

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Rousi, Antti M.; Laakso, Mikael

**Journal research data sharing policies: a study of highly-cited journals in neuroscience, physics, and operations research**

*Published in:*  
Scientometrics

*DOI:*  
[10.1007/s11192-020-03467-9](https://doi.org/10.1007/s11192-020-03467-9)

Published: 01/07/2020

*Document Version*  
Publisher's PDF, also known as Version of record

*Published under the following license:*  
CC BY

*Please cite the original version:*  
Rousi, A. M., & Laakso, M. (2020). Journal research data sharing policies: a study of highly-cited journals in neuroscience, physics, and operations research. *Scientometrics*, 124(1), 131-152.  
<https://doi.org/10.1007/s11192-020-03467-9>

---

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.



# Journal research data sharing policies: a study of highly-cited journals in neuroscience, physics, and operations research

Antti M. Rousi<sup>1</sup> · Mikael Laakso<sup>2</sup>

Received: 5 July 2019  
© The Author(s) 2020

## Abstract

The practices for if and how scholarly journals instruct research data for published research to be shared is an area where a lot of changes have been happening as science policy moves towards facilitating open science, and subject-specific repositories and practices are established. This study provides an analysis of the research data sharing policies of highly-cited journals in the fields of neuroscience, physics, and operations research as of May 2019. For these 120 journals, 40 journals per subject category, a unified policy coding framework was developed to capture the most central elements of each policy, i.e. what, when, and where research data is instructed to be shared. The results affirm that considerable differences between research fields remain when it comes to policy existence, strength, and specificity. The findings revealed that one of the most important factors influencing the dimensions of what, where and when of research data policies was whether the journal's scope included specific data types related to life sciences which have established methods of sharing through community-endorsed public repositories. The findings surface the future research potential of approaching policy analysis on the publisher-level as well as on the journal-level. The collected data and coding framework is provided as open data to facilitate future research and journal policy monitoring.

**Keywords** Data sharing · FAIR data · Journal research data policies · Open data · Research data · Scholarly publishing

---

Antti M. Rousi and Mikael Laakso have contributed equally.

---

✉ Antti M. Rousi  
[antti.m.rous@aalto.fi](mailto:antti.m.rous@aalto.fi)  
  
Mikael Laakso  
[mikael.laakso@hanken.fi](mailto:mikael.laakso@hanken.fi)

<sup>1</sup> Research Services, Aalto University, Otakaari 1, 02150 Espoo, Finland

<sup>2</sup> Department of Management and Organisation, Information Systems Science, Hanken School of Economics, Arkadiankatu 22, 00100 Helsinki, Finland

## Introduction

The connection between a research article and its underlying data is strong and direct for authors involved in the preparation process of article manuscripts, but the immediate link can weaken or become completely absent as the article gets published without any data to support the reported findings being made available for readers to access. Journal policies for research data sharing have been developing actively since over two decades ago (Piwowar and Chapman 2008), but only more recently has uptake among journals gained wider support, with progress so far having been observed to be uneven across disciplines (Resnik et al. 2019). Journal data sharing policies becoming more common and detailed can be perceived as part of the larger push towards open science driven by science policy as well as individual researchers, where open data sharing is a central element for making research more transparent, reproducible, and increasing its potential impact (McKiernan et al. 2016). A summarised way of expressing the various aims of this movement as it concerns research data are the FAIR principles, where research data is urged to be made findable, accessible, interoperable, and reusable (Wilkinson et al. 2016). Due to the large emphasis on journal publications for disseminating new research findings within most disciplines, journal data sharing policies have a large potential influence on what, when, and where researchers make their research data available.

One factor influencing the rate at which research data is shared has been the lack of incentives or requirements for researchers to do so—if data sharing is not necessary for getting funded or published then why do so? Data citations, i.e. citations to published datasets, are still an emerging practice not fully standardised nor a well-established metric in researcher merit systems. However, it has been found that research articles with shared research data receive a higher number of citations on average (Piwowar and Vision 2013; Dorch et al. 2015). As the study reported in the present article will look closer at, some journals require authors to make the data associated with published articles publicly available. This requirement might act as a data sharing incentive for authors (as in “In order to get published in this forum I will have to make my data available, it is worth the extra effort”), but also as a potential deterrent for reluctant authors (as in “I will publish in some other journal that does not require data sharing for its articles”). The influence of data sharing requirements on how researchers select which grants to apply financing from and which journals to publish in is still largely unknown.

In disciplines where infrastructures have been created for aggregating research data from individual studies the process of data sharing is well-integrated into the research process by being predictable, supported, and expected by all stakeholders. In disciplines where research data sharing is more sporadic and emerging, the initiative to share research data can be triggered late in the research process by, e.g. a journal requiring it upon publication of the article. The definition and standards for research data is also still a developing practice within many disciplines where data formats and integration are not already well-established. For the purposes of transparency and reproducibility not all shared data is of equal quality. If there are no discipline-specific standard formats to adhere to the level of detail and quality of documentation might be so low that the data gives limited utility for researchers who want to build upon the work in the future (Mbuagbaw et al. 2017). This is one of the reasons why there has been an ongoing shift towards peer-review processes that also increasingly include review of underlying data so as to get assurance of methodological integrity, the correct conclusions being drawn, and the utility of the dataset in its publicly shared form.

For this study our main research questions is:

What do the journal data sharing policies within the fields of neuroscience, physics, and operations research instruct with regards to what, when, and where?

## Literature review

As briefly alluded to in the introduction, journals are not the only actors that influence research data sharing, organisations such as research funders and research performing organisations (e.g. universities) have also developed policies that recommend or require researchers to make their research available to various degrees. There is a wealth of literature available on the broad landscape of research data sharing, however, we limit our focus to the context of data sharing as it relates to the activity of journal publishing and there specifically on the ways that journals have (or have not) integrated research data sharing as part of their publication policies. Journal research data policies have actively been studied for over a decade, but what has remained a constant obstacle for making the results from various studies over time comparable is the variance in disciplinary focus, sample selection criteria, and the variation in how researchers have coded these policies. There is no universally adopted standard for journals to express their policies, which is one reason for why researchers have developed their own ways of making policies comparable to each other. In an attempt to summarise the diverse existing research on this topic, Table 1 presents a summary of prior studies of journal research data and editorial policies.

