

Forschungsdaten-Richtlinien in wissenschaftlichen Fachzeitschriften



Thomas Seyffertitz & Gertraud Novotny
Wirtschaftsuniversität Wien

researchdata@wu.ac.at

Oktober 17, 2024

STAND: JUNE 2024

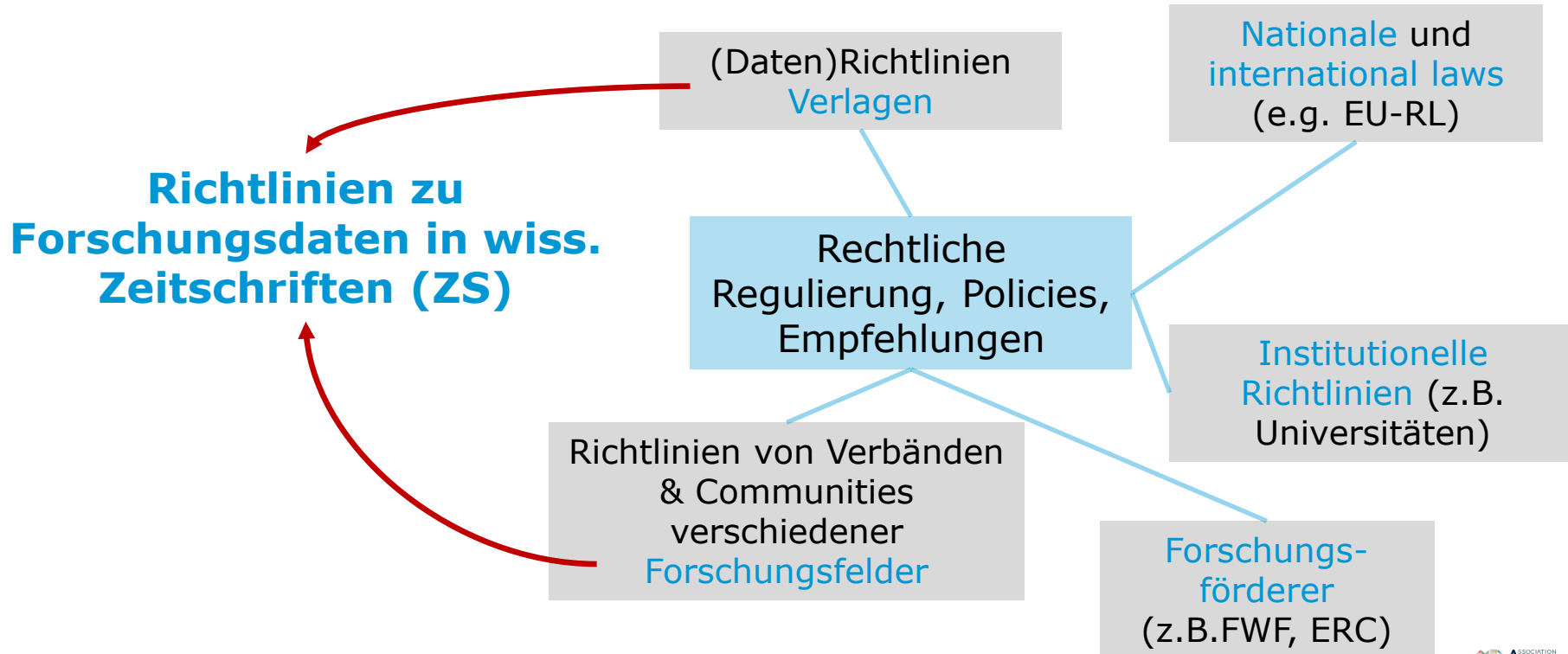


- Einleitung
- Hintergrund
 - Rechtlicher Bezugsrahmen
 - Verlags-Policies
- Fallstudie: Datenrichtlinien der WU Star ZS-Liste
 - Formale Elemente (+Beispiele)
 - Inhaltliche Elemente (+Beispiele)
- Exkurs: KI-Nutzung
- Schlussfolgerung & Empfehlungen

