

Dawning of a New Age? Economics Journals' Data Policies on the Test Bench

Sven Vlaeminck

ZBW Leibniz-Information Centre for Economics, Hamburg, Germany
s.vlaeminck@zbw.eu, orcid.org/0000-0002-7905-4209

Abstract

In the field of social sciences and particularly in economics, studies have frequently reported a lack of reproducibility of published research. Most often, this is due to the unavailability of data reproducing the findings of a study. However, over the past years, debates on open science practices and reproducible research have become stronger and louder among research funders, learned societies, and research organisations. Many of these have started to implement data policies to overcome these shortcomings. Against this background, the article asks if there have been changes in the way economics journals handle data and other materials that are crucial to reproduce the findings of empirical articles. For this purpose, all journals listed in the Clarivate Analytics Journal Citation Reports edition for economics have been evaluated for policies on the disclosure of research data. The article describes the characteristics of these data policies and explicates their requirements. Moreover, it compares the current findings with the situation some years ago. The results show significant changes in the way journals handle data in the publication process. Research libraries can use the findings of this study for their advisory activities to best support researchers in submitting and providing data as required by journals.

Keywords: data policies; journals; economics; open science; reproducibility

1. Introduction

Journals are on the forefront of the scientific ecosystem. Because of their peer-review processes, they are an important instance to ensure scientific quality and integrity. Against the background of an ever growing amount of publications¹ (Johnson et al., 2018) and public debates on 'fake journals' and 'predatory publishing' (cf. Hern & Duncan, 2018), quality assurance mechanisms like the peer-review process should ensure that only well founded research is published in the pages of scholarly journals. Research libraries have also developed a variety of services to help researchers find appropriate journals to publish their findings. Peer review and adherence to good scientific practice are key requirements for selecting a suitable journal.

Nevertheless, there are serious reservations about the role of peer-review procedures when it comes to the inclusion of data and calculations in the publication process. Specifically, discussions about useful editorial procedures in dealing with empirical or other data-driven contributions have been raised frequently in the last decades. Since the seminal paper of Dewald et al. (1986), in which the authors systematically attempted to reproduce the results of published papers, the editorial procedures of journals in dealing with data have been under fire. The study of Dewald et al. suggested that errors in published empirical articles are "a commonplace rather than a rare occurrence" (Dewald et al., 1986, p. 587f). These findings have been widely regarded as a serious issue. However, twenty years later, the US-economist McCullough still noted: "Results published in economic journals are accepted at face value and rarely subjected to the independent verification that is the cornerstone of the scientific method. Most results published in economics journals cannot be subjected to verification, even in principle, because authors typically are not required to make their data and code available for verification" (McCullough et al., 2006, p. 1093f).

To assess the results of empirical publications, reviewers and would-be replicators need the data and some additional information on the methodology used. Also the different steps of the analysis or instructions given in economic experiments are crucial to assess the robustness of research findings. Reproducibility of research is a key pillar of the scientific method. Without the ability to check the findings of applied economic research for robustness or potential methodological errors, publications do not meet a basic requirement for scientific discoveries.

In the light of intensified debates on open science, this paper provides an updated analysis of the editorial policies of economics journals with respect to data. The paper asks how journals in economics today deal with the question of including data and calculations in the publication process. Specifically, the paper asks if there are editorial policies that request these files.

The paper addresses three dimensions: First, the article asks how many journals listed in the Clarivate Analytics' Journal Citation Reports 2017 edition for economics (in the following abbreviated to JCR ECON 2017) have a data policy and classifies the different types of policies found in the sample. Second, the characteristics of these data policies are examined including an analysis of the materials requested from authors, the share of mandatory and voluntary data policies, how propriety and confidential data is handled, and a description of journals' recommendations for publishing research data and replication files. In addition, the paper shows how the requirements of the data policies differ from publishing house to publishing house and characterises these differences. Third, the article compares the findings of this study with a similar survey published in 2015. The goal is to determine potential changes in the number of journals with a data policy and their demands since that point in time.

The paper starts with an outline of previous studies' findings on journal data policies in economics and describes the current state of research. From there, it discusses reasons for which journals hesitate to implement data policies and why authors are reluctant to comply with these policies. Following this theoretical classification, the paper will clarify the methodology of the study and the data collection, which is based on a content analysis. It then presents the outcome of the survey and the specifications of journal data policies before comparing the outcome of this paper with the findings of a survey published in 2015. Finally, it presents summaries and discusses the outcome of the study with respect to the daily work of research libraries.

2. Literature Review

Studies have been dealing with questions of reproducible research and particularly with the data policies and data archives of economics journals since the late 1980s. The publication of Dewald et al. (1986), set a starting point

to the ongoing debate. Their paper presented the findings of a project in which the authors collected programs and data from authors of the *Journal of Money, Credit and Banking* (JMCB). For the project, the JMCB adopted a data policy in which they bound authors to making the programs and data available on request. In one of the author's previous papers, this type of policy is labelled as an 'author responsibility policy' (abbreviated to ARP) as it leaves the responsibility to provide the replication files to other researchers with the original authors (Vlaeminck & Herrmann, 2015a). In their study, Dewald et al. requested the replication files of published papers from the authors of JMCB. They reported a response rate of 67% (within an average response time of 217 days) for authors who already had their papers published. Of these respondents, 48% could not (or did not want to) provide their data and program codes (e.g. the computer programs to run an experiment and to analyse the data). McCullough and Vinod (2003) experienced similar problems, when trying to replicate the findings of a full issue of the *American Economic Review* (AER), which also had an ARP at that time: Half of the authors did not honour the journal's data policy. Also more recent studies report ongoing problems with data policies that rely on author's support: Savage and Vickers (2009) found that only one in ten researchers contacted provided their data. The studies of Krawczyk and Reuben (2012) and Stodden et al. (2018) reached results comparable with Dewald et al.'s study from 1986.

These studies demonstrate a low response ratio by researchers for requests for data and program code. ARPs, like these, do not work due to a lack of incentive to share data with others. Feigenbaum and Levy (1993) have theoretically outlined why there are very limited incentives for authors to comply with such policies. Duvendack et al. (2015) argue that the costs to compile data and code into an easily reusable form are high, while authors receive small or no credit for this time-consuming work. This time could be spent more "productively" on other tasks that offer more benefits in terms of one's academic career. In addition, data sharing also involves the risk of damaging one's own scientific reputation should data or program codes contain errors.

These arguments illustrate why researchers are reluctant to share their data and why parts of the research community delay or even prevent a replication of their research (cf. Krawczyk & Reuben, 2012; Savage & Vickers, 2009).

However, also for journals, the incentives to implement data policies are ambivalent. Editors might fear that a data policy encourages authors to look

for alternative journals, which do not have a data policy. In the event the data contains errors, editors might fear that sharing data and program code can cause similar reputational consequences to journals as it can to authors. Journals can fortify themselves against these by including data and program code review in the review process. However, this has high costs to editors and reviewers: additional communication and workflows need to be set up. For reviewers, going over the data might be more time consuming and challenging than reviewing the paper itself. Consequently, the inclusion of data policies can also have serious obstacles for journals.

