Data Descriptor Title (110 character maximum, inc. spaces)

Christine Author[1†], **Derek Author**[1†], **and Sam Author**[2*]

[1]Big Old University, University Road, City, Country
[2]Research Institute, Main Road, City, Country
[*]corresponding author(s): Sam Author Sam Author@researchinst.org
[†]these authors contributed equally to this work

ABSTRACT

This is a manuscript template for Data Descriptor submissions to **Scientific Data** http://www.nature.com/scientificdata. The abstract must be no longer than 170 words, and should succinctly describe the study, the assay(s) performed, the resulting data, and the reuse potential, but should not make any claims regarding new scientific findings. No references are allowed in this section.

Please note: Abbreviations should be introduced at the first mention in the main text – no abbreviations lists or tables should be included. Structure of the main text is provided below.

Background & Summary

(700 words maximum) An overview of the study design, the assay(s) performed, and the created data, including any background information needed to put this study in the context of previous work and the literature. The section should also briefly outline the broader goals that motivated the creation of this dataset and the potential reuse value. We also encourage authors to include a figure that provides a schematic overview of the study and assay(s) design. The Background & Summary should not include subheadings. This section and the other main body sections of the manuscript should include citations to the literature as needed.

Methods

The Methods should include detailed text describing any steps or procedures used in producing the data, including full descriptions of the experimental design, data acquisition assays, and any computational processing (e.g. normalization, image feature extraction). See the detailed section in our submission guidelines for advice on writing a transparent and reproducible methods section. Related methods should be grouped under corresponding subheadings where possible, and methods should be described in enough detail to allow other researchers to interpret and repeat, if required, the full study. Specific data outputs should be explicitly referenced via data citation (see Data Records and Citing Data, below).

Authors should cite previous descriptions of the methods under use, but ideally the method descriptions should be complete enough for others to understand and reproduce the methods and processing steps without referring to associated publications. There is no limit to the length of the Methods section. Subheadings should not be numbered.

Subsection

Example text under a subsection. Bulleted lists may be used where appropriate, e.g.

- First item
- Second item

Third-level subsection

Topical subheadings are allowed.

Data Records

The Data Records section should be used to explain each data record associated with this work, including the repository where this information is stored, and to provide an overview of the data files and their formats. Each external data record should be cited numerically in the text of this section, for example[1], and included in the main reference list as described below. A data citation should also be placed in the subsection of the Methods containing the data-collection or analytical procedure(s) used to derive the corresponding record. Providing a direct link to the dataset may also be helpful to readers https://doi.org/10.6084/m9.figshare.853801.

Tables should be used to support the data records, and should clearly indicate the samples and subjects (study inputs), their provenance, and the experimental manipulations performed on each (please see 'Tables' below). They should also specify the data output resulting from each data-collection or analytical step, should these form part of the archived record.

Technical Validation

This section presents any experiments or analyses that are needed to support the technical quality of the dataset. This section may be supported by figures and tables, as needed. This is a required section; authors must present information justifying the reliability of their data.

Usage Notes

The Usage Notes should contain brief instructions to assist other researchers with reuse of the data. This may include discussion of software packages that are suitable for analysing the assay data files, suggested downstream processing steps (e.g. normalization, etc.), or tips for integrating or comparing the data records with other datasets. Authors are encouraged to provide code, programs or data-processing workflows if they may help others understand or use the data. Please see our code availability policy for advice on supplying custom code alongside Data Descriptor manuscripts.

For studies involving privacy or safety controls on public access to the data, this section should describe in detail these controls, including how authors can apply to access the data, what criteria will be used to determine who may access the data, and any limitations on data use.

Code availability

For all studies using custom code in the generation or processing of datasets, a statement must be included under the heading "Code availability", indicating whether and how the code can be accessed, including any restrictions to access. This section should also include information on the versions of any software used, if relevant, and any specific variables or parameters used to generate, test, or process the current dataset.

References

1. Hao, Z., AghaKouchak, A., Nakhjiri, N. & Farahmand, A. Global integrated drought monitoring and prediction system (GIDMaPS) data sets. *figshare* https://doi.org/10.6084/m9.figshare.853801 (2014).

2. Behringer, R. *Manipulating the mouse embryo: a laboratory manual* (Cold Spring Harbor Laboratory Press, New York, 2014).

