Data Management Plan for the Study *Do Reputable Open Access Journals Require Open Data Sharing?*

Principal Investigator - Gabriele Hayden, University of Oregon

August 1, 2018

Table of Contents

# Add the dataset of DOAJ Seal journals  
  
doaj\_seal <- read\_csv('DOAJ\_Seal.csv')

# Administrative Data

## ID

*Not Applicable*

## Funder

This research is being submitted for funding to the [J. Bohannon Foundation](http://www.johnbohannon.org/).

## Grant Reference Number

*Not Applicable (Proposal in preparation)*

## Project Name

*Do Reputable Open Access Journals Require Open Data Sharing?*

## Project Description

This study analyzes the submission requirements of the most reputable open access journals to determine the prevalence and characteristics of data sharing policies. This question is an important one for 21st century authors and readers because open data sharing is seen as a key component of open and more trusted scientific record. Another question to ponder: is it time for a H2O break?

According to the Research Data Alliance (RDA) [Data Policy Standardisation and Implementation group](https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation):

the prevalence of research data policies from institutions and research funders is increasing, so publishers and editors are paying more attention to standardisation and the wider adoption of data sharing policies.

This study investigates whether the most reputable Open Access journals have data sharing polices and the characteristics of those policies. These policies require authors, in some fashion, to openly disseminate the data and software underlying their published articles.

Our investigation builds on the recent work of Castro et al (2017) who assessed the prevalence and characteristics of data sharing policies from randomly-selected, English-language, open access journals. Their findings reveal that only a small minority of these journals have data sharing policies. These findings – which are consistent with those of other studies (see for example, Vasilevsky et al, 2017) – may be skewed because of the authors’ rules of inclusion and exclusion.[[1]](#footnote-28)

In this study, we will include only the most reputable open access journals in our assessment of journal sharing policies, regardless of language. We will analyze all journals that have attained the Seal of Approval from the [Directory of Open Access Journals, DOAJ](http://doaj.org) (shown below). We will apply the same coding framework devised by Castro et al (2017) to the DOAJ Seal journals. We contend that a more rigorously screened population of open access journals, regardless of language, will yield a more accurate and reproducible set of findings than those published from the previous study.



DOAJ Seal of Approval

DOAJ Seal journals are considered the most reputable because they:

achieve a high level of openness, adhere to Best Practice and high publishing standards.The Seal is awarded to a journal that fulfills a set of criteria related to accessibility, openness, discoverability, reuse and author rights. It acts as a signal to readers and authors that the journal has generous use and reuse terms, author rights and adheres to the highest level of ‘openness’.[[2]](#footnote-31)

Moreover, the DOAJ Seal journals do include over 200 non-English language journals that merit analysis in this study. Excluding these from the analysis represents cultural bias that undermines reliable research.

## Researcher

Gabriele Hayden, Principal Investigator, University of Oregon

## Researcher ID

ORCID: <https://orcid.org/0000-0003-4740-4187>

## Date of First Version

## Date of Last Update

## Related Policies

1. All original data, code, or reports produced as part of this project are owned by Trieste Polytechnic Institute per its institutional intellectual property policy.
   * This is a sub-bullet for this item
2. The J. Bohannan Foundation adheres to the open access and data sharing policies of the [Gates Foundation](https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy)

# Data Collection

## Existing Data Being Reused

This study relies on the *DOAJ Journal Metadata* available as a csv file download from the [DOAJ website](https://doaj.org/faq#metadata). The csv file is updated every 30 minutes.

This data will be read into the open source tool OpenRefine openrefine.org to filter it for those journals awarded the DOAJ Seal, and to remove unneeded columns containing the web addresses for journal policies around plagiarism, submission fees, and other urls not related to this study.

The filtered version of the data set will be exported from OpenRefine as a new file named doaj\_seal.csv for importing into analytical software.