Nevertheless, as a reaction and consequence of the malfunctions reported with ARPs, journals have slowly begun to implement data availability policies (DAP). Among the first journals to implement a mandatory DAP was the *American Economic Review* (AER) (Bernanke, 2004). The crucial difference compared to ARPs is that authors have to submit their data and program code prior to publication of an article to the editorial office or, nowadays, to a recognised data repository. Accountability for publishing or providing data to would-be replicators is thus shifted from authors to editorial offices or third parties like data repositories. This policy change is a big step forward.

Little by little, other (top-)journals followed the lead of the AER (McCullough et al., 2006). But the number of journals with a solid data policy remained marginal compared to the overall number of journals in economics: McCullough (2009) mentioned 17 journals which implemented mandatory data archives for data and program code since 1993. All of these journals also have data policies in which they inform authors about their requirements.

Since that time, several studies have analysed journal data policies and how their demands have changed over time. While there are many publications about journal data policies in different fields of research (most often in the sciences), there are just a few analyses for economic journals: Vlaeminck (2013) analysed a sample of 141 economics journals and found 29 journals (20.6%) with a DAP, while 11 (7.8%) journals had an ARP. Almost 83% of the DAP were mandatory. The study also includes a discussion of the specific demands of journal data policies. Vlaeminck and Herrmann (2015a) analysed a sample of 346 journals from economics and management for the availability of data policies. They found 49 journals with a DAP (14.2%), another 22 (6.4%) employed an ARP. 61.2% of those DAPs were mandatory. Again, the requirements of the data policies are described in the paper.

Höffler (2017) analysed all economics journals listed in the Journal Citation Reports (JCR) for the year 2015. Of the 343 journals in the study, 26 had a mandatory DAP (7.5%), while 110 had voluntary policies (32.1%). Another 15 (4.4%) employed an ARP and 34 (9.9%) offered both voluntary data deposit and making data available upon request. One of the most striking results of Höffler's study is that a majority of journals in economics holds a policy on data – for the first time, since studies have focused on economic journals. A study by Chin and Dong (2019) generally confirmed the findings of Höffler. They analysed the data policies of 74 journals listed in the Tilburg University Top 100 Worldwide Economics Schools Research Ranking. They found 58 journals with a data availability policy (75.7%), while the remaining 18 journals had no data policy (24.3%). Of those 58 journals, 34 had a mandatory policy (58.6%).

Chang and Li (2015) and Vlaeminck and Podkrajac (2017) have examined how voluntary and mandatory data policies perform. The authors concluded that mandatory data policies perform better, because the probability to find the data necessary to conduct reproductions is higher than for journals with voluntary data policies.

3. Methodology & Data

In order to determine the share of economics journals with a data policy and to illustrate the demands of these policies, this paper will use the methodological approach of a content analysis. According to Neuendorf, "content analysis is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use" (Neuendorf, 2002, p. 18). Neuendorf splits the framework of a content analysis into nine conceptual steps. The first steps -theory and rationale, conceptualisations, operationalisations, creation of a coding scheme, and sampling- deal with the development of the research instrument. The subsequent steps - training and pilot reliability, coding, final reliability, tabulation, and reporting - deal with the collection and analysis of the data.

By using a content analysis, a structured, systematic coding scheme is applied to selected text passages. As a result, latent and manifest conclusions can be drawn about the frequency of certain topics, concepts or meanings. Central to the method is the selection of the text passages to be analysed, a precise

definition of the variables for the analysis, their accurate operationalisation in a codebook, and the careful application of the coding in data acquisition.

While theory and research questions have already been described previously, the construction of the research instrument and the sampling need some explanation. To determine the share of journals with data policies in the most relevant journals in economics, all journals listed in the category 'economics' of the 2017 edition of Clarivate Analytics' Journal Citation Reports (Clarivate Analytics, 2018) – in the following abbreviated to JCR ECON 2017 – have been examined. In total, the JCR ECON 2017 itemises 353 journals in this category.² The publishing houses, the impact factor, and the ranking of the journals within the category 'economics' of the JCR ECON 2017 have been added to the data. Subsequently, the journals have been categorised according to their methodological orientation.

Particularly, it has been of interest whether a journal generally accepts and publishes empirical, applied, or other data-driven contributions (like experiments, simulations or other forms of computational economics). To get this information, the sections 'aims and scope', 'about' and the general introductory text passage of a journal have been checked. In cases of doubt, up to four published issues have been evaluated manually to identify potential empirical papers. Only those journals that publish contributions based on data remain in the sample for further analyses. Journals that do not accept or publish data-driven contributions were no longer regarded as these journals do not need a data policy.

Subsequently, the webpages of the remaining journals have been examined for instructions that regulate the handling of data and other materials essential to reproduce the findings of an article. Typically, this information is part of the 'information for authors' section. In a few cases, this information is also available in the rubric 'duties of author' (or similar named paragraphs). Partially, journals also have a separate section on their webpage which links to their data policy. For the content analysis, these instructions are regarded as sampling units.

If available, not only has the publisher webpage of a journal been searched for this kind of information, but also the website of the editorial office. In the past, the websites of the editorial offices have offered better and more exact information than the publisher's website (cf. Vlaeminck, 2013).³ Specifically

we looked for recommendations to submit specific files and research instruments that are crucial to reproduce the results of a paper.⁴ These recommendations were treated as the coding unit for the content analysis.

All textual information with respect to handling data in the publication process found on journals' websites has been collected in a single document. Policies with an identical wording (or only slight deviations) were grouped to facilitate and accelerate the evaluation. The document served as the base for the content analysis (cf. Appendix B).

The variables needed for the operationalisation in the context of the content analysis have been derived from the literature. McCullough (2007) and the American Economic Association (2005) outline these. Glandon (2011) has shown that the recommendations of the American Economic Association (2005) work in practice. To have a maximum of comparability, the study only incorporates files and variables, which are used by most methodological approaches (e.g. econometrics, simulations, and experiments) in economics.⁵ Thus, they can be regarded as the 'lowest common denominator' for the methods used in economics research.

The variables include:

- the **dataset(s)** used to reach the findings of a paper. Without the data, empirical findings cannot be checked. Therefore, availability of data is essential for reproductions.
- the **program code** (e.g. of statistical analyses in econometric papers, experiments or simulations). Without having the program code of a statistical analysis, the results of a paper and their robustness cannot be assessed satisfactorily. Discussions about methodological choices or details of computations are not possible.
- **descriptions of the data** (e.g. data dictionary, codebooks, documentations) and/or of the entire **research process** (e.g. in readme-file and/or by including the instructions of experiments). Without descriptions of the data, it often is very difficult to make use of the data. In addition, without proper descriptions of the research process, it often remains unclear which steps have been taken to achieve the results of the paper.
- **Intermediate datasets** may help to understand the course and interstation of the research process.

The more of this information a data policy requires, the more robust it is in terms of reproducibility.