As can be seen from Table 1 research on the topic has been active during the last 10 years. Journal research data policies have previously been studied in the fields of environmental sciences (Weber et al. 2010), political science (Gherghina and Katsani-dou 2013), genetics (Moles 2014), social sciences (Herndon and O'Reilly 2016; Crosas et al. 2018), biomedical sciences (Vasilevsky et al. 2017) and through multidisciplinary approaches (Piwowar and Chapman 2008; Sturges et al. 2015; Blahous et al. 2016; Naughton and Kernohan 2016; Castro et al. 2017; Resnik et al. 2019). Within the findings of recent multidisciplinary studies on journal research data policies, circa 50–65% of journals had a research data policy and 20–30% of these policies were either classified as strong policies or mandated data sharing into a public repository (Sturges et al. 2015; Blahous et al. 2016; Naughton and Kernohan 2016). Furthermore, studies suggest that journals with high Impact Factors also have the strongest data sharing policies (Vasilevsky et al. 2017; Resnik et al. 2019).

Prior studies have presented various classification frameworks for evaluation of journal research data policies (Piwowar and Chapman 2008; Moles 2014; Sturges et al. 2015; Herndon and O'Reilly 2016; Blahous et al. 2016; Crosas et al. 2018; Resnik et al. 2019). As an overall trend, the classification frameworks of research data policies have become more detailed and intricate over time. High-level classifications examining a perceived strength of policies, i.e. whether policies were in general seen as strong or weak, were first to emerge and are used in more recent studies as well (e.g. Piwowar and Chapman 2008; Blahous et al. 2016). However, a more intricate line of policy classifications has emerged where coding schemes include up to 24 variables (e.g. Stodden, Guo and Ma 2013; Moles 2014; Vasilevsky et al. 2017; Resnik et al. 2019). In most recent studies, specific data types, such as life science data, are acknowledged as factors affecting journal data policies and included as variables in classification schemes (Vasilevsky et al. 2017; Resnik et al. 2019).

**Table 1** Summary of prior studies of journal research data policies

Study	Year policies studied	Research discipline(s)	Journals studied	Nr of journals that had a data policy	Main findings	Are the following policy aspects studied?		
						What	Where	When
Piwowar and Chapman (2008)	Unknown	Research which includes biological gene expression microarrays	70	42 (60%)	17 of the policies were found to be relatively weak and 23 strong	Yes Focus on publication-related gene expression microarray data	No	No
Weber et al. (2010)	Unknown	JCR categories Ecology, Evolutionary Biology and Environmental Sciences	307	42 (14%)	10% of journals request and 4% require data to be archived. 10% provide directions where to archive, 6% give directions on how the data should be shared	Yes Identifies policies with directions on how to share data	Yes Identifies policies with explicit directions where to archive	No
Stodden et al. (2013)	2011 and 2012 (longitudinal comparison)	WoS categories Mathematical and Computational Biology, Statistics and Probability, and Multidisciplinary Sciences + 4 additional journals	170	2011: 56 (33%) 2012: 64 (38%)	Between 2012 and 2013 there had been an increase of 16% in the number of data policies, a 30% increase in code policies, and a 7% increase in the number of supplemental materials policies	Yes Differentiates between publication-related data policies, software code policies, and supplementary information policies	Yes Differentiates between policies that offer hosting for the object in question	Yes Differentiates between data that should be shared upon submission and policies where this is not required

**Table 1** (continued)

Study	Year policies studied	Research discipline(s)	Journals studied	Nr of journals that had a data policy	Main findings	Are the following policy aspects studied?		
						What	Where	When
Gherghina and Katsanidou (2013)	2011	Political Science	120 Political Science journals from the 2010 Citation Index excluding such without a website or papers not based on data	19 (16%)	Most of the existing policies were found to be vague on when sharing should happen, inclusive of all types of data, specific in the procedures to be followed, and strongly enforced  The study found positive correlation between journal data policy availability and the following factors: journal impact, factor, journal age, and generality of journal audience	Yes Focuses on "What types of data require sharing?"	No	Yes Focuses on "When is data sharing required in the process of manuscript submission?"
Moles (2014)	2014	Genetics PLOS Genetics, Genetics Research, European Journal of Medical Genetics	4	4 (100%)	Focuses on demonstrating how the proposed Data-PE framework can be applied in practice to quantitatively score data policies. The example journals scored between 7 and 17 of the maximum 24 point achievable	Yes Technical format and scope of data is considered in the "Format", "Metadata", and "Dimensions" variables	Yes Location instructions are captured in the "Stewardship" variable	Yes Time of data sharing is considered as part of the "Pre-Existing Conditions" and "Peer review" variables

**Table 1** (continued)

Study	Year policies studied	Research discipline(s)	Journals studied	Nr of journals that had a data policy	Main findings	Are the following policy aspects studied?		
						What	Where	When
Zenk-Möltgen and Lepthien (2014)	Unknown	Sociology	140 Journals listed in the category of sociology as part of the WoS (SSCI) in 2013	101 (72%)	5% of journals were found to have an explicit data policy and recommend data to be made available. Another 67% referred to a common policy provided by the Association of Learned and Professional Society Publishers (ALPSP) where a recommendation to make data available is stated	No	No	No
Sturges et al. (2015)	Unknown	Science and Social Sciences	371, split between top and bottom ranked in JCR	185 (50%)	Adapted classification framework from Piwowar and Chapman (2008) The journal data sharing policies contained patchy and inconsistent information. Of the 230 journal policies found 76% were found to be weak, with the rest being strong	Yes 2 categories: named data types	Yes 3 categories: Vague Unnamed repository Named repository	Yes 3 categories With submission For peer review On publication or later