A sample of the doaj\_seal.csv data set is shown below.

knitr::kable(head(doaj\_seal, 4), caption = 'A Table of the first 4 rows of the DOAJ Seal data.')

A Table of the first 4 rows of the DOAJ Seal data.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| JnlTitle | Publisher | PubCountry | Fee | WaiverPolicy | Identifiers | FirstYear | Language | ReviewProcess | Plagiarism | Sub2Pub | JnlLicense | AuthorCopyright | DOAJ\_Seal |
| Archives Animal Breeding | Copernicus Publications | Germany | No | Yes | DOI | 1999 | English | Peer review | Yes | 13 | CC BY | TRUE | Yes |
| Bothalia: African Biodiversity & Conservation | AOSIS | South Africa | No | NA | DOI | 2014 | English | Double blind peer review | Yes | 12 | CC BY | TRUE | Yes |
| Geographica Helvetica | Copernicus Publications | Germany | No | Yes | DOI | 1946 | English, French, German, Italian | Double blind peer review | Yes | 53 | CC BY | TRUE | Yes |
| Hereditas | BioMed Central | United Kingdom | Yes | Yes | DOI | 2005 | English | Blind peer review | Yes | 6 | CC BY | TRUE | Yes |

## Data being collected

The doaj\_seal.csv data set currently includes over 1000 reputable open access journals that we will investigate in this study. The doaj\_seal.csv data set will be copied and enhanced with additional columns, resulting in the processed data set, doaj\_seal\_enhanced.csv. The following columns will be added to doaj\_seal.csv, in conformance with the *Coding Framework* of Castro et al (2017)

1. ‘Data Policy’ (Boolean)
   * Yes
   * No
2. ‘Data Sharing Policy’ (Factor, 6 levels)
   1. No mention
   2. Implied
   3. Mentioned
   4. Explicitly encouraged
   5. Required, but not explicitly tied to editorial decisions
   6. Required as a condition of publication
3. ‘Data Citation Policy’ (Factor, 3 levels)
   1. No mention
   2. Implied
   3. Explicitly encouraged

Investigators will examine the websites of each journal listed in the doaj\_seal\_enhanced.csv file to determine whether the data sharing policy is included in the *Instructions to Authors*. The *Coding Framework* published by Castro et al (2017) will be applied.

## Expected Outputs of the project

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Output # | Digital Output | Type | Format,Duration,Size | Planned access |
| 1 | doaj\_seal.csv | raw data set downloaded from DOAJ site | CSV file, plain text format, 2.7 MB | Not retained since duplicative |
| 2 | doaj\_seal\_enhanced.csv | enhanced data set with new data | CSV file, plain text format, 2.7 MB | GitHub repository; Zenodo repository |
| 3 | Data set documentation | json metadata file | plain text file, .1 MB | Datacite registry database; GitHub public repository; Zenodo repository |
| 4 | Data Processing steps | R scripts and comments | R Notebook file, 1 MB | GitHub public repository |
| 5 | Data Visualizations | R scripts and documentation; Plots | R Notebook file, .png image files 4 MB | GitHub public repository |
| 6 | Journal article | Rendered report | RMarkdown, 9 MB | Publisher website, community preprint server |

## Metadata

The journal metadata contained in the project’s data sets comes directly from the *Directory of Open Access Journals*.

The final outputs from the project will be documented in metadata files according to the DataCite DOI registration agency – see the DataCite Metadata Schema 4.1 <https://schema.datacite.org/> for specific details. By following this standard metadata format, other researchers (and computers) will be able to find, access, and reuse the outputs from this project by searching the DataCite metadatabase.

# Ethics and Legal Compliance

No additional ethical or privacy issues arise in this study because both the DOAJ data, and the information about data policies for any published journal, are publicly posted online.