Beyond the policies' recommendations for data and other materials, the study also explored some other characteristics of the data policies, which serve as additional variables in the content analysis:

- the **degree of obligation** of a policy (mandatory or voluntary). As discussed, the literature suggests that a voluntary data deposit does not work well in practice. For this reason, the share of mandatory and voluntary data policies was of interest to this study.
- the way journals suggest to **provide the data** for replication purposes and the public (e.g. data repository, website).
- In economics (and specifically in business administration), the use of data from commercial providers is widespread. To increase the reproducibility of research even in cases where restricted (e.g. proprietary or confidential) data was used, journals need a **procedure** that regulates these cases. The goal is to permit reproductions in principle, even if the original data cannot be shared for legal reasons.⁶ Such a procedure might include providing an identifier of the data used, a contact address from where to obtain the data, information on the availability/access conditions of data, and the program code of the analysis.
- With a **data statement**, authors can be transparent about the data they used in their article and indicate its availability. Furthermore they can provide a reason if data is not available to access for others. For this purpose, many of the major publishers offer templates to choose from. By adding such a data statement, the access conditions of the data become transparent and researchers interested in replications can quickly assess the availability of the data.

Each variable was dichotomously classified as “mentioned/not mentioned” in the data policy.⁷ The subsequent coding was done manually by going through all the policies and marking all places in the text in which information on the variables was found. The coding schema experienced two rounds of adaptations after a short pre-test with a limited number of data policies. The final data collection took several weeks. To achieve inter-coder reliability to the greatest possible extent (despite there being only one coder) the coding process was performed twice: After a first run in March 2019, a second pass

of the categorisation process took place in June 2019. The findings slightly differed and resulted in a third pass in which only these ambiguous findings were double checked. Afterwards, the evaluation started.

In order to compare the results of this study with the findings of a previously published survey (see section 5), a dataset compiled by Vlaeminck and Herrmann (2015b), was used. This dataset contained information on journal data policies and their specifications for a sample of 346 journals. 262 of these journals also had an impact factor and almost all were listed in the Social Sciences Citation Index (SSCI).⁸ Those journals serve as a useful comparison group for the findings achieved in this study.

4. Findings of the Study

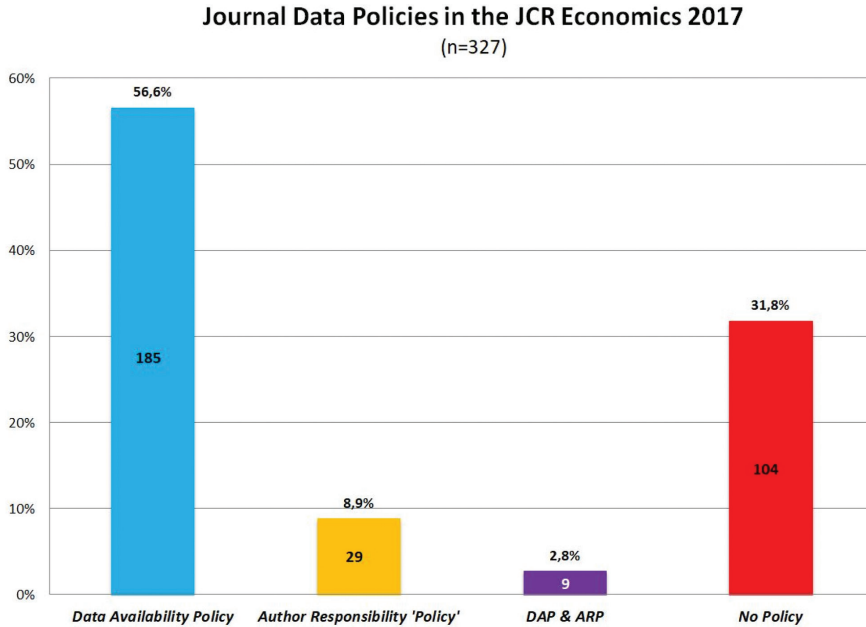
Based on the approach described above this work found that 327 out of 353 journals (92.6%) of the JCR ECON 17 generally accept or at least sporadically publish empirical and other data-driven contributions. The percentage of these 'empirically-oriented' journals is probably higher than the average of all journals in economics. But as the most prestigious journals often like to publish new or 'innovative' results, this high percentage comes as no surprise.

Of these 327 'empirically oriented' journals, 223 have a data policy (68.2%). These shares and numbers refer to all types of policies (DAP, ARP and a combination of the two policy types).

4.1. Types of Journal Data Policies

The first outcome identified by applying the content analysis was the types of data policies used by the journals. The main differentiation was made between data availability policies (DAP) and author responsibility policies (ARP). As mentioned in section two, the main difference between these two policy types is the accountability for providing data to would-be replicators. While DAPs ask or require authors to submit their replication files (most often prior to publication of an article) to a third party (e.g. the publishing house, the editorial office or a recognised/trusted data repository), ARPs leave the responsibility for providing data to other researchers to the author.

Fig. 1: Data Policies found in the sample of 327 journals that accept or publish empirical or other data-driven contributions.



In the sample, 185 policies have been classified as a DAP, while 29 policies have been categorised as an ARP (cf. Figure 1).⁹

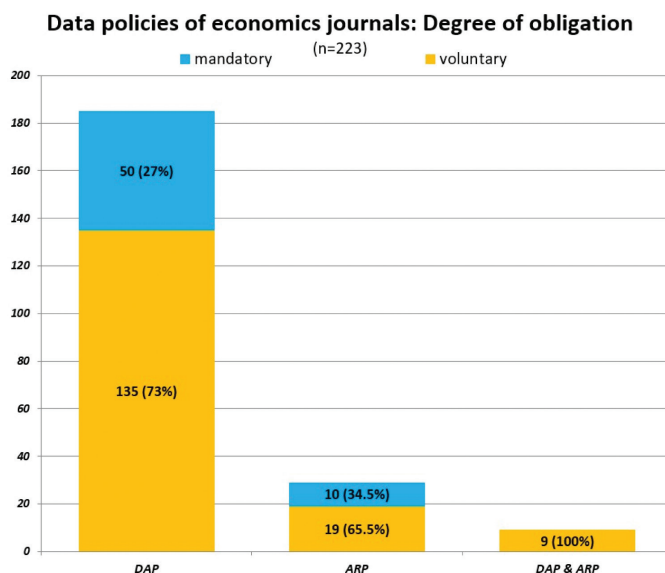
Nine journals offer both a DAP and an ARP (which means that authors may choose between depositing their files in a data repository or providing it to would-be replicators in cases of requests).

4.2. Characteristics of Journal Data Policies

In order to ensure reproducibility as far as possible, the degree of obligation of a data policy plays an important role. As mentioned previously, many studies report issues with voluntary data deposit, while mandatory data policies perform much better.

The results of the content analysis show that 60 out of 223 journals (27%) do have a mandatory data policy (cf. Figure 2). Subdivided into the different

Fig. 2: Degree of obligation of the different types of data policies.



types of data policies, 50 (27%) out of 185 journals with a DAP hold a mandatory policy, while the corresponding number for journals with an ARP is 10 out of 29 (34%). For journals which offer both an ARP and a DAP the respective quantities are zero out of nine (0%).

