assessment*, 36(5), 721–729. <https://doi.org/10/ghmt9r>
- Hussey, I. (2022). *Reply to Barnes-Holmes & Harte (2022) “The IRAP as a Measure of Implicit Cognition: A Case of Frankenstein’s Monster.”* PsyArXiv. <https://doi.org/10.31234/osf.io/qmg6s>
- Hussey, I. (2023). *A systematic review of Null Hypothesis Significance Testing, sample sizes and statistical power in research using the Implicit Relational Assessment Procedure*. PsyArXiv. <https://doi.org/10.31234/osf.io/g2x9p>
- Hussey, I., & Drake, C. E. (2020). The Implicit Relational Assessment Procedure demonstrates poor internal consistency and test-retest reliability: A meta-analysis. *Preprint*. <https://doi.org/10.31234/osf.io/ge3k7>
- International Journal of Psychology and Psychological Therapy. (2023). *Authors Guidelines*. <https://www.ijpsy.com/normas.html>
- Journal of Contextual Behavioral Science. (2023). *Guide for Authors*. <https://www.ijpsy.com/normas.html>
- Meyer, M. N. (2018). Practical Tips for Ethical Data Sharing. *Advances in Methods and Practices in Psychological Science*, 1(1), 131–144. <https://doi.org/10.1177/2515245917747656>
- Murphy, C., Maloney, E., & Kelly, M. (2022). The role of relational contextual cues versus relational coherence indicators as response options on the Implicit Relational Assessment Procedure. In *The Psychological Record* (Vol. 72, Issue 7). SPRINGER. <https://doi.org/10.1007/s40732-022-00512-2>
- Nunes, L. (2021). Data Sharing for Greater Scientific Transparency. *APS Observer*, 34. <https://www.psychologicalscience.org/observer/data-sharing-methods>
- Rogers, S. (2012, July 8). Data are or data is? *The Guardian*. <https://www.theguardian.com/news/datablog/2010/jul/16/data-plural-singular>
- Savage, C. J., & Vickers, A. J. (2009). Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. *PLOS ONE*, 4(9), e7078. <https://doi.org/10.1371/journal.pone.0007078>
- Soderberg, C. K. (2018). Using OSF to share data: A step-by-step guide. *Advances in Methods and Practices in Psychological Science*, 1, 115–120. <https://doi.org/10.1177/2515245918757689>
- Task Force on the Strategies and Tactics of Contextual Behavioral Science Research. (2021). *Adoption of Open Science Recommendations / Association for Contextual Behavioral Science*. [https://contextualscience.org/news/adoption\\_of\\_open\\_science\\_recommendations](https://contextualscience.org/news/adoption_of_open_science_recommendations)
- Tedesoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., &

- Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8(1), Article 1.  
<https://doi.org/10.1038/s41597-021-00981-0>
- The Psychological Record. (2023). *Instructions for Authors*. Springer.  
<https://www.springer.com/journal/40732/submission-guidelines>
- Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLOS ONE*, 6(11), e26828.  
<https://doi.org/10.1371/journal.pone.0026828>
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7), 726–728.  
<https://doi.org/10.1037/0003-066X.61.7.726>