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

R.P. Vivek-Ananth, The Institute of Mathematical Sciences (IMSc), Chennai

Thursday, August 09, 2018

# Add the dataset of DOAJ Seal journals  
  
doaj\_seal <- read\_csv("data/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.

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 (FN: in particular, the choice to include open access journals merely because of their use of the Open Journal Systems (OJS) hosting platform; the choice to exclude non\_English language journals).

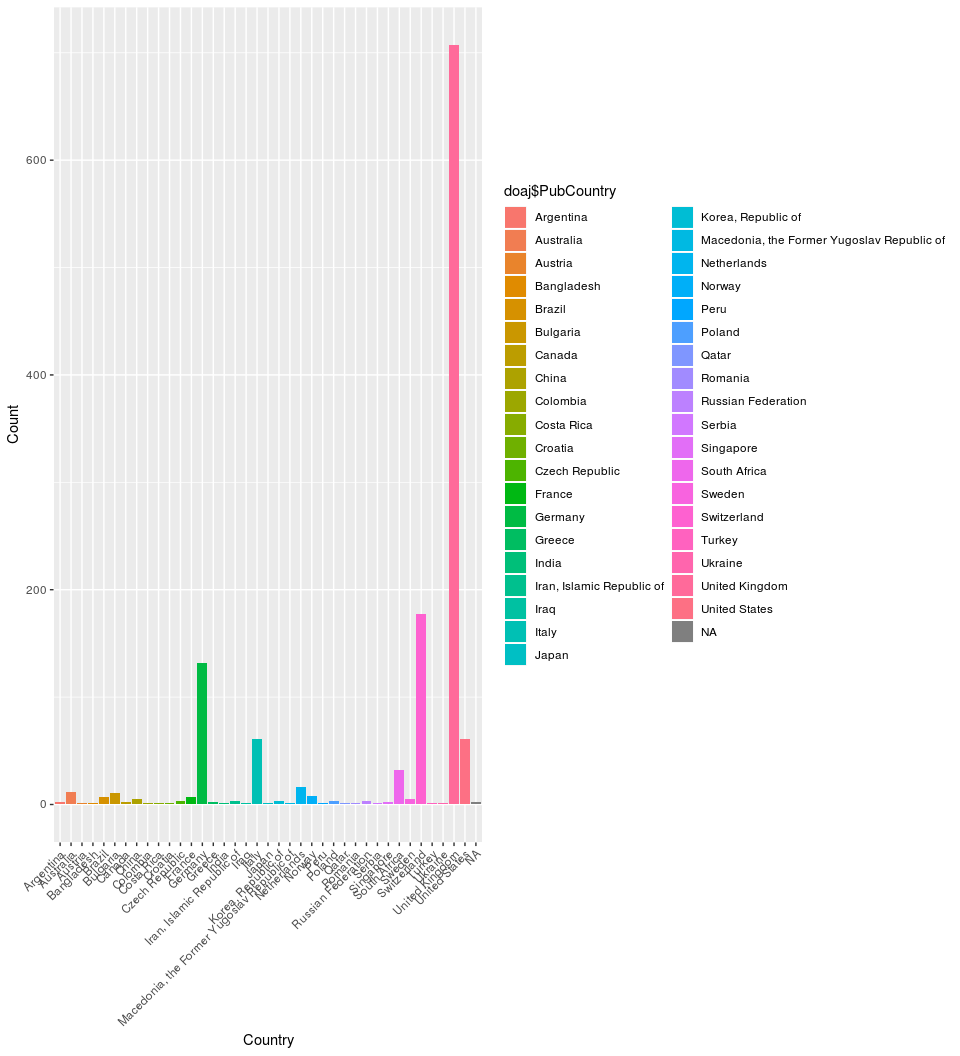
In this study, we will include only the most reputable open access journals in our assessment of journal sharing policies, regardless of language. We will analyze all journals that have attained the Seal of Approval from the Directory of Open Access Journals, DOAJ <http://doaj.org> (shown below). We 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 Castro et al (2017)

||> Insert DOAJ Seal of Approval images/doaj\_seal\_logo.png <||

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’.” (FN: DOAJ selection for Seal Approval is explained in the FAQ at <https://doaj.org/faq#seal>)

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. The following plot of DOAJ Seal Journals by Country indicates the problem.



Finally, our research group has determined that following is true when it comes to reputable open access journals.

According to Xie et al (2018):

Inline

||> Turn this into a code chunk {r equations, child = “equations-child.Rmd”} <||

## Researcher Information

R.P. Vivek-Ananth, Ph.D. student, The Institute of Mathematical Sciences (IMSc), Chennai

Researcher ID

ORCID: ||> Insert your ORCID number here <||

Date of First Version

Two months ago

Date of Last Update Today’s date

Related Policies

All original data, code, or reports produced as part of this project are owned by Your Institution per its institutional intellectual property policy.

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 RStudio, using the tidyverse package (need to cite Tidyverse somehow) 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 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. The complete data set is available in searchable and broweseable format in the [Annex](#annex-table) near the end of this document.

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 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).

‘Data Policy’ (Boolean) Yes No

‘Data Sharing Policy’ (Factor) No mention Implied Mentioned Explicitly encouraged Required, but not explicitly tied to editorial decisions Required as a condition of publication

‘Data Citation Policy’ (Factor) No mention Implied 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.

