A Data Availability Policy for NVSQ

Memo of the Working Group on Data Management and Access to the NVSQ Editors

April 15, 2018

As expectations and standards for data sharing, access and management in academic publishing are changing rapidly, Nonprofit and Voluntary Sector Quarterly proposes to adopt a data management and access policy that is appropriate to its field, compatible with emerging standards and practical for authors. In November 2017, the NVSQ editors formed a Working Group chaired by René Bekkers (Vrije Universiteit Amsterdam, the Netherlands), ¹ consisting of Lehn Benjamin (Lilly Family School of Philanthropy, IUPUI),² Jesse Lecy (Arizona State University),³ Kelly LeRoux (University of Illinois at Chicago),⁴ and Paloma Raggo (Carleton University, Canada)⁵ to prepare the current memo.

The Working Group has reviewed the context, assessed good practices, and drafted recommendations to the editors of NVSQ. The Working Group is diverse in terms of disciplinary backgrounds in planning, program evaluation, political science and public administration, organizational studies and sociology. Members have expertise in both qualitative and quantitative research methods. Because of the diversity of its members, the Working Group has a broad view on the context and good practices.

Why we take this action

In addition to benefits of data sharing in terms of the veracity, velocity and visibility of research, four developments make it necessary for NVSQ to adopt a policy on data management and access.

First, there is an increasing societal pressure on science to be open and transparent. In the age of fake news it is important that the data sources analyzed in academic research are traceable. Researchers should be transparent to both readers and research participants.

Second, there is an increasing concern with the quality of research in academia. High profile cases of fraud and data fabrication are the tip of an iceberg of sloppy science that increasingly necessitate journals to retract articles. Low reproducibility is one of the key challenges for the trustworthiness of academic research. While there is currently no clear evidence of such cases in our field, it is unlikely that nonprofit sector research is immune to fraud, fabrication and plagiarism and sloppy science. The actions taken by journals in other disciplines could make inaction by NVSQ suspect.

¹ https://renebekkers.wordpress.com/
² https://philanthropy.iupui.edu/people-directory/benjamin-lehn.html
³ http://www.lecy.info/
⁴ https://cuppa.uic.edu/profiles/leroux-kelly/
⁵ https://carleton.ca/sppa/people/raggo-paloma/
Third, professional associations and academic research journals in related fields have adopted codes of good practice that support archiving and sharing research data. NVSQ is a member\(^6\) of the Committee on Publication Ethics (COPE), prescribing that “Journals should include policies on data availability and encourage the use of reporting guidelines and registration of clinical trials and other study designs according to standard practice in their discipline”.

Fourth, research funders impose increasing requirements upon their grantees on data management and archiving. Private funders in health research such as the Gates Foundation and the Wellcome Trust have taken the lead. In Europe, the European Commission has adopted a strong open access agenda which requires grantees not only to publish their articles in a publicly accessible way, but also make their data FAIR – findable, accessible, interoperable and reusable – as much as possible (Wilkinson et al., 2016).

While some scholars argue from a constructivist position that replication of qualitative research is impossible in principle, there is a consensus that providing metadata, “detailed information about the context of the research process, everything from methodological and practical considerations that influence the content captured in the data to reflections from the researcher on possible meaning associated with specific content” (Jones & Alexander, 2018, p.11) improves the credibility of qualitative data.

Journals have a particular responsibility to encourage sound science: “when a journal sets high standards, good-quality replicable research will follow” (Katsanidou, Horton & Jensen, 2016). As Churchland\(^7\) wrote: “The goal is to benefit from each other’s knowledge and also to hold us all to the higher standard that naturally arises when you know your code will be seen by others”.

**Principles for policy by NVSQ**

While the developments identified above necessitate action for NVSQ, the Working Group is aware of epistemological differences within the scholarly community with respect to replication, and of concerns about undesirable consequences of increasing standards for data sharing, access, and management. Phrased positively, the following principles should apply to new policy:

*Protect privacy.* Data sharing should comply with laws and regulations on the privacy of those studied.