**Table 1** (continued)

Study	Year policies studied	Research discipline(s)	Journals studied	Nr of journals that had a data policy	Main findings	Are the following policy aspects studied?		
						What	Where	When
Herndon and O'Reilly (2016)	2003 and 2015 (longitudinal comparison)	Social Sciences (Business/ Finance, Economics, International Relations, Political Science, Sociology)	100, top 20 journals listed in JCR 2003 under 5 different SS categories	2003: 10 (10%) 2015: 39 (39%)	Main focus on quantitative longitudinal comparison of policies	Yes	Yes	Yes
						Differentiates between requiring data or reproducing data + code/syntax	Differentiates between instructions for sharing through the journal or external repositories	Differentiates between policies that instruct data sharing with submission and those that have post-publication policies
Blahous et al. (2016)	2014	Journals included from 11 cross-disciplinary Scopus subject categories	534 (Top 20–40 SJR 2012 ranked journals from 11 Scopus Subject Categories) Some individual journals also added manually	347 (65%)	4 policy categories (No, Supplementary Material, Weak, and Strong. 68 (20%) had strong data policies, 66 (19%) had weak data policies, and 212 (61%) had supplementary material policies) Large discrepancy between policies of journals in the sciences versus social sciences, arts and humanities. No breakdown of results per discipline	No	Yes	No
						Differentiates between journal and repositories. All repositories mentioned in the data policies were recorded		



**Table 1** (continued)

Study	Year policies studied	Research discipline(s)	Journals studied	Nr of journals that had a data policy	Main findings	Are the following policy aspects studied?		
						What	Where	When
Naughton and Kernohan (2016)	Unknown	Sciences and social sciences	250 top ranked journals, No further information but reported to be based around a similar set of journals used in Sturges et al. (2015) and split between the sciences and social sciences	130 (52%) overall 81 (65%) among the science journals 50 (40%) among the social science journals	A 7% increase in journal data policy prevalence since an earlier study based around a similar journal set in Sturges et al. (2015), 30% of the journals mandated some data sharing when 'data sharing' is interpreted to imply deposit of data in a public repository (science 45.8%, social science 10.5%)	No	Yes Only considered policies including deposit to public repository	No
Castro et al. (2017)	2015 and 2017 (longitudinal comparison)	All disciplines	100 English-language OA journals randomly selected from the DOAJ (some exclusions apply), of which 50 of the journals was a targeted OJS subset	28 (28%)	Adapted classification framework from Stodden et al. (2013) The study found that OA journals are less or no more likely to have a policy, and much less likely to have a strong one compared to subscription-based journals. Also studied data citation policies	No	Yes Differentiates whether the place of deposit is specified	Yes Differentiates when data sharing is required

**Table 1** (continued)

Study	Year policies studied	Research discipline(s)	Journals studied	Nr of journals that had a data policy	Main findings	Are the following policy aspects studied?		
						What	Where	When
Vasilevsky et al. (2017)	2016	Biomedical journals WoS categories: Biochemistry and Molecular Biology, Cell Biology, Crystallography, Developmental Biology, Biomedical Engineering, Immunology, Medical Informatics, Microbiology, Microscopy, Multidisciplinary Sciences, and Neurosciences	318 Journal lists filtered to the top quartiles by impact factor or number of articles published in 2013. Some individual journals excluded	217 (68%)	Adapted classification framework from Stodden et al. (2013) 11.9% of journals explicitly stated that data sharing was required as a condition of publication. 9.1% of journals required data sharing, but did not state that it would affect publication decisions. 23.3% of journals had a statement encouraging authors to share their data but did not require it. 9.1% of journals mentioned data sharing indirectly. High impact factors journals were particularly strongly related to the strongest data sharing policies Open access journals were not more likely to require data sharing than subscription journals	Yes Separate classification for policies that only address protein, genomic, and/or data sharing	Yes Has 5 different categories for types of locations recommended by policies	No

**Table 1** (continued)

Study	Year policies studied	Research discipline(s)	Journals studied	Nr of journals that had a data policy	Main findings	Are the following policy aspects studied?		
						What	Where	When
Crosas et al. (2018)	Unknown	Social sciences WoS categories Political Science and International Relations, Economics, Sociology, History, Psychology, and Anthropology	291 (top 50 journals listed JCR under 6 separate SS categories)	155 (53%)	Around a third of the journals in economics and political science require data sharing. Many data policies were found to have room to improve in clarity and explicit data sharing requirements	Yes Differentiates whether qualitative data were addressed explicitly	Yes Differentiates where data are supposed to be archived;	Yes Differentiates what the expected timing of data submission was
Resnik et al. (2019)	Unknown	Several scientific disciplines, including biology, clinical sciences, mathematics, physics, and social sciences	447	252 (56%)	Both impact factors and discipline are associated with the presence of a data sharing policy. Journals with higher Impact Factors are more likely to have data sharing policies; use shared data in peer review; require deposit of specific data types into publicly available data banks; and refer to reproducibility as a rationale for sharing data	Yes Separate classification for policies relating to 1) protein, proteomic, and/or genomic data, 2) computer code, and 3) clinical trial data	Yes Has 5 different categories for types of locations recommended by policies	Yes Only has a variable for policies where shared data will be used in peer review

Journal research data policies related to fields of neuroscience, physics and operations research have previously been examined as follows. Vasilevsky et al. (2017) examined research data policies of 318 biomedical journals and found that 21% of journals required data sharing and, in addition, 14.8% journals addressed only sharing of protein, proteomic, and/or genomic data. Of biomedical journals of that addressed research data in their editorial policies, the most recommended methods of data sharing were public repositories (57.6%) and data hosted in the journals' online platforms (20.7%) (Vasilevsky et al. 2017). In general, the journal research data policies of physics and mathematics (operations research is often considered as a field related to applied mathematics) have been previously examined through large multidisciplinary approaches (Resnik et al. 2019). Even though the focus of Womack's (2015) study was not on journal research data policies per se, his findings offer insight into research data practices within the fields of physics and mathematics. These including that even though only 25% of the 50 sampled articles representing mathematics used original data, the share of available data was relatively high (31.6%). Within the field of physics, Womack (2015) observed that even though 88% of the 50 articles representing the field used data (either original data or reused data), only in 8% of articles was the data available for reuse. More detailed domain-specific analyses of journals in the fields of physics and operations research would be beneficial to gain a better understanding of the current state of policies and their intricacies.