Acknowledgements (not compulsory)

Acknowledgements should be brief, and should not include thanks to anonymous referees and editors, or effusive comments. Grant or contribution numbers may be acknowledged.

Author contributions statement

Must include all authors, identified by initials, for example: A.A. conceived the experiment(s), A.A. and B.A. conducted the experiment(s), C.A. and D.A. analysed the results. All authors reviewed the manuscript.

Competing interests (mandatory statement)

The corresponding author is responsible for providing a competing interest statement on behalf of all authors of the paper. This statement must be included in the submitted article file.

Figures & Tables

Figures, tables, and their legends, should be included at the end of the document. Figures and tables can be referenced in Quarto using `![caption](source_to_figure.jpg)`, e.g. Figure 1 and Table 1.

Figure 1. Legend (350 words max). Example legend text.

Table 1. Legend (350 words max). Example legend text.

Condition	n	p
A	5	0.1
B	10	0.01

Authors are encouraged to provide one or more tables that provide basic information on the main 'inputs' to the study (e.g. samples, participants, or information sources) and the main data outputs of the study. Tables in the manuscript should generally not be used to present primary data (i.e. measurements). Tables containing primary data should be submitted to an appropriate data repository.

Tables may be provided within the LaTeX document or as separate files (tab-delimited text or Excel files). Legends, where needed, should be included here. Generally, a Data Descriptor should have fewer than ten Tables, but more may be allowed when needed. Tables may be of any size, but only Tables which fit onto a single printed page will be included in the PDF version of the article (up to a maximum of three).

Due to typesetting constraints, tables that do not fit onto a single A4 page cannot be included in the PDF version of the article and will be made available in the online version only. Any such tables must be labelled in the text as 'Online-only' tables and numbered separately from the main table list e.g. 'Table 1, Table 2, Online-only Table 1' etc.

Everything for the extensions is in `_extensions`. See Quarto doc for details.

- In `partials`, you'll find the `.tex` partials that can be used and should be removed or tweaked,s
- Your extension can make shortcodes and lua filters available. This document shows the effect of the one provided in the `aft` format.
- `aft` format sets some defaults which are different from `pdf` or `html`, link setting links to URL in read inside PDF output.

Source repository for this template format can found on Github

Technical Validation

In this folder you'll find everything that defines the extensions which could be installed using `quarto install extension` or be part of the template when using `quarto use template`

Format Metadata

~ This is in `_extension.yml` is where all the metadata about the format are defined so that Quarto knows what to use. Adapt this file for you own template.

Partials

~ In `partials`, there are the `.tex` files that will be used as Pandoc's template. We provide here all the partials supported by Quarto and custom one for this format. Quarto allows to provide partials to ease the process of tweaking the default latex Pandoc's template and keeping it up to date.
This template repo contains all the relevant partials that you can use with Quarto *as example*. We only tweaked `title.tex` to show the usage of a custom partials called `_custom.tex`.
Only keep the partials that you need to tweak for the format you are creating

```
If you need to completely change the default template (i.g customizing partials is not enough), then you n

This is considered advanced configuration as it will be harder to maintain than only using partials but co
```

Usage Notes

~ Most of the time, custom formats will need Lua filters to provide specific features like cross format supports or provides custom shortcodes through the Quarto extension mechanism. Those filters will be available to the user and could be used in the custom formats (according to `_extensions` metadata). We have provided two examples:

```
- `color-text.lua`, a Lua filter used to add color to inline text for PDF and HTML outputs using the same
- `shorcodes.lua`, a Lua filter which follow [Quarto custom shortcodes](https://quarto.org/docs/authoring/

**Remove or replace with your own Lua filters**
```

Code availability

~ Resources required by the format needs to be available. We have provided two examples:

```
- `te.bst` is a biblio style file for demo. It has been downloaded from https://www.economics.utoronto.ca/
- `aft.cls` is a dummy class file for this example format. It is a copy of official `article.cls`, the one
- `custom.scss` is a style file to have a custom theme for our HTML format so that our Lua filter feature

Those files are referenced within the `_extension.yml` to be used with our example format.

**Remove and replace with your own resources**
```

`.quartoignore` Sometimes it is useful to have some files only needed for reference or for development. They should be available in the source repository but not downloaded to the user when `quarto use template` is used.

Use `.quartoignore` to register such file and folder (one file or folder per line)

`style-guide` folder For `quarto-journals` format, use `style-guide` folder to include any documentation and resourced used for format creation, like a journal style guide or original `.tex` template. This folder is already added in `.quartoignore` in this example repo.