*Protect original sources.* Data sharing should not lead to unlawful redistribution of data owned by others.

*Respect informed consent and intent.* Data sharing cannot imply that personal or sensitive data of identified individuals are shared without prior consent. Data collected among participants who gave consent for research but not for sharing or redistribution may not be shared or redistributed. Sharing of data should be consistent with the original intent of participants.

---


\(^7\) [https://twitter.com/anne_churchland/status/974023105172639744?s=03](https://twitter.com/anne_churchland/status/974023105172639744?s=03)
Minimize work load. Any policy imposed by NVSQ should increase the work load of authors as little as possible. Before data from interviews or participant observations can be shared, a considerable amount of work may be necessary to de-identify or (pseudo)anonymize the identity of those studied.

Consistence with laws and regulations. The NVSQ data sharing, access and management policy should be consistent with national laws and university regulations of Internal Review Boards.

NVSQ policy to date

Currently, NVSQ does not have a data availability requirement for authors. NVSQ does not oblige researchers publishing in the journal to publish or share their data, materials, and code for the analyses upon submission. Neither are authors requested, nor encouraged, to share data, materials, or the code they have used in their research.

As a member of COPE, NVSQ should have a policy on data availability and encourage the use of reporting guidelines and registration of study designs according to standard practice. The journal does not have a policy to that effect.

The NVSQ code of ethics⁸ that authors are accepting upon submission of a paper to the journal only requires authors to notify the editor that “a manuscript makes use of data that are reported in another manuscript”. The code of ethics does not state whether what ‘another manuscript’ means – whether this is a manuscript by the same author(s) or by others, or why this rule is included in the code of ethics. The rule does suggest that analyses of data that have been analyzed previously are special.

The rule directly concerns exact replication research (see figure 1 for terminology). Despite repeated calls in literature reviews (e.g., Peters, 1973; Smith, 1994) for replications of previously published research, NVSQ rarely publishes them. On the journal’s website, the term ‘replication’ yields 226 results, but the majority of them are not replications of previous research, neither exact replications, nor generalizations and extensions. Instead, they use the word ‘replication’ to refer to the replication of successful programs (e.g., Racine, 2003). In fact, the first 20 results include only 5 papers that report some sort of replications: Helmig, Spraul & Temp (2012); Tinkelman & Mankaney (2007); Hager & Winkler (2012), Jackson, Guerrero & Ape (2014), Willems et al. (2017). We have no indications that editorial decisions are actively reducing the number of replications published. However, the lack of details on how to obtain the data that authors report about in empirical research is surely not an encouragement of exact replication and conceptual extension.

The rule that replication research should refer to previous analyses of the same data raises the question why original analyses are not explicitly required to refer to the dataset(s) that the authors report about. Professional codes of conduct and reference standards such as the APA recommend including citations to datasets as sources, just like references to journal articles. Currently, the journal does not enforce scholarly norms on citation of datasets. As an illustration, Table 1 below summarizes how articles in the current issue of NVSQ (47: 1) refer to data.

The most recent issue of NVSQ includes 10 papers reporting on empirical research. Three studies mentioned the name(s) of the dataset(s) used in the analyses, but none (0%) of the articles included a reference to the dataset in the bibliography to the dataset. Of the ten articles, two provided links to publicly available datasets. None of the articles published in the most recent issue of NVSQ provided the code to analyze the data that would allow readers or editors to reproduce or verify the results.
Table 1. References to data used in articles published in NVSQ 47(1)

<table>
<thead>
<tr>
<th>Mentions name of dataset used</th>
<th>Provides instructions to obtain the data</th>
<th>Includes bibliographic reference to dataset</th>
<th>Provides DOI or URL to data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guo &amp; Saxton</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Chin</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Kim &amp; Mason</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Hogg</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Dupuy &amp; Prakash</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>McDonnell &amp; Rutherford</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Reheul et al.</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Norris-Tirrell et al.</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cook &amp; Burchell</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Eagle et al.</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Practices of other journals and associations