A recent paper by Jones et al. (2019) provides an overview of research data policies for all journals published by Taylor & Francis and Springer Nature. Taylor & Francis launched their data sharing policy initiative for journals at the start of 2018 offering journals five different standardised policies of various policy strength. By the end of the year, the average uptake of their basic data policy (encouragement to share data when possible) ranged between 71 and 83% across journals in various disciplines. Springer Nature started rolling out standardised research data policies for their approximately 2600 journals in 2016, having four different types of standard statements ranging from encouragements (Type 1 and 2) to requirements (Type 3 and 4). As of November 2018 more than 1500 Springer Nature journals have adopted one these four policies, some slightly modified to accommodate for disciplinary specificity. The distribution of policies in ascending order of policy strength: Type 1 (39%), Type 2 (34%), Type 3 (26%), and Type 4 (<1%). Based on the overview of Jones, Grant and Hrynaszkiewicz (2019), which authors are employed at one of the two publishers, there seems to be momentum in more journals adopting data sharing policies that are expressed in standardised way across the publisher-level.

Looking beyond these two publishers and very recent developments, prior longitudinal studies confirm growth in scientific journals adopting research data policies. In their longitudinal study of 170 journals representing computational sciences, Stodden, Guo and Ma (2013) observed that the amount of research data policies increased by 16% between 2012 and 2013, along with increases in code (30%) and supplementary materials policies (7%). Herndon and O'Reilly (2016) observed that within the period of 2003–2015, the amount of high Impact Factor social science journals having a research data policy increased from 10 to 39%. Furthermore, their findings also suggest that on average the policies of 2015 are more exact and demanding than their 2003 counterparts (Herndon and O'Reilly 2016, 229). Castro et al. (2017) found that research data policies of a sampled set of multidisciplinary open access journals did not show signs of adoption of stronger data policies observed during 2 years interval of 2015–2017. Even though the amount of longitudinal studies is limited and earlier domain specific studies have not been replicated recently to observe change in the journal policies, the observations of Stodden et al. (2013) and Herndon and O'Reilly

(2016) combined with the editorial efforts reported by Jones et al. (2019) suggests that scientific journals are more frequently adopting research data policy and that these policies are becoming more demanding. Previous studies also support the notion that it is more common for journals with higher Impact Factor ranking to have data policies.

The development of common frameworks for journals to express, and by extension for researchers to study journal data policies, has been gradually improving over the last decade. In Table 1 we summarised how 14 previous studies had approached development and use of journal selection and coding frameworks. The first study directly on journal data policies by Piwowar and Chapman (2008) relied on relatively crude classification of 70 journals into categories of “No”, “Weak”, or “Strong”, while the most recent by Resnik et al. (2019) incorporated an extensive 24-point framework in a study of 447 journals. For this study we will incorporate a scaled down 14-point framework that is capable of registering the central variables relating to the questions of what, where, and when data is to be shared based on policies from highly cited journals from different disciplines. Previously, an analysis on what, where, and when dimensions of research data policies has been done by at a multidisciplinary level by Sturges et al. (2015). However, Sturges et al. (2015, p. 2449) focused on providing recommendations regarding research data policies of scientific journals and presented their empirical findings at a very general level without domain specific differences, for example. The present study takes the approach of what, where, and when data is to be shared based on coding categories that are mutually exclusive, which helps comparisons between journals representing different fields of science, for example. The framework and data collection methodology is described more closely in the following section.

## Methodology

### Research aim of the empirical study

To answer the research question of the present study, an empirical study was conducted. Given the multitude of previously developed journal research data policy frameworks and quickly developing field of research data related editorial policies (e.g. Jones et al. 2019), the present article seeks to gain recent in-depth observations by conducting a qualitative analysis of current journal policies within the fields of neuroscience, physics and operations research. The fields of physics and operations research were included in the analysis due to lack of detailed studies of journal data policies of these fields. By incorporating neuroscience journals into the analysis, the present study seeks to understand the research data policies of physics and operations research journals in relation to fields whose data policies are affected by inclusion of life science data (e.g. protein, proteomic, genetic and genomic data). Although the above three fields are science, technology and medicine related disciplines, previous studies suggest that they could have different approaches to underlying research data (Womack 2015; Vasilevsky et al. 2017).

## Data collection

Since data collection is a manual process due to the lack of machine-readable policies, the sampling strategy focused on capturing the policies of the most highly cited journals within each discipline, which is a similar approach to many previous studies due to such journals more reliably having an explicit data policy. For selection of journals within the field of neuroscience, Clarivate Analytics' InCites Journal Citation Reports database was searched using categories of neuroscience and neuroimaging. From the results, journals with the 40 highest Impact Factor (for the year 2017) indicators were extracted for scrutiny of research data policies. Respectively, the selection journals within the field of physics was created by performing a similar search with the categories of physics, applied; physics, atomic, molecular and chemical; physics, condensed matter; physics, fluids and plasmas; physics, mathematical; physics, multidisciplinary; physics, nuclear and physics, particles and fields. From the results, journals with the 40 highest Impact Factor indicators were again extracted for scrutiny. Similarly, the 40 journals representing the field of operations research were extracted by using the search category of operations research and management.