- **Datenrichtlinien in wiss. Zeitschriften – Arbeitsdefinition**
 - **Eigenständiges** Dokument oder **Teil eines** größeren **Leitfadens**
 - **Inhalt:** Angaben, Anweisungen und Empfehlungen zum Umgang mit Forschungsdaten/Code, auf denen die eingereichte Arbeit basiert.
 - **Ursprung** der Richtlinie
 - **Verlag** (z.B. Springer Nature), für bestimmte oder alle ZS im Portfolio
 - Forschungsorganisation/Verband: z.B. AEA, AERA, INFORMS, APS
 - **ZS-spezifisch** (z.B. durch Herausgeberboard)

Synonyme Verwendung der Begriffe:
(Forschungs)Datenpolicy und (Forschungs-) datenrichtlinie

Hintergrund - Rechtlicher Bezugsrahmen



Hintergrund – Merkmale und Beispiele von Verlags- Policies

14 journal research data policy features arranged as 6 policy types (tiers)

	Policy 01	Policy 02	Policy 03	Policy 04	Policy 05	Policy 06
Definition of the research data	○	○	○	○	○	○
Exceptions to policy	○	○	○	○	●	●
Embargoes	○	○	○	●	●	●
Supplementary materials	○	○	○	●	●	●
Data repositories	○	○	○	●	●	●
Data citation	○	○	○	○	●	●
Data licensing	○	○	○	○	○	○
Researcher/ author support	●	●	●	●	●	●
Data availability statements		○	●	●	●	●
Data formats and standards				○	○	●
Mandatory data sharing (specific data types)				●	●	●
Mandatory data sharing (all papers)				○	●	●
Peer review of data				○	○	●
Data Management Plans (DMPs)				○	○	○

Source: Hrynaszkiewicz I, et al. 2020. Developing a Research Data Policy Framework for All Journals and Publishers. *Data Science Journal*, 19: 5, pp. 1–15. DOI: <https://doi.org/10.5334/dsj-2020-005>

Publisher	Minimum of information required on data sets as stated by publisher/journal	
	(all features encouraged)	(all features required)
Wiley	<ul style="list-style-type: none"> Data availability statement (=DAS) 	<ul style="list-style-type: none"> DAS Peer review of data
Taylor & Francis	<ul style="list-style-type: none"> DAS 	<ul style="list-style-type: none"> DAS PID for data set data citation
Elsevier	<ul style="list-style-type: none"> Data deposit in a relevant data repository Citing this dataset in the article 	<ul style="list-style-type: none"> data deposit data citation & linking (or DAS) Peer review of data prior to publication
Springer	<ul style="list-style-type: none"> Data sharing encouraged DAS required (new in 2023) 	<ul style="list-style-type: none"> Data sharing required Peer review of data on request

Fallstudie: Datenrichtlinien in den WU Star Journals

WU Star-Journal-Liste

- Anreiz zum Publizieren in Top-Journalen in den relevanten Forschungsgebieten (Prämie)
- basiert auf Auswahlverfahren aus internationalen Rankings
- mehr als 270 Publikationen von WU-Forscher/innen von 2017-2024 in diesen Zeitschriften
- → Liste von **84 Fachzeitschriften**
- → **111 Dokumente** mit Informationen zu Forschungsdaten

Forschungsfeld	Anzahl der ZS
Economics	17
Business	13
Management	12
Psychology	9
Environmental Research	7
Political Science	7
Math./Stat./Comp.Sci.	6
Educational Research	4
Information Systems	3
Multidisciplinary Science	3
Sociology	2
Communication	1

Wir werfen einen Blick auf ausgewählte Merkmale von Datenrichtlinien (mit praktischen Beispielen)

- **Formale Elemente** in Bezug auf Präsentation, Aufbereitung und Klarheit
- **Inhaltliche Elemente** der Datenrichtlinien für den Umgang mit Forschungsdaten
- zeigen Beispiele aus verschiedenen Zeitschriften

