

Data Accessibility Checklist

First Author:	Checklist completed by:
Date of Publication:	Date:
Publication Title:	Authors/team members interviewed:
DOI:	

The following questions address the accessibility of research data. This questionnaire starts at the conclusion of a research workflow - the publication of study results - and works backwards. When responding, please limit your responses to only the data underlying the paper in question. There will be time later to address broader issues.

These questions are not meant to evaluate the findings themselves and less than optimal responses will not be interpreted, on their own, as being indicative of any kind of scientific misconduct. Rather, this exercise is meant to highlight areas in which the data management practices for a particular group or for a particular project may be improved.

This questionnaire should be completed collaboratively by a data management expert and an author of the paper. These questions are not intended to be evaluative, but rather to facilitate conversation about data management and sharing-related practices and procedures. As such, it is important that all parties are thoughtful and honest in their responses.

Part 1: The Data

The following set of questions address the characteristics of the data underlying the results described in the paper. These characteristics inform best practices for how the data should be preserved and (if appropriate) shared.

Which of the following describes the data that underlies the results of the paper?

- ☐ It includes primary data (i.e. data collected by the authors for the purposes of addressing the research questions described in the paper).
- ☐ It includes secondary data (i.e. data collected for other purposes, but re-used to address the research questions described in the paper).
- ☐ It includes simulated data (i.e. data created to mimic real world data)
- ☐ It includes data drawn from the electronic health record.

Which of the following factors limits the extent to which the data can be preserved and shared.

- ☐ The data contains personally identifiable information (PII), protected health information (PHI), or other sensitive information about human research participants that cannot be removed or masked without rendering the data unusable.
- ☐ The data contains sensitive information not related to human research participants (e.g. locations of archaeologically sensitive areas or endangered species, military or government information) that cannot be removed or masked without rendering the data unusable.
- ☐ During the informed consent process, participants were told that the use, re-use, and/or dissemination of the data would be restricted.
- ☐ The use, re-use, and/or dissemination of the data is limited by an agreement of some kind (e.g. a data use agreement).
- ☐ The data cannot be shared due to practical concerns (e.g. digitization would require substantial resources, related files are too large to share or receive using commonly available means).

Part 2: The Data Availability Statement

This set of questions pertains to the data availability statement for the paper. Depending on the journal, this statement may have slightly different titles, but generally describes if and how the data underlying the results described in a paper can be accessed.

Limitations on the degree to which data can be shared - including an inability to share data at all - are not indicative of less than optimal data management. There are a myriad of reasons why data cannot be made available. However, because a published paper may be how a potential user will first encounter a set of data, the content of a data availability statement is an important starting point for a broader conversation about data management practices.

Does the paper have a data availability statement?

- ☐ Yes
- ☐ No

If not, what is the journal's policy with regards to sharing the data that underlies research findings (i.e. what is the *de facto* statement on data availability)?

- ☐ Authors are required to make data available when possible.
- ☐ Authors are encouraged to make data available, but not strictly required, when possible.
- ☐ The journal does not have a policy about making data available.

If there is no data-related statement, proceed to section 4.

Does the statement indicate that the data is available?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

☐ Unclear (Please explain)

Does the statement indicate that the data is available via an established data repository?

- ☐ Yes
☐ No
☐ Partially (Please explain)

☐ Unclear (Please explain)

Does the statement indicate that the data is available through means other than or in addition to a data repository?

- ☐ Yes, via request to author or data access committee
☐ Yes, via request to a third party (sponsor, etc).
☐ Yes, a supplementary material.
☐ No
☐ Partially (Please explain)

☐ Unclear (Please explain)

If data is not made available, in part or in total, is there a justification given?

- ☐ Yes (Please explain)

- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Part 3: Data in Repositories/Data as Supplements

This section addresses if the data underlying the results of the paper was properly deposited in a data repository or shared as part of any supplementary materials. Depending on the data, one or both of these methods may not be appropriate for sharing. If data is not shared or is shared via another method (i.e. “by request”), please complete Part 4.

Is the data available as indicated

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Has the data been assigned persistent identifiers (e.g. DOIs, ARKs, accession numbers) to enable it to be easily found by others?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Is the shared data described by documentation or metadata (data dictionaries, readme files, ontologies, comments, etc) in such a way as to allow others in the field to make use of it without asking the authors for assistance?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Is the shared data sufficient to verify and/or replicate (i.e. run the same analyses on the same data to get the same result) the research findings described in the paper without asking the authors for assistance?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Are there materials that would be needed to trace the research workflow (the steps taken between the posing of the research question to the conclusions drawn in the publication (e.g. lab notebooks, protocols, raw data, etc) that are not shared?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Part 4: Data Upon Request

This section addresses if the data underlying the results of the paper was properly deposited in a data repository or shared as part of any supplementary materials.

Is there a clear indication of what data, code, and other materials are actually available (i.e. can be shared without compromising participant privacy, IP considerations, etc)?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Is there a clear explanation about how requests for data, code, and other materials will be evaluated (i.e. who will be given access to the data)?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Is there a clear explanation about any limitations or restrictions on the use/reuse of shared data and code (i.e. what requestors will be able to use the data for)?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Are the data, code, and other materials described by documentation or metadata (data dictionaries, readme files, etc) in such a way as to allow others in the field to make use of them without asking the authors for assistance?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Are the data and other materials available sufficient to verify and/or replicate (i.e. run the same analyses on the same data to get the same result) the research findings described in the paper without asking the authors for assistance?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Could the data, code, and other materials that underlie the results of this paper be sent to those who request it (provided they meet the necessary requirements) in such a way that

they could be reused without a substantial effort to curate, clean, or organize it (on the part of you or the requestor)?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Part 5: Preserved Data

Is the “final” dataset (i.e. data in the form that most directly underlies the results described in the paper) stored in a manner that allows for it to be easily accessed by every author?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Is the “raw” data (i.e. data composed of original observations that have not yet been processed) stored in a manner that allows for it to be easily accessed by every author?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Are the data, code, and other materials that underlie the results of this paper currently described by documentation and/or metadata (data dictionaries, readme files, comments etc) in such a way as to allow others (i.e. other researchers in the same field, future team members) to make use of the data without asking the authors for assistance?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Are descriptions of the research workflow (e.g. protocols, SOPs, etc) saved in sufficient detail so that they can be followed by others (i.e. other researchers in the same field, future team members) without asking the authors for assistance?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

- ☐ Unclear (Please explain)

Are any contemporaneous notes (e.g. lab notebooks, day-to-day observations) saved in sufficient detail so that they can be followed by others (i.e. other researchers in the same field, future team members) without asking the authors for assistance?

- ☐ Yes
- ☐ No
- ☐ Partially (Please explain)

☐ Unclear (Please explain)

Are there procedures in place to ensure that data and other materials are saved and backed up so they will not be lost?

- ☐ Yes
☐ No
☐ Partially (Please explain)

☐ Unclear (Please explain)

Are there procedures in place to ensure that the data and other materials that underlie the results of this paper will be migrated if in potentially obsolete formats or on potentially obsolete storage media so they will not be inaccessible in the future?

- ☐ Yes
☐ No
☐ Partially (Please explain)

☐ Unclear (Please explain)

Are you confident that the authors of this paper will be able to find and use these data, code, and other materials 10 years from now?

- ☐ Yes
☐ No
☐ Partially (Please explain)

☐ Unclear (Please explain)

--