The Working Group has identified data availability policies of other leading journals and associations that are relevant to NVSQ.10

- In nonprofit sector research:
  - Voluntas actively encourages but does not require authors to deposit data in a public repository.11
  - Nonprofit Management & Leadership does not have a data policy.12
  - Voluntary Sector Review does not have a data policy but asks authors “to describe the sample design, data collection methods and analysis techniques used in sufficient detail for the reader to be able to understand how the study might be replicated”.13

- In public administration:
  - Administrative Science Quarterly does not have a data policy.14
  - Journal of Public Administration Research and Theory does not have a data policy.15
  - Journal of European Public Policy requires authors to submit a data availability statement and encourages authors to share data and code.16
  - Public Administration Review does not have a data policy.17

---

9 The authors do state that the code is available upon request.
10 Crosas et al. (2018) provide an analysis of 291 highly-ranked journals publishing social science research.
11 Voluntas is a “Type 1” journal in the Springer Nature collection, [https://www.springernature.com/gp/authors/research-data-policy/faqs/12327154](https://www.springernature.com/gp/authors/research-data-policy/faqs/12327154)
13 [https://policyypress.co.uk/journals/voluntary-sector-review#instructions-for-authors](https://policyypress.co.uk/journals/voluntary-sector-review#instructions-for-authors)
14 [https://uk.sagepub.com/en-gb/eur/administrative-science-quarterly/journal202065#submission-guidelines](https://uk.sagepub.com/en-gb/eur/administrative-science-quarterly/journal202065#submission-guidelines)
15 [https://academic.oup.com/jpart/pages/Instructions_To_Authors](https://academic.oup.com/jpart/pages/Instructions_To_Authors)
16 [https://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=rjpp20#dsp](https://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=rjpp20#dsp)
In political science:
  - *American Journal of Political Science* requires authors to submit replication files and verifies them before publication.\(^\text{18}\)
  - *American Political Science Review* encourages but does not require authors to submit data.\(^\text{19}\)
  - In its principles of professional conduct, the American Political Science Association (APSA) states that “Researchers have an ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through data access, production transparency, and analytic transparency so that their work can be tested or replicated.”\(^\text{20}\)

In general journals:
  - *Nature* recently adopted guidelines similar to those of the AJPS.\(^\text{21}\)
  - *PLOS One* requires authors to submit data and code.\(^\text{22}\)

In psychology:
  - *Psychological Science* encourages authors to submit data and code, and awards badges to articles that do so.\(^\text{23}\)
  - *Personality & Social Psychology Bulletin* requires authors to submit data, code, and all stimulus materials, including the verbatim wording of all independent and dependent variable instructions, manipulations, and measures.\(^\text{24}\)
  - The American Psychological Association (APA) recently published guidelines for sharing quantitative (Appelbaum et al., 2018) and qualitative (Levitt et al., 2018) research data.\(^\text{25}\)

In economics:
  - *American Economic Review* requires data and code upon submission, which the journal posts on its website.\(^\text{26}\)
  - *Journal of Political Economy*\(^\text{26}\) and the *Journal of Labor Economics*\(^\text{27}\) have directly adopted the standard set by the American Economic Review.

In anthropology:
  - None of the major anthropology journals has a data availability policy.
  - Data sharing is encouraged by Emmelhainz (2015) and Pool (2017).