Datenrichtlinien – Formale Elemente

Überblick

Kategorie	Beschreibung
Verortung	<ul style="list-style-type: none">○ Webseite des Verlages (Zeitschrift)○ Verbandswebseite (e.g. AMA) oder Universität
Benennung	<p>Wie werden die RL benannt? Beispiele:</p> <ul style="list-style-type: none">○ "Code Sharing Policy"○ "Reporting standards and ..."○ "Data and Code Availability Policy"○ "Guide for authors" ..., etc.
Präsentation, Format, Klarheit	<ul style="list-style-type: none">○ Separates Dokument, oder○ Integriert in Autoren-/Einreichungsrichtlinien (meist HTML oder pdf)
Zeitpunkt	<p>Wann werden Datenrichtlinien (oder Teile) für Autoren schlagend: Einreichung, Akzeptanz oder Publikation?</p>

Datenrichtlinien – formale Elemente: Nicht eindeutige Verortung

- „**missing link**“: auf der [Wiley website](#) kein Link zur Policy auf der [Webseite des ZS-Eigentümers](#) (University of Pennsylvania)

The screenshot displays the website of the International Economic Review (IER), published by Penn Arts & Sciences. The top navigation bar includes links for HOME, ABOUT, CONTRIBUTE, and BROWSE. The CONTRIBUTE dropdown menu is open, showing options: Author Guidelines, Open Access, Submit a Manuscript, and For Referees. A red arrow points from the 'Open Access' option to the 'Data Availability Policy' link in the right-hand column. The right-hand column features a section titled 'Data Availability Policy' with the text: 'EFFECTIVE January 1, 2022. To promote consistent standards and requirements among general reviewers, the IER has adopted the AER data availability policy. We are, therefore, requiring authors to provide data availability statements for all empirical papers. The policy is effective for all papers accepted for publication on or after January 1, 2022. For more information, see the IER Data Availability Policy page.' Below this, a link for 'Data Availability Policy' is provided. The bottom of the page shows a 'Most Recent' and 'Most Cited' section.

→ schwierig zu finden, wenn keine Vorinformation

Datenrichtlinien – formale Elemente: Beispiel zur Darstellung und „Klarheit“

- **Verlagsempfehlung:** <https://www.elsevier.com/researcher/author/submit-your-paper/sharing-and-promoting-your-article> **und**
- [Forschungsdaten-RL dieser ZS auf Elsevier's Webseite](#)
- und **eine separate Zeitschriften-Webseite:** <https://www.jfinec.com/news/data-and-code> → JFE's „News“ **führt zur gesamten RL:**

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.

Achtung:
verschiedene RL an
verschiedenen
Stellen für dieselbe
ZS !!

Datenrichtlinien – inhaltliche Elemente I

Kategorie	Beschreibung
Datenverfügbarkeits- erklärung (DAS)	<p>Erwähnt oder vorgeschrieben? Templates?</p> <p>In Publikationen verwendet, um zu beschreiben, wo die der Veröffentlichung zugrunde liegenden Daten zu finden sind und unter welchen Bedingungen auf sie zugegriffen werden kann(*)</p>
Datenrepositorium	<p>Informationen über die Anforderungen an die Datenarchivierung und Hinweise auf empfohlene Repositorien</p> <p><i>Beispiel:</i> <u>QJE Dataverse repository for replication packages and raw data</u></p>

Datenrichtlinien – inhaltliche Elemente II

Kategorie	Beschreibung
Teilen von Daten & Code ("sharing")	Angaben zum Teilen/Nachnutzbarkeit von Daten/Codes im allgemeinen, z.B. obligatorisch oder nur empfehend
Pseudo-Daten/Synthetische Daten	Keine Weitergabe der Originaldaten aufgrund bestimmter Einschränkungen (Vertraulichkeit, kommerzielle Daten)
<u>Reproduzierbarkeit</u> <u>Replikation</u>	<p>Aussagen über die Reproduzierbarkeit, Anforderungen, Empfehlungen und Anleitungen (Replikationspaket)</p> <p>Ist sie zwingend und wann? Was muss ein Replikationspaket enthalten?</p>
Daten Editor	Gewährleistet Einhaltung der Datenpolitik, Beratung der Autoren in datenbezogenen Aspekten; zuständig für Fragen der Reproduzierbarkeit

Inhaltselement: Datenverfügbarkeit

Beispiele

Datenverfügbarkeit	Vorlagen für Datenverfügbarkeitserklärungen
Offen zugängliche Daten in Repositorien (Datensätze mit 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].
Daten aus öffentlich zugänglichen Quellen	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]
Daten unterliegen Einschränkungen durch Dritte	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 aufgrund kommerzieller Beschränkungen	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.

