What you should know about data policies in economic and social science journals

Thomas Seyffertitz
Vienna University of Economics and Business
Library Center – Research Data Management

researchdata@wu.ac.at

https://doi.org/10.5281/zenodo.10391047

ENGAGE.EU Webinar December 15, 2023

Agenda

- Introduction & Objectives

- Background: Content analysis of journals' data policies

- Data policies of publishers and journals

- Formal features of data policies

- Content characteristics of data policies

- Conclusions & Recommendations

Introduction & Objectives

- Journal data policies – a working definition:

  - document (or part thereof) provided by a journal or its editor/owner that

  - contains statements on various aspects of how research data (and/or code) and information about the data should be made available (such as data availability, data provision, reproducibility, data sharing, data depositing, etc.)

  - underlying a submitted paper.

- Different levels of data policies

  - framework policies: publisher/research organization/associations; valid for specific group or all journals in journal portfolio (Springer Nature, AEA)

  - journal specific data policy (e.g. created by the editorial board)

Introduction & Objectives ctd.

- Formal and content features of data policies (DPs)

  - Formal characteristics relating to presentation, preparation and clarity.
  - Content characteristic relating to the scope, minimum requirements, comprehensibility and "strength" of the policies

- Objectives:

  - present some key features of DPs,
  - provide practical examples and
  - raise awareness for data policies of scholarly journals

Background – Content analysis of journals' data policies

- sample of 34 prestigious journals
- relevant publication outlets for the main research areas at WU
- emphasis on economics, business, management
- content analysis in February 2023
- new list to be released in 2024

Subjects*	No. of journals
Business/Management	12
Business/Finance	7
Economics	9
Economics/Math. Methods	2
Management/Info. Science	2
Science	2

Summary of the results in Novotny, G., & Seyffertitz, T. (2023). Research Data Policies in Scientific Journals – a Case Study. In V. Heuveline, N. Bisheh, & P. . Kling (Hrsg.), *E-Science-Tage 2023: Empower Your Research – Preserve Your Data* (S. 73-88). heiBOOKS. https://doi.org/10.11588/heibooks.1288.c18066

*) according to SSCI Core classification (Web of Science Master Journal List)

Framework policies of selected publishers

Publisher	Minimum of information required on data sets as stated by publisher/journal		
	(all features **encouraged**)		(all features **required**)
Wiley	• Data availability statement (=DAS)	………	• DAS • Peer review of data
Taylor & Francis	• DAS	………	• DAS • PID for data set • data citation
Elsevier	• Data deposit in a relevant data repository • Citing this dataset in the article	⟷	• data deposit • data citation & linking (or DAS) • Peer review of data prior to publication
Springer	• Data sharing encouraged • DAS required (new in 2023)	………	• Data sharing required • Peer review of data on request

Formal features of data policies

Category	Description
Location and clarity	• publisher's (journal) website • website of an association (e.g. AMA)/learned society or university
Naming	How are the data policies termed? **Examples** • "*Code Sharing Policy*" • "*Reporting standards and* …" • "*Data and Code Availability Policy*" • "*Guide for authors*"
Presentation	• separate data policy documents, or • part of author guidelines or other policy documents
Format	• plain HTML, pdf, Word etc.
Timing	When do data guidelines become effective for the author: • submission, acceptance or publication

Location of data policy – the example of IER

- „**missing link**": at the Wiley website there is no link to the policy at the journal owner's website auf die Policy auf Journal Website (University of Pennsylvania)

→ may be challenging

Clarity & presentation: Journal of Financial Economics

- Publisher: https://www.elsevier.com/researcher/author/submit-your-paper/sharing-and-promoting-your-article and

- https://www.sciencedirect.com/journal/journal-of-financial-economics/publish/guide-for-authors

- and at a separate journal's site:
  https://www.jfinec.com/news/data-and-code → JFE's „News" page leads to the entire policy:

different guidelines at different locations for the same journal (?!)

Effective April 3, 2023, all manuscripts that receive an "accept pending uploads" decision must follow the new data and code policy, which is identical to the old policy except for two additions. The data editor will review all code and data packets to ensure that the packets are complete. In those cases in which the actual data cannot be disclosed, a pseudo data set must be provided. Articles will not be accepted outright until the data editor approves them. Articles that have received an "accept pending uploads" decision before April 3, 2023, must comply with the old data and code policies. Articles that were initially submitted before July 1, 2021, are exempt.