Data file formats and standards

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

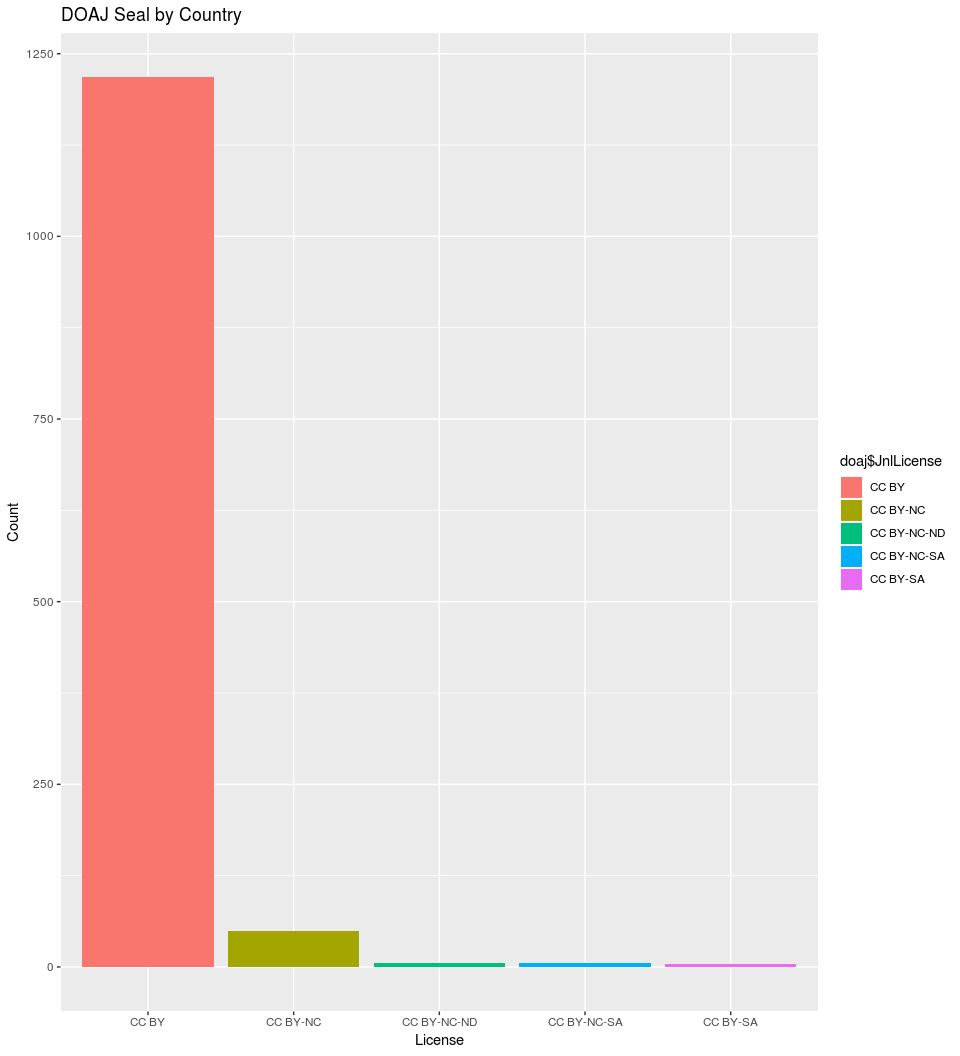
All data policies culled from the web sites of the DOAJ seal journals will be saved as .txt files. The data generated by applying Castro et al’s (2017) Coding Framework will be stored in csv file format.

Analysis, visualization, and summarization of the study’s findings will be performed in the open source software R and RStudio using the tidyverse package (cite tidyverse package). Reports produced from the study will be also be created in RStudio using the open source text format Rmarkdown (cite rmarkdown package) and output to HTML documents, slides, and MS Word documents for submission to funders or publishers (cite Xie et al, 2018, RMarkdown guide from CRC)

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.

### 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 |  |
| 2 | doaj\_seal\_enhanced.csv | enhanced data set with new data | CSV file, plain text format, 2.7 MB |  |
| 3 | Data set documentation | json metadata file | plain text file, .1 MB |  |
| 4 | Data Processing steps | R scripts and comments | R Notebook file, 1 MB |  |
| 5 | Data Visualizations | R scripts and documentation; Plots | R Notebook file, .png image files 4 MB |  |
| 6 | Journal article | Rendered report | RMarkdown, 9 MB |  |



# Documentation and 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 CC 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. (FN: <https://www.rd-alliance.org/rda-codata-legal-interoperability-research-data-principles-and-implementation-guidelines-now>

# 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 My 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

# Selection and long-term preservation

||> To be completed by the participants! <||

# 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. (FN: Zenodo policies are available online at <http://about.zenodo.org/policies/>

# 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 maintain all files under version control using git and GitHub.

The Data Management Specialist assigned to the project as an in-kind contribution from the My Institution 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.

# Annexes

#### Complete dataset doaj\_seal.csv

*This data table was compiled using the DT package by Xie (2018)*

||> Turn this into a code chunk

{r data-table} doaj\_seal %>% datatable(rownames = FALSE, colnames = c("Title", "Publisher", "Country", "Fees", "Waivers", "Identifiers", "Start Year", "Language(s)", "Review Process", "Plagiarism check", "Time to Press", "License", "Author owns Copyright", "DOAJ Seal"), class = "cell-border stripe", caption = "Journals with DOAJ Seal", filter = list(position = "bottom"), extensions = 'Buttons', options = list(dom = 'Bfrtip', buttons = c('colvis', 'csv', 'pdf')) )

<||

#### Principal Investigator’s BioSketch

This is auto-populated from your ORCID profile using the rorcid package. (cite the package).

||> insert code from insert\_orcid.R <||

# References

Castro et al, 2017, blahblahblah

Rmarkdown package citation of some kind, blahblahblah

Vasilevsky et al, 2017, blahblahblah

Tidyverse package citation of some kind, blahblahblah

Rorcid package citation of some kind, blahblahblah

Xie (2018), DT Table package, CRAN, blahblahblah

Xie et al (2018) Markdown book published by CRC