Another important regulation of journal data policies is whether they offer a procedure for research based on restricted data. Using restricted data is widespread in economics. Such data could be purchased/proprietary, protected due to privacy restrictions or might not be accessible due to reasons of confidentiality. The analysis shows 38 journals (17%) in total with specific regulations for research based on restricted data. Thirty-seven of these are journals with a DAP. One is a journal with an ARP (cf. Figure 3).

As a further result of the content analysis, the demands of the data policies in relation to specific files and data could be determined. The table below details how often the different files and information have been requested by the data policies in the sample (cf. Table 1).

All policies ask for data. This comes as no surprise, as the term “data” has been the selection criterion to determine a data policy. There are other

Fig. 3: Journals with procedures for research based on restricted data.

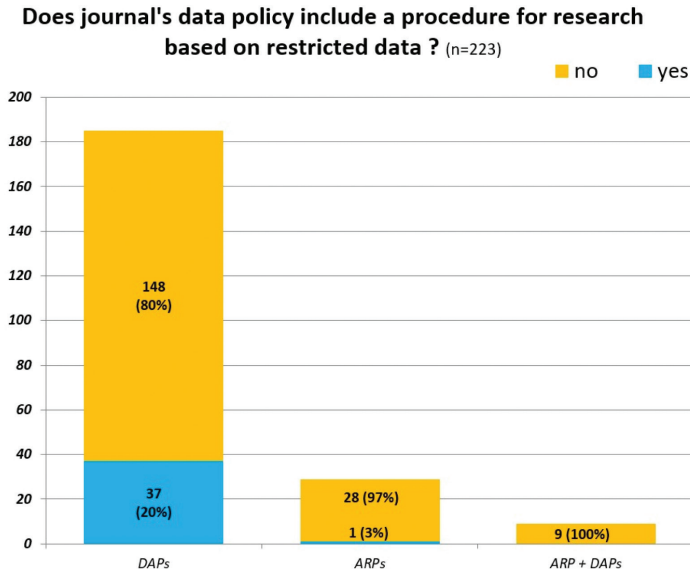


Table 1: Specifications and demands of journal data policies

Specification	DAP (n=185)	ARP (n=29)	DAP & ARP (n=9)	Total (n=223)
Data/sets	185 (100%)	29 (100%)	9 (100%)	223 (100%)
Program Code	122 (65.9%)	26 (89.7%)	2 (22.2%)	150 (67.3%)
Descriptions/documentations	34 (18.4%)	22 (75.9%)	1 (11.1%)	57 (25.6%)
Intermediate datasets	17 (9.2%)	0 (0%)	0 (0%)	17 (7.6%)
Data statement	129 (69.7%)	0 (0%)	9 (100%)	138 (61.9%)

findings, which astonish more: If we only look at the results for the DAPs, the first thing that stands out is the very small number of policies that demand descriptions or documentations. Without proper descriptions, it often is difficult to make use of that data. In addition, a comparatively low percentage, two thirds, of the policies ask for the program code. In view of the fact that the program code is essential to be able to reproduce the published results, this proportion is not satisfactory. Authors of 17 journals are also invited to submit intermediate datasets. All of these journals have adopted the robust data policy of the American Economic Association or use a modified version of it.

Data statements represent a new requirement of journal data policies and are widespread. Although they create transparency with regard to the availability

of the data used, they do not help reproduce the results, strictly speaking. Almost 70% of all journals ask their authors to submit a data statement.

If we compare the specifications of the DAPs with those of the ARPs several differences can be observed: While all policies ask for data, the average demands for program code and descriptions are higher for ARPs than for DAPs. However not a single journal with an ARP asks for data statements or intermediate datasets. The nine journals that offer both policy types are the group with the weakest policies in the sample. While all of these ask for data and data statements all the other information, which is crucial to reproduce the results of a paper, is rarely requested.

In summary, many journals lack strong or detailed data policies. Often, the data policies do not mention fundamental requirements to ensure reproducibility (e.g., documentation and descriptions or - to a lesser degree - program code). Also the implementation of procedures, which should help in reproducing the findings of papers based on restricted data, is not widespread. To submit intermediate datasets is definitely a useful recommendation to ensure good scientific practice, but it also is a demanding requirement.

In order to make data available, most of the data policies suggest storing the data in a recognised repository (65%). All Elsevier journals offer to deposit the data in their in-house product Mendeley (28.7%). For 22% putting the files online on a personal or institutional website is an acceptable solution (cf. Table 2).

Also with regard to data deposit, things have changed over the last years. In the past, journals typically attached the data to the article on the publisher's

Table 2: Recommendations for depositing and disclosure of research data and replication files made by journals (multiple answers possible)

Specification	DAP (n=185)	ARP (n=29)	DAP & ARP (n=9)	Total (n=223)
Recognised data repository	136 (73.5%)	0 (0%)	9 (100%)	145 (65%)
Dataverse	4 (2.2%)	0 (0%)	0 (0%)	4 (1.8%)
Website	47 (25.4%)	1 (3.5%)	1 (11.1%)	49 (22%)
Mendeley data	63 (34.1%)	0 (0%)	1 (11.1%)	64 (28.7%)
On request	0 (0%)	26 (89.7%)	9 (100%)	35 (15.7%)
Not stated	6 (3.2%)	1 (3.5%)	0 (0%)	7 (3.1%)
Other	7 (3.8%)	1 (3.5%)	0 (0%)	8 (3.6%)

website as the default (cf. Vlaeminck, 2013). This was not a useful practice as data often went behind paywalls. A few journals offered their own data archives or recommended data repositories like Dataverse. Now some of the big publishing houses maintain lists of recommended data repositories on their webpages. Most often, these lists break down the repositories by the different scientific domains.

4.3. Journal Data Policies and Publishers

A breakdown of the most important publishing houses of the 353 journals listed in the 2017 JCR ECON shows that Elsevier (70; 19.8%), Wiley (67; 19%), Springer (48; 13.6%), Taylor & Francis (38, 10.8%) and Oxford University Press (24, 7.6%) are the publishers with the highest number of journals in the prestigious JCR ranking. Altogether, these five publishers account for more than 70% of all JCR ECON journals in the year 2017, what clearly shows their dominance and market power among the top journals in the field.

Figure 4 details the absolute number of the 327 journals that accept data-based contributions. The figure is grouped along the publishing houses and the data policies their journals employ.