Content characteristics of data policies I

Category	Description
Data availability statement (DAS)	Is a DAS required or not and what information must be provided? It confirms the presence or absence of shared data. It is used in publications to describe where the data directly underpinning the publication can be found and under what conditions they can be accessed(*)
Data repository	Type of repositories recommended or required; Example: QJE Dataverse repository for replication packages and raw data
Data citation	About the provision of the citation of underlying data in the submitted paper (i.e., DOI required)?

(*)see, for example https://library.bath.ac.uk/research-data/archiving-and-sharing/data-access-statements)

Examples: Data availability statement (DAS)

Availability of data	Templates for DAS
Data openly available in a public repository that issues datasets with DOIs	The data that support the findings of this study are openly available in [repository name e.g "figshare"] at http://doi.org/[doi], reference number [reference number].
Data derived from public domain resources	The data that support the findings of this study are available in [repository name] at [URL/DOI], reference number [reference number]. These data were derived from the following resources available in the public domain: [list resources and URLs]
Data subject to third party restrictions	The data that support the findings of this study are available from [third party]. Restrictions apply to the availability of these data, which were used under license for this study. Data are available [from the authors / at URL] with the permission of [third party].
Embargo on data due to commercial restrictions	The data that support the findings will be available in [repository name] at [URL / DOI link] following an embargo from the date of publication to allow for commercialization of research findings.

Source: Above examples taken from Wiley's Standard Templates for author use

Content characteristics of data policies II

Category	Description
Data sharing	Statements about data sharing in general, mandatory or only recommending
Pseudo-data sets	If original data can not be provided due to certain restrictions
Reproducibility/ Replication	Statements about reproducibility, its requirements, recommendations, and how-to descriptions Is it required? What must a reproducibility package contain?
Data editor	Their roles comprise i.a. securing the DP compliance, advising authors in several data related aspects and sometimes may verify for reproducibility of provided data

Data sharing

- Beispiel aus dem „Guide for Authors" des **Journal of Accounting and Economics** (Elsevier):

  > ***Research data***
  > This **journal encourages** and enables you to share data that supports your research publication where appropriate, and enables you to interlink the data with your published articles. **Research data refers to** the results of observations or experimentation that validate research findings. **To facilitate reproducibility** and data reuse, this journal also **encourages** you to share your software, code, models, algorithms, protocols, methods and other useful materials related to the project.

- Example from„Guide for Authors" **Accounting, Organizations and Society** (Elsevier):

  > This **journal requires** and enables you to share data that supports your research publication where appropriate, and enables you to interlink the data with your published articles. Research data refers ……….. project.

Pseudo-data sets

- Journal of Finance: "… Authors are also encouraged to include the data along with the source code if public posting of the data does not violate copyright or confidentiality agreements. If the authors choose not to provide the data, **they must include a pseudo-data set** that illustrates the format of the files read **by the code so that users can understand and check the functionality of the code**. …" See also their FAQs

- Journal of Financial Economics: "… For those cases in which the **data cannot be disclosed**, the authors **must supply pseudodataset(s)** to **demonstrate that the code runs**. It is the authors' responsibility to ensure that the code works on the pseudo-dataset or the actual dataset if the data can be disclosed. …".

- Review of Financial Studies: "… We **encourage** authors to include the data along with the source code. However, if the authors **choose** not to provide the data, or if they are **restricted** from doing so because of copyright or confidentiality agreements, they are **required to include a pseudo-data** set to illustrate the format of the files read by the code so that users can better understand the code. …".
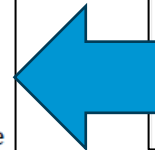
Example: Replication package

- Problem: licenced data not to be shared

- Solution: in-depth description (4pp.) how data can be derived

- contact data editor at submission (!)