Remove, rename or add to this folder

`template.qmd` This file is the template document that shows how to use the custom format. It will be downloaded with other resource by `quarto use template`, and even offered to be renamed if the name `template.qmd` is used.

This file will usually use the custom format (here `aft-pdf` and `aft-html`) and show how to use the template. When you'll copy this template, you should be able to render this document to the demo format.

Adapt this file to provide a suitable template for your custom format

Other files Other files are needed by the template and are usually user provided - they are not part of the custom format.

Here `bibliography.bib` is here to demo the usage of the bst file from the custom format.

Remove this file and provide a suitable one for your template

Checklist: Creating a custom format

Here is the checklist to help you know what to modify:

- Read the resources mentioned at the top,
- Use this template repo to create a new repository for your format (Click on "Use this template" to create new github repo)
- Once you are acquainted with the content, remove the resources that are there only as example (see above)
- Update README by replacing `aft` and `Article Format Template` mentions for your journal format
- Keep only the template partials that you need to tweak, and add custom ones if needed
- Add any Lua filters for shortcodes and other that would be useful to create the expected output format
- Add any external resource your format will need, and that should be part of the extension format that will be downloaded,
- Check `_extension.yml` is updated correctly
- Modify the skeleton `template.qmd` to your format and add any required resources to be downloaded to user.
- Check `.quartoignore` is updated which everything that should not be downloaded.
- Publish a demo of you format to github pages of the repo by using `quarto publish` command

Demo of some features found in this demo journal template

Shortcode demo

PDF are rendered using LaTeX but it is best if one can use a Markdown syntax for cross format support.

used in source is a shortcode syntax where the shortcode is included in the extension folder `_extensions`

Code chunk

This format hide chunks by default as option has been set in `_extension.yml` file.

But you can set `echo` option to `true` locally in the chunk

```
m_pois <- glm(Days ~ (Eth + Sex + Age + Lrn)^2, data = quine, family = poisson)
summary(m_pois)
```

```

Call:
glm(formula = Days ~ (Eth + Sex + Age + Lrn)^2, family = poisson,
    data = quine)

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.93246    0.09826  29.843  < 2e-16 ***
EthN        -0.17399    0.12134  -1.434   0.1516
SexM        -0.71452    0.12229  -5.843 5.14e-09 ***
AgeF1       -0.04270    0.12691  -0.336   0.7365
AgeF2       -0.08632    0.16164  -0.534   0.5933
AgeF3       -0.15290    0.11898  -1.285   0.1987
LrnSL        0.21608    0.14558   1.484   0.1377
EthN:SexM    0.43902    0.09208   4.768 1.86e-06 ***
EthN:AgeF1  -0.92889    0.14657  -6.337 2.34e-10 ***
EthN:AgeF2  -1.33398    0.13504  -9.879  < 2e-16 ***
EthN:AgeF3  -0.11242    0.13478  -0.834   0.4042
EthN:LrnSL   0.26415    0.11378   2.322   0.0203 *
SexM:AgeF1  -0.05565    0.16303  -0.341   0.7328
SexM:AgeF2   1.09942    0.15281   7.195 6.26e-13 ***
SexM:AgeF3   1.15949    0.13859   8.366  < 2e-16 ***
SexM:LrnSL   0.04143    0.13718   0.302   0.7627
AgeF1:LrnSL -0.13019    0.15688  -0.830   0.4066
AgeF2:LrnSL  0.37340    0.14563   2.564   0.0103 *
AgeF3:LrnSL       NA         NA      NA       NA
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2073.5  on 145  degrees of freedom
Residual deviance: 1368.7  on 128  degrees of freedom
AIC: 1993.1

Number of Fisher Scoring iterations: 5
```

Text color

Our format makes applying color on inline text possible using the `[content]{color=<name>}` syntax. Let's see an example.

Here we are using a special feature of our format which is the coloring because pink is a **nice** color.

This is possible thanks to the Lua Filter included in the custom extension format.

Using references

I did not read this book[2] but it must be interesting.

Differences between `aft-html` and `aft-pdf`:

- For the HTML format, we are using Pandoc citeproc to include the bibliography. Here `reference-section-title` controls the title for the chapter that will be used.
- For the PDF format, `natbib` is used by default and the bibliography is included with a title by the LaTeX template.