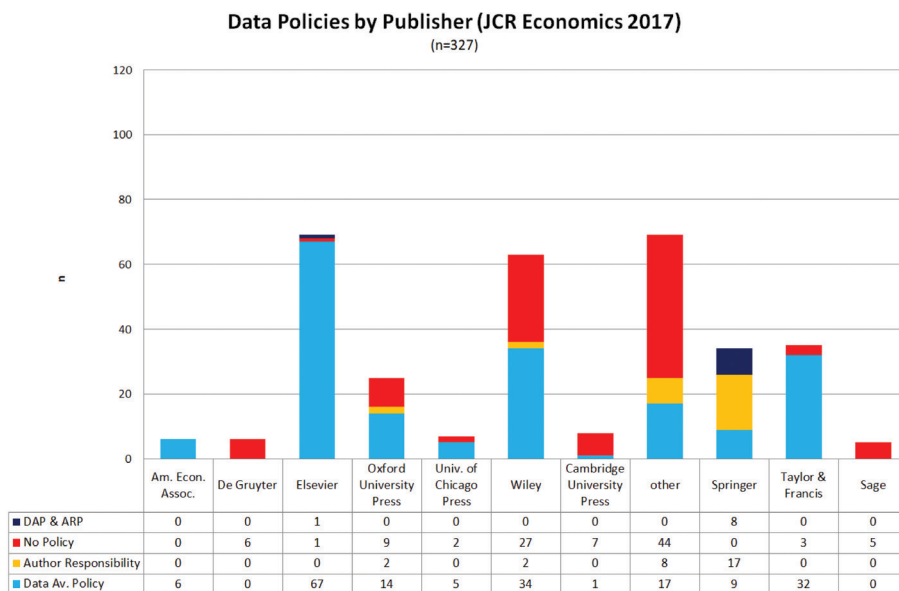
At first glance, we observe very high rates of journals with data availability policies at the publishers Elsevier and Taylor & Francis, while journals with ARP often belong to journals of the SpringerNature group. The SpringerNature group also has the highest number of journals in its holdings that offer both policy types.

As we worked through the data policies, we noticed that the guidelines of journals from the same publisher are often very alike. They are often structured very similarly and even their wording is also identical in many cases.

Some publishers (e.g. SpringerNature) use almost the same wording for their voluntary and their mandatory data policy. Sometimes only a few words have been changed (for instance, authors are not *encouraged* but *required* to follow the data policy).

Many publishers seem to have developed such a standard data policy. This seems to be the case for most major publishers. However, the policies differ quite substantially between different publishing houses.

Fig. 4: Journals with empirical contributions sorted by publishing houses and data policy.



In the past (cf. Vlaeminck, 2013; Vlaeminck & Herrmann, 2015a) most data sharing policies were often individual. One exception was the American Economic Review’s data policy, which had become a quasi-standard at that time (although only a handful of journals outside the American Economic Association used their policy).

Below, a more detailed summary for the five biggest publishers in the JCR ECON 2017 is given. Here, some major differences in how journals of these publishers structure their data policies are described (for details c.f. appendix B and the output file of the statistical analysis).

4.3.1. Elsevier

All but one of Elsevier’s journals that accept data-based contributions have a research data policy (98.6%). Almost all of them can be characterised as a DAP (98.5%). One journal has an ARP. At the point of evaluation, more than 60 of their journals in the JCR ECON used more or less the same data policy. It appears to be the ‘standard policy’ for most of the publisher’s journals. It consists of five paragraphs in the guide for authors that relate to the handling

and disclosure of research data. All of these paragraphs appear to be generic and not specific to the situation in economics. In most cases, the data policy is voluntary, but there are nine journals (13.2%) with a mandatory data policy. Nevertheless, the policy mentions the most important information to ensure reproducibility of published research (software, code, models, algorithms, protocols, methods and ‘other useful materials’). The policies rarely mention data descriptions (2.9%). Data statements are requested by 67 data policies (97.1%). Elsevier provides templates for these. Here, authors can choose the template that best fits to their data and intentions. A defined procedure in cases where authors used restricted data for their scientific findings is available in only four cases (5.9%).

In terms of data deposit, all but two journals (97.1%) suggest storing the data in a data repository. All but four (94%) also recommend Mendeley Data, an in-house product. Simply attaching the data to the article on the journals’ websites is accepted by only seven journals (10.3%).

4.3.2. SpringerNature

Per default SpringerNature included a standardised data policy (ARP) for all of their journals. It is located in the ‘ethical responsibilities of authors’ section, which might not be read by all of their authors. The policy states: *“Upon request authors should be prepared to send relevant documentation or data in order to verify the validity of the results.”* Beyond this sentence, 26 out of 43 (60.5%) journals have an additional DAP. Eight of these 26 also offer an ARP.

Less than half of SpringerNature’s data policies mention data documentation (44.2%) or disclosure of program code (48.8%). The policies are voluntary most often: Only three journals (7%) are mandatory. Very few journals have an individual data policy. Among these journals, there is just one journal (2.3%) with a defined procedure in the event researchers use restricted data. 24 (55.8%) journals recommend depositing research data in a recognised repository.

4.3.3. Wiley

Journals published by Wiley are more reluctant with implementing research data policies. Thirty-six out of 57 journals (57.1%) possess such a policy. That is the lowest share among the big publishing houses in the sample. Among

these, 34 (94.4%) have a DAP. In contrast to journals published by Elsevier or SpringerNature, many of Wiley's journals still have an individual data policy. Frequently, these policies are very detailed. Fourteen of these policies (38.9%) are mandatory. The same number of journals has a procedure in the event authors use restricted data. That is the highest percentage among major publishing houses. 58.3% of the policies recommend depositing the data in a data repository. Two-thirds (66.7%) of the policies ask for data statements.

Wiley also has a joint data policy for some of their journals. It consists of a short paragraph with two sentences in which the publishers asks to deposit research data in a public repository and to provide a data accessibility statement.

4.3.4. Taylor & Francis

88.9% (32) of journals published by Taylor & Francis have a data policy. Almost all of these policies are identical. It consists of three paragraphs. Core is the "Basic Data Sharing Policy" (Taylor & Francis also offers stricter data sharing policies that are used primarily by journals in the sciences). All of these policies are DAPs, but not one single policy is binding. In addition, there is not a single policy with a defined procedure for research based on restricted data. Taylor & Francis' standard policy is also weak in other aspects: recommendations to submit program code (6.3%) or data documentation (3.1%) are below average (equivalent to two and one journals, respectively). In contrast, all data policies require data statements (100%) and almost all policies (96.9%) recommend depositing data in a repository.

4.3.5. Oxford University Press

The data policies of journals published by Oxford University Press remain highly individual. An overarching standard data policy of the publisher does not appear to exist to date.

Only 16 (66.7%) of their journals have a data policy. However, fourteen of these (87.5%) have a DAP. 12 of these data policies are mandatory (75%) - the highest share of all publishers in the sample. In addition, all data policies require the program code of calculations (100%) and six (37.5%) require

data documentation. The same number of journals have a policy on the use of restricted data. Surprisingly, not a single journal recommends depositing data in a repository. Most guidelines (81.3%) refer to publishing the data on the publisher's website along with the article.