Inhaltselement: Datenrepositorium

- Journal of Applied Psychology (APA):
 - "Authors must state whether data, code, and study materials are posted to a trusted repository and, if so, how to access ... and any limitations on use. If they cannot ... available, authors must state the legal or ethical reasons why they are not available. Recommended repositories include [APA's repository](#) on the Open Science Framework ([OSF](#)), or authors can access a [full list of other recommended repositories](#)."
- Science Journals (Amer. Ass. for the Advancement of Science):
 - ... In cases where datasets do not yet fall under the purview of a dedicated [subject repository](#), they should be archived in a [permanent, independent repository](#) such as [Dryad](#), [Dataverse](#), or [Zenodo](#). Sites such as Dropbox that give authors discretion to alter or remove files are not acceptable. ..."
- Beispiel: [Springer empfiehlt](#) Repositorien für verschiedene Forschungsfelder
- American Journal of Political Science:
 - „... All verification files [must be stored](#) in a Dataset within the [AJPS Dataverse](#) authors also can make their verification files available elsewhere ... as long as all of the necessary files are included in the Dataset on the [AJPS Dataverse](#)...."

Zeitschriften
können
Repositorien
empfehlen -
oder ein eigenes
obligatorisches
Dataverse
anbieten!

Inhaltselement: Angaben zum Teilen von Daten & Code

- **Beispiel 1** „Guide for Authors“ des [Journal of Accounting and Economics](#)

... 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. To **facilitate reproducibility** and data reuse, this journal also **encourages** you to share your software, code, models, algorithms, protocols, methods and

- **Beispiel 2** „Guide for Authors“ [Accounting, Organizations and Society](#)

... 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. ...

Term	Vorkommen*
data sharing	36
code sharing	8
share data	33
share code	2
sharing policy	12

*) Term kommt mind. einmal in n Policy-Dokumenten vor ($N=111$)

- Journal of Finance: "... Authors ... encouraged to include the data... If 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**. ..."
- Journal of Financial Economics: "... data cannot be **disclosed**, the authors **must supply pseudodataset(s)** to **demonstrate that the code runs**. ..."
- Review of Financial Studies: "... **encourage** authors to include the data ... source code. ... authors **choose** not to provide ..., or ... **restricted** ... because of copyright or confidentiality ..., ... **required** to **include a pseudo-data** ...".

Verursacht u. U.
erheblichen
Mehraufwand für
Forschenden

Also, the term “synthetic data” may occur, in different contexts:

- AERA (Educational researcher):
 - “... If in order to use certain forms of data or information consonant with guarantees of **confidentiality**, specific techniques have been used to mask or perturb the data **or generate synthetic data from the original data**, these processes should be noted. (**confidentiality issues**)
- Marketing Science – INFORMS Replication and Disclosure Policy:
 - “...data disclosure in such cases **would prevent the authors from using their data set beyond the first paper** and thus serve as a disincentive for such data collection efforts...”(**data collection incentive**)
 - “... Our policy preserves the **incentives for ambitious data collection** projects by allowing the authors to **meet the data requirements** by providing either **a synthetic data set** developed from the estimated models or a **slice of the data** set containing a randomly drawn subset of observations....”

Inhaltselement: Reproduzierbarkeit/Replikation

Beispiel für den Fall geschützter Daten

- **Problem:** lizenzierte/proprietäre Daten, die nicht weitergegeben werden dürfen
- **Lösung:** Detailbeschreibung, wie Daten hergeleitet werden können
- **Erfordert** Kontakt mit (data) editor bei oder vor Einreichung(!)

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 verify 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].

VOL. 107, NO. 10, OCTOBER 2017
(pp. 3119-45)

[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.

Citation

Hébert, Benjamin, and Jesse Schreger. 2017. "The Costs of Sovereign Default: Evidence from Argentina." *American Economic Review*, 107 (10): 3119-45.