Journal-specific data policies were sought from journal specific websites providing journal specific author guidelines or editorial policies. Within the present study, the examination of journal data policies was done in May 2019. The primary data source was journal-specific author guidelines. If journal guidelines explicitly linked to the publisher's general policy with regard to research data, these were used in the analyses of the present article. If journal-specific research data policy, or lack of, was inconsistent with the publisher's general policies, the journal-specific policies and guidelines were prioritized and used in the present article's data. If journals' author guidelines were not openly available online due to, e.g. accepting submissions on an invite-only basis, the journal was not included in the data of the present article. Also journals that exclusively publish review articles were excluded and replaced with the journal having the next highest Impact Factor indicator so that each set representing the three field of sciences consisted of 40 journals. The final data thus consisted of 120 journals in total.

## Data analysis and coding

The journals' author guidelines and/or editorial policies were examined on whether they take a stance with regard to the availability of the underlying data of the submitted article. The mere explicated possibility of providing supplementary material along with the submitted article was not considered as a research data policy in the present study. Furthermore, the present article excluded source codes or algorithms from the scope of the paper and thus policies related to them are not included in the analysis of the present article. The coding scheme of journal data policies utilized in the present article was created through qualitative analysis the above data by means of constant comparative method (Silverman 2005). Preliminary coding was done incorporating the dimensions of what, where and when (for such analyses of journals' article sharing policies, see Laakso 2014). This preliminary coding was further enhanced through comprehensive data treatment until no new variants of the above dimensions could be inferred from the data. Once the journal research data policy framework of the present study was formed, the entire data was rescrutinized to ensure the consistency of the findings.

## Journal research data policy coding framework

The journal research data policy coding framework derived from the empirical data of this study is presented in Table 2. The framework utilises terminology that is further specified here. The term ‘data availability statement available’ refers to whether the journal editorial policies required or encouraged the submitting researcher to issue a statement regarding the sharing, or lack thereof, of the primary data underlying the submitted work.

‘Public deposition’ refers to a scenario where researcher deposits data to a public repository and thus gives the administrative role of the data to the receiving repository. ‘Scientific sharing’ refers to a scenario where researcher administers his or her data locally and by request provides it to interested reader. Note that none of the journals examined in the present article required that all data types underlying a submitted work should be deposited into a public data repositories. However, some journals required public deposition of data of specific types. Within the journal research data policies examined in the present article, these data types are well presented by the Springer Nature policy on “Availability of data, materials, code and protocols” (Springer Nature 2018). These specific data types included DNA and RNA data; protein sequences and DNA and RNA sequencing data; genetic polymorphisms data; linked phenotype and genotype data; gene expression microarray data; proteomics data; macromolecular structures and crystallographic data for small molecules. Furthermore, the registration of clinical trials in a public repository was also considered as a data type in this study. The term ‘specific data types’ used in the custom coding framework of the present study thus refers to above life sciences related data types, including reporting of clinical trials. These data types have community-endorsed public repositories where deposition was most often mandated within the journals’ research data policies. The term ‘location’ refers to whether the journal’s data policy provides suggestions or requirements for the repositories or services used to share the underlying data of the submitted works. The category of ‘immediate release of data’ examines whether the journals’ research data policy addresses the timing of publication of the underlying data of submitted works. Note that even though the journals may only encourage public deposition of the data, the editorial processes could be set up so that it leads to either publication of the research data or the research data metadata in conjunction to publishing of the submitted work.

The data of the present article is available at <https://doi.org/10.5281/zenodo.3268351>.

## Findings

### Research data policies and availability statements

Research data policies were common within the editorial policies of the neuroscience, physics and operations research journals examined in the present study: from all 120 journals, 92 (c. 77%) had incorporated a research data policy into their editorial processes. Research data policies were most common for journals within neuroscience as 35 (85%) of these journals had a research data policy. Respectively, within the three field of sciences examined in the present study, they were least common in physics: 26 (65%) of the sampled journals in physics having a research data policy. Editorial processes that incorporated data availability statements were not as common as only 61 (51%) of all studied journals

required or encouraged such statements. Tables 3 and 4 summarise the findings on implemented research data policies and data availability statements.

### Sharing of previously unpublished data

Within the neuroscience, physics and operations research journals that presented a research data policy, the highest proportion of journals encouraged, but did not require, authors to share their data via public repositories: 12 of the neuroscience, 11 of the physics and 30 of the operations research journals encouraged authors to share their data via public repositories. This total of 53 journals encouraging public deposition of research data constituted c. 58% of all examined journals that included a research data policy in their editorial policy guidelines ( $n=92$ ).

The research data policy of some journals in the fields of neuroscience ( $n=7$ ) and physics ( $n=7$ ) addressed only specific data types related to life sciences. As stated previously, these included data types such as DNA and RNA data; protein sequences and DNA and RNA sequencing data; genetic polymorphisms data; linked phenotype and genotype data; gene expression microarray data; proteomics data; macromolecular structures, crystallographic data for small molecules and clinical trials data. However, instead of a mere encouragement to publicly share data, these journals required the public deposition of these data types. Furthermore, some journals from neuroscience ( $n=6$ ) and physics ( $n=4$ ) similarly required the open deposition of the above specific data types and also in addition encouraged open deposition rest of the research data used in the submitted article. These 24 journals that had adopted either one of the above policy types consisted 26% of all journals that presented a research data policy ( $n=92$ ).