Download Full Text PDF

Article Information

Abstract

We estimate the causal effect of sovereign default on the equity returns of Argentine firms. We identify this effect by exploiting changes in the probability of Argentine sovereign default induced by legal rulings in the case of NML Capital, Ltd. v. Republic of Argentina. We find that a 10 percent increase in the probability of default causes a 6 percent decline in the value of Argentine equities and a 1 percent depreciation of a measure of the exchange rate. We examine the channels through which a sovereign default may affect the economy.

Choose Format:

Additional Materials

Replication Package (72.84 MB)

Online Appendix (773.00 KB)

Author Disclosure Statement(s) (100.80 KB)

Data for Replication of Hebert and Schreger "The Costs of Sovereign Default: Evidence from Argentina"

This document outlines how to collect the replication data for the paper. Much of the data used in the paper is from commercial data vendors and cannot be shared. In this case, this document will outline how to acquire the data and a skeleton file will be posted so a reader can very the acquired data is in the correct format for the replication files to run. In the below instructions, this will be indicated as [**Raw Data Unavailable**].

When publicly available data is used, the raw file is posted along with information on where it was acquired. This case will be noted as [**Raw Data Available**].

Example: Replication package - Exemption

- Submission guidelines of *The Econometrics Journal* refer to the **replication policy**

- …We will only publish papers if they are accompanied by a complete replication package. …

- but you may request exemption !!

Covering letters

Authors should (only) include a cover letter if they wish to:

- disclose conflicts of interest
- request exemption from the replication policy
- request exemptions from other submission guidelines
- request that previous editors' letters and referee reports a

- **→ read guidelines in advance; answer to your request may need some time!**

Data editors: three examples

- **Review of Economic Studies**: "… The Review of Economic Studies endorses DCAS, the Data and Code Availability Standard [v1.0], and its data and code availability policy is compatible with DCAS. … Requests for **exemptions** should be clearly stated **when** the **article** is **first submitted**. The article will then be **reviewed** at the discretion of the Managing Editors and the **Data Editor**. Exceptions will not be considered later …."

- **Management Science**: "*…* The task of the Management Science **Data Editor** is to help authors to get their published research **compliant** with the Management Science **Data Policy**. … **advises authors of published papers** on the data, materials, and information to be provided to allow other researchers to replicate the original results. …"

- **American Economic Review and Journal of Economic Literature**: "… AEA **data editor** will **assess compliance** with this policy, and will **verify** the **accuracy** of the information **prior to acceptance** by the Editor. … private (not to be published) version of the data should be provided to the AEA Data Editor and/or a designated third-party replicator who can provide a third-party **reproducibility** report. … **After the data and code deposit is accepted** by the AEA Data Editor, it will become the version of record associated with the paper. … Data Editor will **assess suitability** of any such **repositories and archives**. …"

Conclusions & Recommendations

➢ Authors may not be aware of a journal's specific data sharing requirements; requiring authors to deposit the underlying data in a particular repository, which, if overlooked, could result in complications during the publication process.

➢ Neglecting data guidelines could increase the likelihood of submitted paper being rejected by leading journals.

➢ Lack of awareness about data policies can lead to significant time and effort wasted in revisions and resubmissions and delay publication.

Conclusions & Recommendations ctd.

- **Start as early as possible** to carefully read the data guidelines of the journal in which you intend to publish; at least **before submitting** the paper, but better at the **beginning of the research project**

- **Research community**: Participate in academic discussions, conferences, or forums related to data-sharing and reproducibility.

- **Journal Websites**: Explore the websites of specific journals, especially those known for stringent data-sharing policies.
  - Some journals provide information on their editorial decisions, including instances where papers were rejected due to data-related issues (or see, e.g. https://retractionwatch.com/)

- **Contact the editor** or data editor prior to submission, e.g. if not sure whether some policy features may apply to your data

In-depth Studies

Authors	Disciplines/Subjects (sample size)	Scope
Andreoli-Versbach & Mueller-Langer (2014)	economics (147)	data sharing policies
Vlaeminck & Herrmann (2015)	economics (346)	characteristics of journals with data policies
Herndon & O'Reilly (2016)	social-sciences (100)	data sharing policies comparison (2003/2015)
Crosas et al. (2018)	social-sciences (291)	data policies and instructions
Rousi & Laakso (2020)	operation research (40)	data sharing policies (2019)

Thank you very much!
Questions & Discussion