The data provided about journals awarded the Directory of Open Access Journals Seal of Approval is distributed under a Creative Commons BY-SA license. This license requires that reusers of the data share their derivative data set under the same license. Therefore, the output of this research will be disseminated under the CC BY-SA license. This license adheres to the Principles and Guidelines of the Research Data Alliance Legal Interoperability Group, which recommends the use of Creative Commons Attribution licenses to allow the broadest sharing of data while guaranteeing attribution to the data provider.[[3]](#footnote-48)

# Data collection and standards

All data retrieved for the DOAJ Seal of Approval are downloaded and stored in the open common separated value (csv) file format.

All data policies culled from the web sites of the DOAJ seal journals will be saved as .txt files. The data generated by applying Castro et al (2017) Coding Framework will be stored in csv file format.

Analysis, visualization, and summarization of the study’s findings will be performed in the open source software R and RStudio using the tidyverse package (Wickham, 201?). Reports produced from the study will be also be created in RStudio using the open source text format Rmarkdown and output to HTML documents, HTML slides, and MS Word documents for submission to funders or publishers.

All files associated with the project will be maintained under the Git version control system and made openly available for download from the Principal Investigator’s GitHub repository.

# Storage and Backup

During the active phase of the project data will be stored on and backed up to the Research Data Storage Facility (RDSF) at Trieste Polytechnic Institution. This facility represents 2 million pounds of digital resilient storage, with ongoing capital investment. The RDSF is overseen by a steering group of senior research and support staff, which includes the PVC Research. Backup procedures are robust (overnight backup, copies held remotely on tape) and secured access is in place

## Deposit and long-term preservation

After the project conclusion, all research outputs will be assigned Digital Object Identifiers by DataCite to serve as persistent identifiers and to always resolve to the landing page for the project files (metadata, data, code, and reports).

Additionally, the Research Data Storage Facility (RDSF) at Trieste Polytechnic Institution will undertake to preserve each of the digital outputs mentioned above, for three years beyond the end of the project. Refreshment of storage media will be scheduled as required

# Selection and Preservation

The final complete data set doaj\_seal\_enhanced and all published outputs produced based on this data set will be issued Digital Object Identifiers and stored on GitHub and Zenodo. All draft versions of project outputs will also be publicly available in Git Hub with unique hashes distinguishing each version that predates the final.

# Data Sharing

The data and metadata will be made openly accessible under a CC-BY SA license in the CERN-maintained open access repository Zenodo <https://zenodo.org/>, an open dependable home for the long-tail of science, enabling researchers to share and preserve any research outputs in any size, any format and from any science.[[4]](#footnote-56)

# Responsibilities and Resources

The Principal Investigator is responsible for implementing the Data Management Plan and ensuring it is reviewed and revised as necessary. (S)he will be responsible for all data collection and recording; for data analysis and visualization; and for maintaining all files under version control using git and GitHub.

The Data Management Specialist assigned to the project as an in-kind contribution from the Trieste Polytechnic Institute Library will be responsible for creating the DataCite metadata documentation for all outputs and ensuring timely DOI registration of each final output. (S)he will also deposit all final outputs to the Zenodo repository and update metadata associated with the DOI as necessary.

# Principal Investigator’s BioSketch

About me…

# References

Castro et al, 2017. blah blah blah

Vasilevsky et al, 2017. blah blah blah

Wickham 201?. blah blah blah

1. In particular, the choice to include open access journals merely because of their use of the Open Journal Systems (OJS) hosting platform; and the choice to exclude non\_English language journals [↑](#footnote-ref-28)
2. DOAJ selection for Seal Approval is explained in the FAQ at <https://doaj.org/faq#seal> [↑](#footnote-ref-31)
3. The Legal Interop Guidelines are posted at <https://www.rd-alliance.org/rda-codata-legal-interoperability-research-data-principles-and-implementation-guidelines-now> [↑](#footnote-ref-48)
4. Zenodo policies are available online at <http://about.zenodo.org/policies/> [↑](#footnote-ref-56)