5. A Comparison with the Situation in 2014

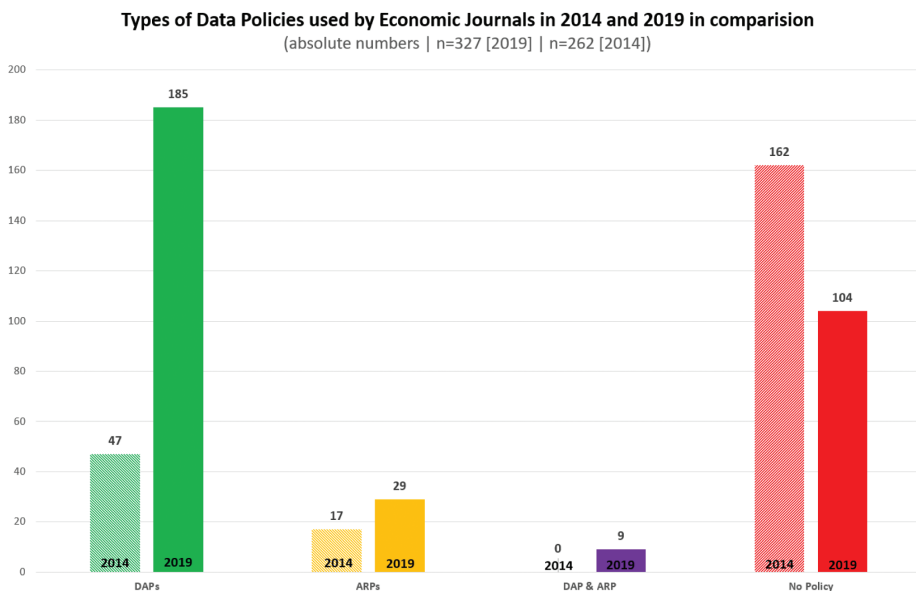
In order to classify the findings of this study, a comparison with previously published results was made. For this purpose, this work reused a dataset compiled by Vlaeminck and Herrmann (2015b). The dataset contains information on data policies of economics journals and their specifications for a sample of 346 journals. Of these journals, 262 also had an Impact Factor and almost all of them appeared in the Social Sciences Citation Index (SSCI). The Journal Citation Reports (JCR) for the specific disciplines can be seen as (non-exclusive) sub collections of the SSCI (e.g. JCR Economics, Business, Management, etc.). The 262 journals therefore serve as a useful comparison group for an in-depth analysis of the types of data policies in use. In addition, it helps to determine potential differences in the specifications of these policies.¹⁰

The comparison suggests fundamental changes in the way journals deal with data (cf. Figure 5): While in 2014 only 47 journals (17.9% out of 262) held a DAP, the corresponding number in 2019 was 185 (56.6% out of 327). This discrepancy suggests a paradigm shift in the academic publishing sector. For the other policy types, the changes have not been substantial as the growth of journals with ARPs and of a combination of DAP and ARP is much smaller.

However, the sheer number of data policies in economics journals does not necessarily say much about the policies' quality. A useful data policy should request data, program code, and descriptions as these elements are crucial to reproduce the results of an empirical article.

When examining the specifications and demands of the data availability policies, significant changes can be observed: While in 2014 83% of the journals asked to provide the program code of an analysis, the share diminished to 66% in the recent study. The request to post descriptions of the data and the research process was mentioned by 74.5% of the data policies in 2014 compared to only 18.4% of the policies in 2019.¹¹

Fig. 5: Data policies of economics journals in 2014 and 2019 compared.



Also concerning the policies' degree of obligation the numbers vary considerably: While the overall number of journals with mandatory data availability policies has grown, their share has dropped from 63.8% to only 27%. In 2014, 51.1% of the journals had a procedure that specifies which data and information authors have to provide in the event they used restricted data. This percentage plummeted to 20% in 2019 (cf. Table 3).

Summarised, these numbers suggest two trends: On the one hand, we observe a massive increase of journal data policies among the most prestigious periodicals in economics. Specifically, the rate of increase of journals with DAPs is tremendous. In 2019, the absolute number of journals with such a data policy is almost four times as high as five years ago. Also the number of journals with an ARP or a combination of the two policy types has grown, but to a much lesser degree.

On the other hand, the average quality of these policies has not improved. Indeed, in absolute numbers more journals ask for the program code. More of these policies are obligatory and have a procedure for the use of restricted data. However, on average, the share of journals that are asking for program

Table 3: Specifications and demands of data availability policies in 2014 and 2019 in comparison

Specification	DAPs 2019 Sample: JCRECON2017 (n=185 of 223)	DAPs 2014 Sample: SSCI2013 (n=47 of 262)	Changes (in %) between 2014 and 2019
Dataset(s)	185 (100%)	47 (100%)	0%
Program Code	122 (65.9%)	39 (83%)	-17.1%
Description/documentation	34 (18.4%)	35 (74.5%)	-56.1%
Intermediate datasets	17 (9.2%)	14 (29.8%)	-20.6%
Mandatory data policies	50 (27%)	30 (63.8%)	-36.8%
Procedure for restricted. data	37 (20%)	24 (51.1% ^{1a})	-31.1%

^aThree journals discouraged the submission of papers that rely on restricted data. For this reason, these journals have not been included in the count.

code, descriptions or intermediate datasets has diminished considerably over time. In addition, the share of mandatory data policies and of the policies that offer a specific procedure for research based on restricted data has fallen rapidly.

6. Summary and Discussion

The aim of this paper was to answer the question of how economics journals today deal with the inclusion of underlying data and analysis of an article in the peer review and publication process. To this end, we analysed how many of the journals listed in the 2017 edition of JCR ECON have specific rules (data policies) that address the use of data in the peer review or publication processes.

The study found that of the 353 journals listed in the 2017 edition of JCR ECON 327 journals (92.6%) publish empirical or data-driven research articles at least sporadically. Of these 327 journals, 223 have a data policy (68.2%).

The types of data policies found can be categorised into 185 journals (56.6%) with a data availability policy (DAP), 29 journals (8.9%) with an author responsibility policy (ARP) and nine journals that both offer a DAP and an ARP (2.8%). The remaining 104 journals (31.8%) do not have a data policy.

In light of the findings of previous studies on data policies of journals in economics, the results indicate a clear trend towards the adoption of data policies. In recent years, there seems to have been a veritable paradigm shift among journals and publishers when it comes to data policies. Moreover, it is satisfying that the overwhelming majority of journals has implemented DAPs rather than ARPs. As described in section 2, ARPs do not work well in practice.

The paper also illustrated some characteristics of the data policies found in the sample. As one important characteristic, the study determined the share of mandatory and voluntary data policies. As a result of the content analysis, it was found that only a minority of the journals have mandatory policies: 50 (27%) of the DAPs are mandatory, 10 (34.5%) of the ARPs but not a single journal that offers both an ARP and DAP. These numbers are comparatively low, especially when we take into account how much journals enforce other parts of their editorial policies (e.g. the use of style sheets), that are less important (from a scientific point of view) than the reproducibility of an article's results.

Regarding whether the guidelines include a process for research that relies on proprietary or confidential data, the study found 37 journals with a DAP (20%) that have such specific requirements (the corresponding numbers for journals with an ARP is one (3.4%) and zero for journals that support both types of policies). Journals without specific rules for such articles often grant exemptions from their data policies. However, to exempt these articles from basic scientific quality criteria does not seem to be a useful approach to me. Since economists often work with such proprietary data, the low percentage of policies with specific rules in this area is not satisfactory.

