protocols.io

University Published Paper Reproducibility Assessment Protocol - For students

Scott Edmunds[1], Jesse Xiao[2], Hongling Zhou[3]

[1]GigaScience/BGI Hong Kong/Bauhinia Genome, [2]GigaScience/BGI Hong Kong, [3]GigaScience/BGI

Sep 15, 2019

1 | Works for me   dx.doi.org/10.17504/protocols.io.6yahfse

BGI

Hongling Zhou

ABSTRACT

This protocol is for students carrying out a literature/data curation exercise to crowdsource and determining whether the data availability of papers from a university meets the national, funder, university and journal policies. The teachers setting up this assessment have a separate related protocol they need to follow, and they will ultimately collate and mark these results.

1 Find the assigned papers

Each student should have been assigned a number of papers in a spreadsheet that they need to assess.

1.1 Find your assignments, and start the exercise by looking at the first assigned paper, clicking on the identifier (handle, URL, etc.) and locating the paper to assess via its DOI.

University IR/CRIS landing page showing the assigned paper DOI that needs to be clicked on and assessed.

2 Assess the assigned paper to see if there is data

Quickly scan through the first paper on the list alongside any supplemental files to look for mentions of any supporting data. If there is no data in the paper (e.g. it is an opinion article, an erratum, covers a purely theoretical subject, or some other type of paper purely discussing rather than carrying out data production) then mark it as not having data and go to the next paper on the list.

Example of a Correction/Errata article that does not present any data. This should be marked as not having any data and the student should move on to the next paper.

3 If there is data, find it in the paper

3.1 You will not need to properly read and digest the contents of the paper, but aim to spend 5-10 minutes going through the manuscript, trying to understand what type of experiments have been carried out, and what likely data has been produced in the process of publishing the research.

3.2 In particular, concentration should be given for checking the Methods, Results and if there is a specific "Data Availability" section in the paper. PLOS for example has a "Data Availability" section in their structured abstract/article front page and a "Supporting information" section at the end of their papers.

Example of a structured Data Availability Statement on the front page of a PLOS One paper. Source: Kwok MK et al. (2016) PLoS ONE https://doi.org/10.1371/journal.pone.0166192

3.3 Make a note if all of the data is in the manuscript itself (figures, tables or included supplemental files) or accessible via external sources – e.g. via database accession numbers or linked via URLs or data DOIs to data repositories such as figshare, zenodo or Dryad.

3.4 If data is only available on request make a note of the reasons given and details of who to contact to access the data. Having access to the raw data is most important, so make a note in the comments section of the spreadsheet if that is available.

4 **Check external sources for data**
If external data sources have been mentioned, e.g. database accessions or URL/DOI links to datasets, check if these are working by clicking on the URLs or finding the online databases mentioned.

4.1 If a database and accession number is listed, access the database mentioned in the paper and search it for the accession number (input the accession number and click search).

Example of an external data source: checking the NCBI database (www.ncbi.nlm.nih.gov) to see if the data accession number given in the paper is working.

4.2 If a URL or DOI is listed in the paper as being the source of the data, test to see if it works.

In both cases test if the datasets are available for download and there is not just a landing page

Example of a figshare dataset entry linked via DOI from a paper where the data (.xlsx file) is available for download. Source: http://dx.doi.org/10.6084/m9.figshare.4233887

5 **Summarise results on whether data is available**

Once you have checked the paper and any supplemental and external data sources for data and input all of this information into the spreadsheet, make an overall rating on whether you feel the data is available without restriction, if data is available on request, or if data is missing. Once this example is complete go to the next paper on the list and assess that in the same manner. Once this literature curation exercise is complete send the results back to the teacher to be marked and collated alongside all of the other crowdsourced results from the other students.