The percentage of journals that mandated the sharing of all data types underlying a submitted work was highest in neuroscience. Scientific sharing of all data, including requirement to publicly deposit the specific life science related data types, was mandated by 10 of the high impact neuroscience journals. Within the journals of physics, 4 journals similarly required sharing of all underlying data. Respectively, 1 of the high impact journals of operations research required sharing of all data types underlying a submitted work. These most demanding research data policies were similar across the examined disciplines. The policies again explicitly required the open deposition of life science related data types such as DNA and RNA data; protein sequences and DNA and RNA sequencing data; genetic polymorphisms data; linked phenotype and genotype data; gene expression microarray data; proteomics data; macromolecular structures, crystallographic data for small molecules and clinical trials data. Furthermore, also sharing of other data types used in the submitted article was required. For other data types, the requirement to share could be fulfilled either via public deposition of data or by giving the data for the scrutiny of any interested reader. This most demanding research data policy type was adopted by c.16% of the examined journals that included a research data policy in their editorial guidelines ( $n=92$ ).

It is important to note that journals who had adopted this last most demanding data policy type represented major commercial publishers, i.e. Elsevier, Sage and Springer Nature (incl. PubMedCentral journals). Within the findings of the present article, only Springer Nature (incl. PubMedCentral journals) required either public deposition or scientific sharing of all data in all of their journals examined, whereas Elsevier and Sage had adopted also less-demanding policies in their high impact journals. The four high



**Table 2** Journal research data policy coding framework

Data points	Coding
Research data policy	Yes/no
Data availability statement	Yes/no
<i>Sharing of previously unpublished data</i>	
Sharing not required or no data policy	Yes/no
Public deposition of data encouraged	Yes/no
Public deposition of specific data types required	Yes/no
Public deposition all data encouraged and for specific types required	Yes/no
Scientific sharing of all data required and public deposition of specific types required	Yes/no
<i>Location recommendations or requirements</i>	
No data policy	Yes/no
No stated location recommendations	Yes/no
Location recommendations	Yes/no
Location requirements for specific types	Yes/no
Location recommendations and requirements for specific types	Yes/no
<i>Immediate release of data</i>	
No data policy	Yes/no
Timing of release not addressed	Yes/no
Immediate release on publication required	Yes/no

Specific data types refer to life sciences data or reporting of clinical trials

impact journals of physics and the one operations research journal, which had this most demanding research data policy type were all published by Springer Nature.

Figure 1 summarises the field-specific differences in research data policies with regard to sharing of previously unpublished data.

### Location recommendations or requirements

Whether the scope of the journal included works utilizing data types of DNA and RNA data; protein sequences and DNA and RNA sequencing data; genetic polymorphisms data; linked phenotype and genotype data; gene expression microarray data; proteomics data; macromolecular structures, crystallographic data for small molecules and clinical trials was also affected the where dimension of the research data policies. Journals that published research related to aforementioned data types also required that their deposition into specific community-endorsed public repositories. It was only with the data type of registering clinical trials that examined journals sometimes did not issue location requirements but mere recommendations. The approach where public deposition of clinical trials data was required but only locations recommendations were made was found exclusively from the neuroscience journals examined in the present study. Within the empirical data of the present study, if a journal research data policy did not address the above data types, only recommendations regarding the services used for data sharing were included in the research data policies.

The examined journals recommended diverse locations for research data sharing, such as discipline specific repositories, general purpose repositories and data journals. Given the

**Table 3** Presence of journal research data policies

Field of science	Data policy defined	No data policy
Neuroscience ( <i>n</i> = 40)	87.5% (35 journals)	12.5% (5 journals)
Physics ( <i>n</i> = 40)	65% (26 journals)	35% (14 journals)
Operations research ( <i>n</i> = 40)	77.5% (31 journals)	22.5% (9 journals)
Total ( <i>n</i> = 120)	77% (92 journals)	23% (28 journals)

**Table 4** Presence of data statements as part of editorial processes

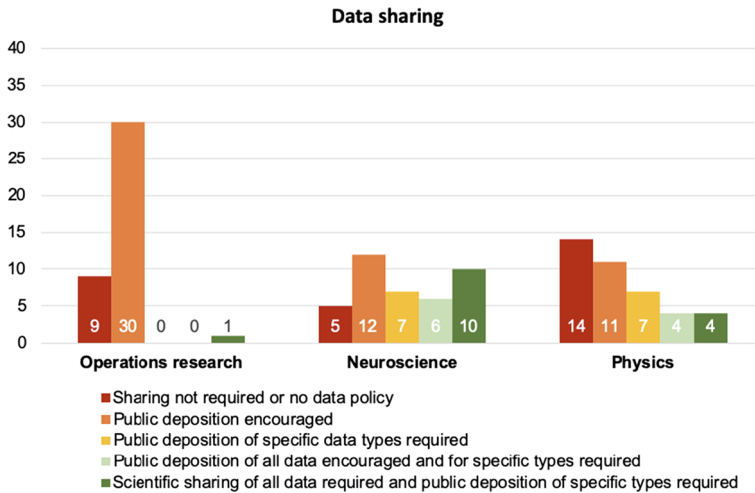
Field of science	Data availability statements	No data availability statements
Neuroscience ( <i>n</i> = 40)	55% (22 journals)	45% (18 journals)
Physics ( <i>n</i> = 40)	37.5% (15 journals)	62.5% (25 journals)
Operations research ( <i>n</i> = 40)	60% (24 journals)	40% (16 journals)
Total ( <i>n</i> = 120)	51% (61 journals)	49% (59 journals)

large variance in approaches, exact codings of location recommendations seemed unfeasible. For example, Springer Nature's (2019) list of recommended data repositories classifies circa 110 different repositories into domain specific categories. However, three main general approaches were identified from the altogether 70 journals (c. 58% of all examined 120 journals) including location recommendations in their research data policies. First, 28 (c. 23% of all examined journals) journals had adopted an approach where the main recommendation was a repository supported by the journal platform. Second, 25 journals (c. 21% of all examined journals) had adopted the approach where recommended locations were presented through domain specific categories listing suitable repositories. Finally, 16 journals (c. 13% of all examined journals) recommended unspecified public repositories as locations of data sharing.

Figure 2 summarises the field-specific differences in research data policies with regard to locations for sharing of previously unpublished data.