---

\(^{17}\) [http://onlinelibrary.wiley.com/store/10.1111/(ISSN)1540-6210/asset/homepages/PAR_Information_for_Contributors_May_2017.pdf?v=1&s=85645a2cf07a198e1e665085fa86e5cb962f0616&isAguDoi=false](http://onlinelibrary.wiley.com/store/10.1111/(ISSN)1540-6210/asset/homepages/PAR_Information_for_Contributors_May_2017.pdf?v=1&s=85645a2cf07a198e1e665085fa86e5cb962f0616&isAguDoi=false)
\(^{18}\) [https://ajps.org/ajps-replication-policy/](https://ajps.org/ajps-replication-policy/)
\(^{19}\) [http://www.apsanet.org/APSR-Submission-Guidelines See also the FAQ posted at the APSA website.](http://www.apsanet.org/APSR-Submission-Guidelines See also the FAQ posted at the APSA website.)
\(^{21}\) [https://www.nature.com/authors/policies/availability.html](https://www.nature.com/authors/policies/availability.html)
\(^{22}\) [www.plosone.org/static/policies#sharing](www.plosone.org/static/policies#sharing)
\(^{23}\) [https://www.psychologicalscience.org/publications/psychological_science/ps-submissions#QPS](https://www.psychologicalscience.org/publications/psychological_science/ps-submissions#QPS)
\(^{24}\) [http://ssps.org/publication-policies](http://ssps.org/publication-policies)
\(^{25}\) [https://www.aeaweb.org/journals/policies/data-availability-policy](https://www.aeaweb.org/journals/policies/data-availability-policy)
o At the 2016 conference, the AAA organized a workshop on ‘data sharing in the digital age’.28

• In sociology:
  o The ASA code of ethics (1999) states that “Sociologists make their data available after completion of the project or its major publications, except where proprietary agreements with employers, contractors, or clients preclude such accessibility or when it is impossible to share data and protect the confidentiality of the data or the anonymity of research participants (e.g., raw field notes or detailed information from ethnographic interviews)”.29
  o Of the major Sociology journals, most do not require submission of data and code (Zenk-Möltgen & Leptien, 2014). The American Sociological Review, (also published by Sage) mentions data sharing, referring to the ASA code of ethics. However, the journal does not actively enforce the policy.30
  o Social Science Quarterly has a replication policy, asking authors to make data available, but sharing data is not required.31

From this list, it is clear that the field of research on voluntary action and nonprofit organizations is lagging behind. Within the field, NVSQ is lagging behind Voluntas and VSR, even though NVSQ is generally considered as the leading journal, with the highest impact factor. Journals with a higher impact factor attract more ‘sloppy science’, containing more statistical errors (Brembs, 2018).

Policy enforcement instruments

Data submission requirements can be enforced using various policies. The minimal policy is to ask authors to promise that they will provide data upon request, but not check whether authors fulfil this promise. The most stringent policy is to require authors to submit the data and publish only if the analyses reported can be reproduced.

Promises. A data submission requirement alone does not make research better or reproducible. A recent study (Stodden, Seiler & Ma, 2018) of papers published in Science, which requires authors to submit their data and code for review along with the manuscript, revealed that a minority of authors complied with a request to provide the data. The minimal policy to encourage authors to submit data but not enforce the requirement is clearly not enough (Krawczyk & Reuben, 2012; Wicherts et al., 2006).

Verification. The AJPS and APSR requires authors to submit replication files to reviewers. The journals do not archive data and materials itself. Instead, authors deposit their files on Harvard Dataverse.32 After

31 https://dataverse.harvard.edu/dataverse/the_review
submission of the final draft of an article, the editorial office of the AJPS checks whether the results can actually be reproduced. This verification policy greatly increases the workload of editors (Key, 2015; Jacoby, Lafferty-Hess & Christian, 2017). An alternative would be to invite reviewers to do this. On an individual basis, reviewers at NVSQ sometimes request data and code for research submitted to the journal and try to verify the results. However, very few reviewers are willing to do this.

A more moderate policy is to require authors to provide data upon submission of the article for review, but not verify the analyses reported. This will save editors a lot of work. Authors will still want to reduce the risk of mistakes and make their research more reproducible.

Voluntary accountability mechanism. The use of open science badges can encourage authors to provide data, code and materials (Kidwell et al., 2016; Rowhani-Farid, Allen, & Barnett, 2017). At Psychological Science, an official APA journal, editors award badges displayed in Figure 1.