DOI: 10.1257/aer.20151667

Choose Format:



Additional Materials

[Replication Package \(72.84 MB\)](#)

[Online Appendix \(773.00 KB\)](#)

[Author Disclosure Statement\(s\) \(100.80 KB\)](#)

Inhaltselement: Beispiele für Replikations-Vorlagen und Werkzeuge

Vorlage, Checklisten, Readme

- Template README für Replikation in Sozialwiss.*
- *Management Science* Journal: Code and Data Disclosure Policy + Replication Package checklist
- American Journal of Political Science: Infos zu Replikationsdateien
- *Beispiel*: Aufbereitung des Datenmaterials für Abgabe an ein Repository (dataverse)

Werkzeuge & Portale

- Code ocean: computational science platform
- Penn Wharton Credibility Lab:
 - ResearchBox - platform helping researchers to share research files (pre-registration, materials, data, code ...)
 - Groundhog is an R package (statistical software) that makes it easy for researchers to write R scripts that will be reproducible
- Austrian Social Science Data Archive (AUSSDA) – Hostet auch Replikationsdaten

*) see <https://zenodo.org/records/7293838> for latest version

Inhaltselement: Data Editor – Beispiele

- **Review of Economic Studies**: "... 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**: "... task of ... **Data Editor** is to help authors to get their published research **compliant** with the ... **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**: "... **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. ..."

Exkurs: Richtlinien in Zeitschriften zur Nutzung generativer KI ?

- regulieren Anwendung/Nutzung von KI-Werkzeugen bezüglich **Verfassens eines Manuskripts** (“... *generative editorial work and autonomous content creation.*”)
- **KI** (z.B.) **nicht** als **Autor** zulässig
- Quellen vorlegen, die zur Erstellung von Inhalten/Zitaten verwendet wurden, die von LLM generiert wurden.
- Hinweis auf die Verwendung des **Sprachmodells** (welches LLM, Zweck)
- **Beispiele:**
 - [Guide for authors - Global Environmental Change](#) (Elsevier)
 - [Association for Psychological Science - Editorial Policies](#) (APS)
 - Generative [AI Policy](#) (Springer)
 - [ChatGPT and Generative AI](#) (SAGE)

- **Umfassendste** Richtlinien von **Forschungs-Communities**: z.B. AEA, AERA, INFORMS, APS
- Verschiedene **Disziplinen** → **unterschiedliche Prioritäten** in den Policies
- Kommerzielle Verlage haben teilweise **Rahmenrichtlinien** – *Beispiel*:
 - Verweis in ZS auf Verlags-Policy bezüglich Datenverfügbarkeit: Review of Finance von OUP: Level1- bis Level4)
- **Datenverfügbarkeitserklärung** quasi-Standard in Datenrichtlinien
- **Pseudo/synthetische** Daten erfordern zusätzlichen Aufwand
- **Mangelnde Kenntnis** der betreffenden Datenrichtlinie kann zu erheblichem Zeit- und Arbeitsaufwand führen
 - Revision, Verzögerungen beim Erscheinen oder Gefahr einer Ablehnung

- **Datenrichtlinien** frühestmöglich, aber spätestens vor der Einreichung der Arbeit lesen!
- Kontaktieren Sie die **FDM-Beratung** ihrer Institution (!)
- **Community**: Fragen Sie Kollegen, Teilnahme am akademischen Diskurs, Konferenzen; Portale/Foren zum Thema Datenrichtlinien/Reproduzierbarkeit
- **Zeitschriften**: Erkunden Sie die Websites von ZS, insbesondere von Top-Journals, die i.d.R. „strenge“ Datenrichtlinien haben.
 - Einige Zeitschriften stellen Informationen über ihre redaktionellen Entscheidungen zur Verfügung, einschließlich der Fälle, in denen Beiträge aufgrund von Datenproblemen abgelehnt wurden (siehe z.B. <https://retractionwatch.com/>)
- Machen Sie sich mit (relevanten) **Werkzeugen** vertraut (z.B. code checking)
- **Kontaktiere Dateneditor/Editor**, bei Unsicherheit, welche Teile der Richtlinie zutreffen könnten.

Danke!

Q&A