## Policy regarding the release of data

Figure 3 summarises the field-specific differences in research data policies with regard to when data should be shared. With regard to timing of sharing the underlying data, the defining factor within research data policies examined in this study appeared to be the editorial processes that were set up so that they lead to publication of the research data in conjunction to publishing of the submitted work. The highest share of journals requiring immediate sharing or, with specific data types, public deposition on publication, where again within the fields of neuroscience and physics. It is noteworthy that often the journals that published research utilizing aforementioned life science data types often required data bank accession number and clinical trial registration number information to be included already in the manuscripts submitted to the journal.

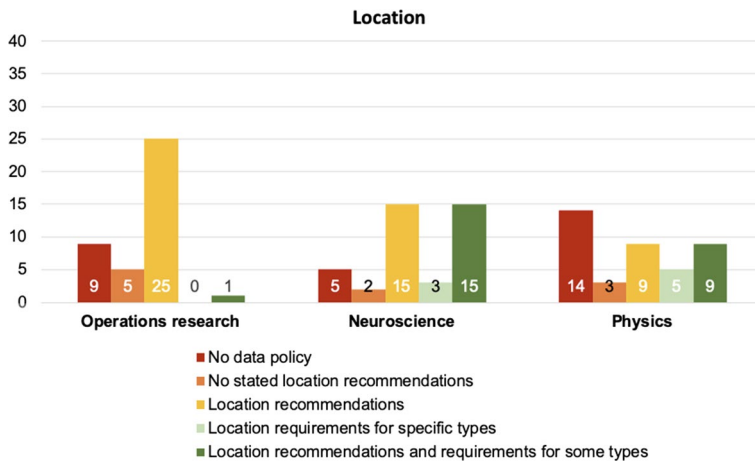


**Fig. 1** Journal policies on data sharing

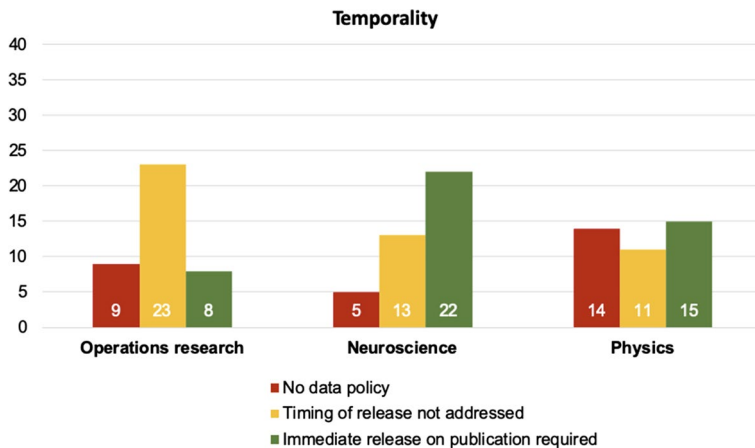
## Discussion

The literature review revealed that research in this domain has been active over the last two decades with various journal population sampling strategies being employed to gain either deep single-discipline results or results that cut across multiple disciplines. A recurring common theme in both previous research and the study reported on in this article is the unevenness of policies between research disciplines, and surprisingly large shares of journals that still do not have research data policies (e.g. in this study a third of physics journals were found to not have one). Prior longitudinal studies confirm growth in scientific journals adopting research data policies. From the review, which was summarized in Table 1, it became clear that many different approaches have been used to standardize heterogeneously expressed policies. A trend which goes in tandem with journals increasingly having data sharing policies and also expressing them in a more detailed way, the classification frameworks have also become more detailed and intricate over time. While some re-use and iterations of frameworks was found to occur it has been common for studies to design granularity and classification criteria to specifically support the research agenda of a specific study. Within the findings of the present study, the requirement of depositing open data to public repositories was given only to data types related to life sciences. This supports the recent tendency (Vasilevsky et al. 2017; Resnik et al. 2019) to separately code these data types within classifications examining journal research data policies in fields where they are of relevance and in multidisciplinary studies.

The empirical findings revealed that one of the most important factors influencing the dimensions of what, where and when of research data policies was whether the journal's scope included specific data types related to life sciences which have established methods of sharing through community-endorsed public repositories. As mentioned previously, these specific data types included DNA and RNA data; protein sequences and DNA and RNA sequencing data; genetic polymorphisms data; linked phenotype and genotype data; gene expression microarray data; proteomics data; macromolecular structures, crystallographic data for small molecules, and clinical trials data. If above data types were included



**Fig. 2** Journal policies on locations of data sharing



**Fig. 3** Journal policies on when data should be shared

in the scope of the journal, it affected the what dimension so that public deposition of these data types were required whereas most often research data policies only encouraged data sharing through public repositories. Furthermore, if above data types were included in the scope of the journal, the research data policy often mandated the use of specific community-endorsed repositories, thus affecting the where dimension of research data sharing. Finally, the pertinence of these life sciences related data types also affected the when dimension of data sharing as some journals required data bank accession number and clinical trial registration number information to be included already in the manuscripts submitted to the journal.

Although no prior detailed analyses of research data policies of neuroscience, physics and operations research journals existed, in general the findings are in line with the prior studies of high impact journals (Vasilevsky et al. 2017; Resnik et al. 2019). The most notable difference between the analysed fields of sciences was that the neuroscience journals,

being part of life sciences, most often included in their scope specific data types, which have established methods of sharing through community-endorsed public repositories. This in turn led to more specific and demanding research data policies than found from the two other fields of sciences examined. The inclusion of these same life science related data types also influenced research data policies of high impact physics journals, albeit not to the same extent as in neuroscience. As exemplified in the findings, these differences between the fields may diminish over time as publishers are adopting standardized policies across their journals (Jones et al. 2019). However, what matters perhaps more is establishing new cultures of open sharing and community-endorsed public repositories around novel specific data types. This in turn might encourage scientific journals to move from mere recommendations of data sharing to requiring open deposition of new kinds of data outside of life sciences.