With respect to requested files and information mentioned in the data policies, all 185 DAP mentioned datasets (100%). Two-thirds (66%) also ask for program code/code of computation (e.g., of statistical/econometric analyses). Documentations of the data or research process were mentioned by less than one-fifth (18.4%). Intermediate datasets were mentioned in only 9.2% of the guidelines. Some journals also have specific requirements for experiments. In these cases, authors have to submit additional information (e.g. instructions or information about subject eligibility).

In comparison, journals with an ARP were more frequently asking for program code (89.7%) and descriptions of the data and/or research process

(75.9%). All journals offering both types of guidelines asked for data sets (100%), but only two (22.2%) mentioned program code. Only one of these journals (11.1%) asked for descriptions or documentation of the data and/or research process.

A relatively new development is that journals have started to request data statements. These statements were demanded by 69.1% of journals with a DAP, by none of the journals with an ARP, and by all journals offering both policy types. Strictly speaking, a data statement does not really help in reproducing published results. However, it clarifies the accessibility of the data used by researchers.

With respect to data deposit, it is very positive that journals have begun to recommend trusted data repositories for storing replication data. Some years ago, journals often hosted the data next to the research article or on a separate webpage. This practice often resulted in issues with data accessibility (for instance after changes to the webpages and URLs) or data has been locked behind paywalls.

The study also contrasted the data policies of the journals of the five largest publishers within JCR ECON. In contrast to the situation a few years ago, most publishers now have a standard data policy that is used by most of their journals. The publishers' standard data policies are not similar to each other. Frequently, these policies are not particularly detailed and they often lack the precision necessary to ensure reproducibility. The only commonality among all publishers' data policies is that they are generic, whereas most journals with an individual data policy are subject-specific and thus often more robust in terms of reproducibility. Elsevier journals offer the most detailed standard data policy. Their guidelines name all files and materials that are crucial for reproducing the results of a paper.

The significance of this study becomes particularly evident when comparing the results with previous studies. With respect to the amount of data policies in economics journals, the last few years have brought a fundamental change: This study reports a 14% higher share of journals with data policies compared to the figures mentioned by Höffler (2017) for more or less the same set of journals. Compared to the situation in 2014, the increase is massive: While Vlaeminck and Herrmann (2015a) found a total of 71 journals with a data policy in a sample of 346 economics journals, this study found a total of 223

journals in a sample of 327. Specifically the number of journals with DAPs has almost quadrupled.

While on the one hand the study shows a massive increase of newly established data policies, the numbers also indicate that not all policies can be considered robust. The code of computation for instance is crucial to understanding the research process, cleaning the data and the assumptions and decisions made during the analysis. This program code is requested by too few journals. The lack of documentation is also a serious concern regarding the reproducibility of published findings.

One reason why many of these data policies are still relatively weak could also be that publishers are initially adopting policies that are easy to comply with. At a later stage, after authors got more used to these policies, journals and publishers might state more precisely the requirements towards more strict and/or domain specific rules.

The massive increase in journal data policies may also be rooted in the science policy debates of recent years. While in general, good scientific practice and research integrity have always been crucial topics in academia, the debate has become more visible within the last few years. Discussions on open science but also reports on fraudulent research practices by some researchers have triggered an intensified debate in academia and the public. The publishing houses and journals seem to be responding to these debates by implementing data policies. It will be interesting to see how journal data policies evolve in the future. Therefore, a subsequent study on this set of journals might be useful in a few years to determine if publishers and journals have tightened their data policies. In addition, it might be of interest to investigate to what degree authors comply with these newly introduced data policies and if more data and other replication files become available.

Research libraries could benefit from the results of this study in several ways: If they have not yet begun to extend author advisory services to include data and data submission, the results of this study suggest that it is time to develop such services as soon as possible. Since the majority of prestigious journals in economics have data policies, research libraries should be prepared to respond to the potential increase in requests from researchers in the social sciences and beyond.

For those involved in advising researchers, the results of this study may also help to quickly assess and compare the requirements of different journals and publishers in this field. Going through the results and materials of this study, consultants quickly get an idea of the most important files and materials requested by economics journals and their publishers. This might be particularly helpful if the advisors are not from the field of social sciences or economics.

For those who provide guidance to researchers, we recommend using the American Economic Association (2021) requirements as a guide (American Economic Association, 2021). Their requirements listed in their data and code availability policy will almost certainly meet the expectations of any journal in the field and offer a good overview on what is needed to ensure reproducibility.

Furthermore, the results also indicate that research libraries should think about offering (hands-on) workshops for researchers on how to make their research reproducible. At least in scientific disciplines in which these skills are not taught in undergraduate education, this makes a lot of sense. Specifically young researchers are a useful target group for such seminars.

In addition, the results of the study suggest that already established services-like supporting researchers in finding a trusted repository to deposit their data- might become more important.

7. Appendices

Appendices A and B can be viewed and downloaded in PDF format from the journal article website.

References

American Economic Association. (2005). *Previous data availability policy (2005 - July 10, 2019)*. <https://www.aeaweb.org/journals/policies/data-code/archive/2005>

American Economic Association. (2021). *Data and code availability policy*. <https://www.aeaweb.org/journals/data/data-code-policy>

Bernanke, B.S. (2004). Editorial statement. *The American Economic Review*, 94(1), 404. <http://www.jstor.org/stable/3592790>

- Chang, A. C., & Li, P. (2015). Is economics research replicable? Sixty published papers from thirteen journals say "Usually not". *Finance and Economics Discussion Series*, 2015(83), 1–26. <https://doi.org/10.17016/FEDS.2015.083>
- Chin, M., & Dong, D. (2019, May 6–7). *The quest for replicability: A review of research data policies in economics journals* [Conference presentation]. INCONECSS, Berlin, Germany. https://ink.library.smu.edu.sg/library_research/147
- Clarivate Analytics. (2018). *2017 Journal Citation Reports Economics*. <https://jcr.clarivate.com/jcr/browse-journals>
- Dewald, W. G., Thursby, J. G., & Anderson, R. G. (1986). Replication in empirical economics: the journal of money, credit and banking project. *The American Economic Review*, 76(4), 587–603. <http://www.jstor.org/stable/1806061>
- Duwendack, M., Palmer-Jones, R. W., & Reed, W. R. (2015). Replications in economics: A progress report. *Econ Journal Watch: Scholarly Comments on Academic Economics*, 12(2), 164–191. <https://econjwatch.org/articles/replications-in-economics-a-progress-report>
- Feigenbaum, S., & Levy, D.M. (1993). The market for (ir)reproducible econometrics. *Accountability in Research*, 3(1), 25–43. <https://doi.org/10.1080/08989629308573828>
- Glandon, P. (2011). *Report on the American Economic Review data availability compliance project. Appendix to American Economic Review editors report 2011*. Vanderbilt University. https://digital.kenyon.edu/economics_publications/20/
- Hern, A., & Duncan, P. (2018, August 10). *Predatory publishers: The journals that churn out fake science*. The Guardian. <https://www.theguardian.com/technology/2018/aug/10/predatory-publishers-the-journals-who-churn-out-fake-science>
- Höffler, J. H. (2017). Replication and economics journal policies. *American Economic Review: Papers & Proceedings*, 107(5), 52–55. <https://doi.org/10.1257/aer.p20171032>
- Johnson, R., Watkinson, A., & Mabe, M. (2018). *The STM Report. An overview of scientific and scholarly publishing*. International Association of Scientific, Technical and Medical Publishers. https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf
- Krawczyk, M., & Reuben, E. (2012). (Un)Available upon request: Field experiment on researchers' willingness to share supplementary materials. *Accountability in Research: Policies and Quality Assurance*, 19(3), 175–186. <https://doi.org/10.1080/08989621.2012.678688>
- McCullough, B. D. (2007). Got replicability? The journal of money, credit and banking archive. *Econ Journal Watch: Scholarly Comments on Academic Economics*, 4(3), 326–337. <https://econjwatch.org/articles/got-replicability-the-journal-of-money-credit-and-banking-archive>
- McCullough, B. D. (2009). Open access economics journals and the market for reproducible economic research. *Economic Analysis and Policy*, 39(1), 117–126. [https://doi.org/10.1016/S0313-5926\(09\)50047-1](https://doi.org/10.1016/S0313-5926(09)50047-1)