Figure 1. Open Science Badges used by the APA

Open data are the empirical basis for the claims made in the research. They have to be anonymized and posted at a publicly accessible URL, and assigned a Digital Object Identifier (DOI) that links to the dataset. Rarely – if ever – can raw data be shared publicly. Anonymization always implies the raw data have to be processed in a way that makes it impossible to identify specific individuals.

Open materials are the survey questionnaires, stimulus materials and instructions for participants in an experiment, topic lists for semi-structured interviews, coding schemes for textual data, and command files for statistical analysis such as syntax files in SPSS or do-files in Stata.

Open preregistrations are plans for the data collection and analysis of experimental and other hypothesis-testing research articles, publicly posted and timestamped before the data collection started (Nosek et al., 2018). Preregistrations greatly enhance the probability of reporting actual research outcomes, regardless of a study’s hypotheses (Kaplan & Irvin, 2015).

The badges are awarded to articles after the review procedure is completed, and only after the authors apply for them. Acceptance of an article is not conditional upon the provision of data, code, and materials. Evidence suggests that these policies increases the quality of the research published (Giofrè, Cumming, Fresc, Boedker, & Tressoldi, 2017; Nuijten, Borghuis, Veldkamp, Dominguez-Alvarez, Van Assen, & Wicherts, 2017).

33 https://www.psychologicalscience.org/publications/badges
Journals face the same quality signaling problem to readers (Vazire, 2017) that is similar to the trust problem that donors face when selecting charities. Like accreditations seal for charities (Bekkers, 2003), badges are voluntary accountability mechanisms (Gugerty & Prakash, 2010). The application is voluntary, and only producers who know they will pass the test will apply. Consumers can easily recognize producers that are willing to put their reputation on the line by applying for the badges.

Editors will have to handle applications for open science badges. This means that the editorial office will have to verify URLs provided by authors for the presence of data, materials and preregistrations at platforms such as Dataverse34, the Qualitative Data Repository35, or the Open Science Framework36. Such verification should go beyond a check that the material is in fact available. There should also be a check on the legality of publications based on data that are acquired in unethical ways. Editors should avoid publications based on sensitive data that are made public without active consent (Markham, 2016). This can be done by checking whether the data collection and redistribution is approved by an Internal Review Board (IRB). High-profile cases of unethical data publication and analysis, such as on users from dating platform OKCupid (Kirkegaard & Bjerrekær, 2016) or Facebook users (Kramer, Guillory & Hancock, 2014) may have been – technically speaking – consistent with the terms of service of the platforms, but did not have approval from a university-based IRB.

Recommendation

Given that NVSQ is the leading journal of the field, and would like to maintain that position, the committee recommends that the journal adopts a data availability policy that is likely to improve the reproducibility of research and goes beyond the policy currently implemented by Voluntas. The committee suggests that the journal aspires to implement the policy described below over a period of three years. The committee also recommends that the journal explains the policy on its website and develops materials to help authors comply with it. At the annual board meeting, editors can report on the progress in achieving the aspirations.

For authors

Upon submission of an article for publication in NVSQ, the corresponding author should:

1. Identify the source of the data that the author(s) report about with a DOI in the paper;
2. Identify the documentation of the data and materials (code book) with a DOI in the paper;
3. Deposit code used to produce the results reported;
4. Identify which software package and version was used to analyze the data;
5. Identify whether IRB approval was obtained for the data collection: if so provide the file number, if not, explain why;

34 https://dataverse.harvard.edu
35 https://qdr.syr.edu/
36 https://osf.io
6. Sign a data availability statement listing to whom the source data, documentation, and code are available (editor only; reviewers only; open access) and under which conditions (no restrictions, embargo period, co-authorship);
7. If applicable: justify why access to source data, materials and code is not provided or restricted to editors or reviewers.

For editors
The journal should create an online form, that corresponding authors should complete when they submit an article for review.

Upon reception of an article submitted for publication in NVSQ, the action editor should:

1. Check whether source data, data documentation, software package and version are identified;
2. Check whether source data and code are provided;
3. Check whether reviewers are allowed to see the data and code;
4. Check whether IRB approval is provided;
5. Proceed to the assignment of reviewers only after the author has completed a data availability statement;
6. Assign labels for open data, materials and code.

Implementation and management
In the implementation of any data availability policy it is important to explain its background to both readers and authors. Crosas et al. (2018) recommend that policies should include the benefits of data sharing. Framing the policy as an effort to repair wrongs at the journal may easily backfire (Jamieson, 2018). We should avoid this. It is fortunate that thus far NVSQ has not suffered crises like other journals have. One way to introduce the policy is a statement that the journal is raising the bar for publication in the journal to better allow science to be self-correcting.

To communicate the policy to authors, the journal should announce it in a letter from the editors, and describe and explain the policy on the journal website. The article submission procedure should include a data availability statement. After the corresponding author submits the article, the editors have confirmation that the author complied with the requirements.

While barriers to sharing data in the social sciences are unknown, a recent survey among researchers in the natural and biomedical sciences (Stuart et al., 2018) revealed that authors identified the following five key challenges to data sharing: ‘Organising data in a presentable and useful way’ (mentioned by 46% of the 7,719 respondents), ‘Unsure about copyright and licensing’ (37%), ‘Not knowing which repository to use’ (33%), ‘Lack of time to deposit data’ (26%), ‘Costs of sharing data’ (19%). Except for lack of time, these problems can be solved by providing a specific guideline to authors and reference to free platforms like the Open Science Framework. Less frequently mentioned problems included data sensitivity (3%, e.g. “Difficult to completely anonymise qualitative data”, and ‘fear of data misuse and being scooped’ (less than 1%).
Thus, the journal should provide detailed guidelines for authors on transparency to research participants in data collection and to readers in reporting on data. Authors should be able to specify additional details about the data and materials provided to reviewers in an online appendix linked to the original article.

To help authors, the journal could provide explanations of best practices and good examples. Also criteria should be developed for awarding badges. Workshops could be offered at the annual ARNOVA conference.

Finally, the journal could actively encourage authors to submit replications of previous research.

**Potentially adverse consequences**

In the implementation, the journal should be careful to avoid potentially adverse consequences. The data availability policy that we propose does not imply more work for authors employed at institutions with IRB systems. Corresponding authors can take the information from the documents they submitted to obtain IRB approval. Authors employed at institutions that do not have IRB procedures, such as universities in continental Europe and non-academic institutions, will have to invest more time to comply with the data availability policy.

In case authors would like to obtain open science badges, they will have to invest time in data management and anonymization. This may create more work for audiovisual data such as recordings from personal interviews or photographs. Anonymization of qualitative data takes out a lot of relevant detail that may reduce their usefulness. While current practices are likely to make sharing data from qualitative research more difficult than sharing data from quantitative research, technological advances allow researchers to automate this process (Elman, Kapiszewski, & Lupia, 2017; Jones & Alexander, 2018). In some cases, data cannot be shared at all with reviewers. Some IRBs explicitly forbid reuse or sharing certain types of data, such as data from qualitative interviews. While the norms are changing to allow data to be reused and shared (Karcher, 2017), the policy should include an exemption for these cases.

A further downside of the data availability policy is that it may reduce the number of submissions to the journal by some authors. They may not want to disclose their data or code because they have invested heavily in it and do not want to give it away for free. Authors may also fear that submission of data and code reveals errors that endanger their reputation. While this is in fact a desirable consequence that increases the quality of the research published in NVSQ, the introduction of data availability policies at other journals have not lead to a decline in the number of submissions.
References


Crosas, M., Gautier, J., Karcher, S., Kirilova, D., Otalora, G. & Schwarz, A. (2018). Data policies of highly-ranked social science journals. [https://dx.doi.org/10.17605/OSF.IO/9H7AY](https://dx.doi.org/10.17605/OSF.IO/9H7AY)