- McCullough, B. D., & Vinod, H. D. (2003). Verifying the solution from a nonlinear solver: A case study. *American Economic Review*, 93(3), 873–892. <https://doi.org/10.1257/000282803322157133>
- McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2006). Lessons from the JMCB archive. *Journal of Money, Credit and Banking*, 38(4), 1093–1107. <http://www.jstor.org/stable/3838995>
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Sage Publications.
- Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS Journals. *PLoS One*, 4(9), Article e7078. <https://doi.org/10.1371/journal.pone.0007078>
- Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, 115(11), 2584–2589. <https://doi.org/10.1073/pnas.1708290115>.
- Vlaeminck, S. (2013). Data management in scholarly journals and possible roles for libraries – Some insights from EDaWaX. *Liber Quarterly*, 23(1), 48–79. <https://doi.org/10.18352/lq.8082>
- Vlaeminck, S., & Herrmann, L.-K. (2015a). Data policies and data archives: A New paradigm for academic publishing in economic sciences? In B. Schmid & D. Dobрева (Eds.), *New avenues for electronic publishing in the age of infinite collections and citizen science: Scale, openness and trust* (pp. 145-155). IOS Press. <https://doi.org/10.3233/978-1-61499-562-3-145>
- Vlaeminck, S., & Herrmann, L.-K. (2015b): Data policies and data archives: A new paradigm for academic publishing in economic sciences? (Replication data; Version 1), [Data set]. ZBW Journal Data Archive. <http://journaldata.zbw.eu/dataset/data-policies-and-data-archives-a-new-paradigm-for-academic-publishing-in-economics>
- Vlaeminck, S., & Podkrajac, F. (2017). Journals in economic sciences: paying lip service to reproducible research? *IASSIST Quarterly*, 41(1–4), 16. <https://doi.org/10.29173/iq6>
- Wharton Research Data Services. (2021, August 2). *Terms of use*. University of Pennsylvania. <https://wrds-www.wharton.upenn.edu/users/tou/>

Notes

¹ Estimations range up to 33,100 active peer-reviewed English-language journals in 2018, which have collectively published approximately 3 million articles (Johnson et al., 2018, p. 25).

² For the field of economics also other JCR subsets are of potential interest, e.g. Business (140 journals), Business and Finance (98 journals), Management (210 journals) and Operations Research and Management (84 Journals). All journals included in these subsets have an Impact Factor and are part of the SSCI (Social Sciences Citation Index) which is –like the Impact Factor– a proprietary product of Clarivate Analytics.

When looking at how many journals from these rankings are also represented in JCR ECON 2017, the following figures emerge: 13 journals (9.2%) from JCR Business, 40 (39.3%) from JCR Business and Finance, 12 (5.7%) from JCR Management, and another 13 (15.5%) from JCR Operations Research. For this reason, the results of this article relate primarily to journals found in the JCR ECON 2017 and, to a lesser extent, to business and finance journals.

³ We did not examine the printed issues of the journals, although we are aware of one case where the data policy is not available online, but only in the printed issue. Therefore, the results of the analysis describe the lower limit of the data policies found in our sample. The actual number of journals with a data policy is likely to be higher.

⁴ It was not always easy to distinguish between editorial policies that aim to achieve reproducibility of results and editorial notes that only explain what types of files can generally be processed in the submission process. In cases in which only the possibility to also submit data sets was mentioned in general, but no further instructions in terms of requirements for reproducibility were given, we did not consider such a note as a data policy.

⁵ For instance, some experimental designs require other replication files than simulations do. To have a comparability among the different methodologies in economics, we only looked for those files that are important for most methodological approaches.

⁶ Commercial data providers generally do not allow any sharing of research data, even for the purpose of reproducing published results. For example, Wharton Research Data Services (WRDS) - a major aggregator of access to numerous databases from commercial providers - states in its terms of use: "Except as provided under the terms of the Subscription Agreement, you may not reproduce, distribute, modify, adapt, create derivative works of, display, transmit, broadcast, sell, license or in any way exploit the Proprietary Material, in whole or in part, without our advance written consent." (Wharton Research Data Services, 2021).

⁷ Details of the operationalisation are available in appendix A.

⁸ The SSCI is a set of journals of several Clarivate (formerly Thomson Reuters) Journal Citation Reports (JCR) for the social sciences. It includes more than 3400 journals and is subdivided into 58 (non-exclusive) disciplinary rankings. The JCR ECON is one of these disciplinary rankings. For the field of economics also other JCR subsets are of potential interested, e.g. Business (140 journals), Business and Finance

(98 journals), Management (210 journals) and Operations Research and Management (84 Journals).

⁹ Please check appendix A for a detailed description of how a data policy has been assigned to one of the types. Some special cases are journals published by the SpringerNature group: As part of the 'ethical responsibilities of authors', these journals include a short paragraph on 'data disclosure upon request'. In addition, a substantial part of these journals has a dedicated data policy. Journals with an additional data policy have been categorised as journals with a DAP, while those without a dedicated data policy have been categorised as journals with an ARP. Always the more strict policy type was decisive for the categorisation.

¹⁰ On average, the impact factor of the 262 journals was 1.98 (SD: 3.36). The 25% percentile was 0.86, the median 1.37 and the 75% percentile 2.4. For a detailed description of the sample, please cf. Vlaeminck and Herrmann (2015a).

¹¹ One hundred twenty six of the 223 journals from the 2014 study were also listed in the JCR ECON. The average impact factor of these 126 journals was 1.49 (SD: 1.2). Thirty-three journals had a DAP (26.2%), eight journals (6.4%) had an ARP. Also for this reduced subsample, the results remain comparable. The numbers for the 126 journals listed in the JCR ECON are even higher than reported for the sample of all the 262 IF-journals. For comparison: datasets (100%), program code (82.9%), descriptions (75.6%), intermediate datasets (34.2%), mandatory data policies (78.1%), procedure for restricted data (51.2%).