When the custom coding framework created in the present study is compared to recent classification frameworks presented by Resnik et al. (2019) and Vasilevsky et al. (2017), the following observations can be made. The categorization of data sharing into public deposition and scientific sharing may be seen as a strength of the present study's classification. Even though there are valid limitations to data sharing, such as protecting the identity of participants, public deposition of shared data may be seen to better reflect the FAIR principles in data sharing (see Wilkinson et al. 2016) than the scenario where researchers themselves administer the scientific sharing of data. Thus, the above distinction may be seen as beneficial when examining research data policies of scientific journals. Furthermore, within the present study's coding framework the categories are mutually exclusive, which helps comparisons between journals representing different fields of science.

## Limitations

This study is limited to the most highly cited journals of each discipline included in Clarivate Analytics' Web of Science database, would the sample selection have been wider the results would likely differ considerably based on indications from earlier research. Furthermore, the present article excluded source codes or algorithms from the scope of the paper and thus policies related to them are not included in the analysis of the present article. It is also important to note that the fields of sciences examined in the present study represented science, technology and medicine related fields. Further studies should be conducted using the policy-coding framework of the present study to examine the research data policies of in social sciences and humanities, for example. This approach would make visible whether journals representing these fields of science encourage data sharing, require public deposition of research data or require scientific sharing of research data.

## Conclusion

Our findings continue the consistent trend observed by previous research of considerable disciplinary differences in presence and strength of journal data policies. Publicly deposited open data was found to be a characteristic requirement for neuroscience as part of life sciences, where this concerns in particular deposit of protein, proteomic, genetic, genomic and clinical trial data. For other data types even journals in neuroscience were more lenient in their policies, encouraging or requiring data sharing among researchers but not making

it available as open data. In research disciplines where there is no prevalent experiment or data type that many studies make use of, e.g. operations research, the existence of strong and specific journal data policies is also less prevalent even among the highly-ranked journals in the disciplines. Something which also has an influence on this is research funder requirements, where life sciences is backed by big funders where many have open science as part of their policies, thus trickling down commonly specified requirements to the top journals in the discipline.

As part of the data collection process we could observe that the availability and format through which the journal research data policies were expressed were publisher-dependent, e.g. our results corroborate that Springer Nature has already for several years worked towards the publishers journals having policies expressed in a standardized way (Hrynaskiewicz et al. 2017). It would be useful for future research to include the publisher level in addition to the individual journal level of observation and analysis since there seems to be consolidation happening within several of the large publishers.

**Acknowledgements** Open access funding provided by Aalto University.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Blahous, B., Gorraiz, J., Gumpenberger, C., Lehner, O., & Ulrych, U. (2016). Data policies in journals under scrutiny: Their strength, scope and impact. *Bibliometrie—Praxis Und Forschung*. <https://doi.org/10.5283/bpf.269>.
- Castro, E., Crosas, M., Garnett, A., Sheridan, K., & Altman, M. (2017). Evaluating and promoting open data practices in open access journals. *Journal of Scholarly Publishing*, 49(1), 66–88. <https://doi.org/10.3138/jsp.49.1.66>.
- Crosas, M., Gautier, J., Karcher, S., Kirilova, D., Otalora, G., & Schwartz, A. (2018). *Data policies of highly-ranked social science journals*. SocArXiv Preprint. <https://doi.org/10.31235/osf.io/9h7ay>.
- Dorch, B., Drachen, T., & Ellegaard, O. (2015). The data sharing advantage in astrophysics. *Proceedings of the International Astronomical Union*, 11(A29A), 172–175. <https://doi.org/10.1017/S1743921316002696>.
- Gherghina, S., & Katsanidou, A. (2013). Data availability in political science journals. *European Political Science*, 12(3), 333–349. <https://doi.org/10.1057/eps.2013.8>.
- Herndon, J., & O'Reilly, R. (2016). Data sharing policies in social sciences academic journals: Evolving expectations of data sharing as a form of scholarly communication. In L. Kellam & K. Thompson (Eds.), *Databrarianship: The academic data librarian in theory and practice* (pp. 219–242). Chicago: American Library Association.
- Hrynaskiewicz, I., Birukou, A., Astell, M., Swaminathan, S., Kenall, A., & Khodiyar, V. (2017). Standardising and harmonising research data policy in scholarly publishing. *International Journal of Digital Curation*, 12(1), 65–71. <https://doi.org/10.2218/ijdc.v12i1.531>.
- Jones, L., Grant, R., & Hrynaskiewicz, I. (2019). Implementing publisher policies that inform support and encourage authors to share data: Two case studies. *Insights the UKSG Journal*, 32(1), 1–11. <https://doi.org/10.1629/uksg.463>.

- Laakso, M. (2014). Green open access policies of scholarly journal publishers: A study of what, when, and where self-archiving is allowed. *Scientometrics*, 99(2), 475–494. <https://doi.org/10.1007/s11192-013-1205-3>.
- Mbuagbaw, L., Foster, G., Cheng, J., & Thabane, L. (2017). Challenges to complete and useful data sharing. *Trials*. <https://doi.org/10.1186/s13063-017-1816-8>.
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., & Kenall, A. (2016). How open science helps researchers succeed. *Elife*. <https://doi.org/10.7554/eLife.16800.001>.
- Moles, N. (2014). Data-PE: A framework for evaluating data publication policies at scholarly journals. *Data Science Journal*, 13(27), 192–202. <https://doi.org/10.2481/dsj.14-047>.
- Naughton, L., & Kernohan, D. (2016). Making sense of journal research data policies. *Insights*, 29(1), 84–89. <https://doi.org/10.1629/uksg.284>.
- Piwowar, H. A., & Chapman, W. W. (2008). A review of journal policies for sharing research data. In *Sustainability in the age of web 2.0—Proceedings of the 12th international conference on electronic publishing*. Toronto: ELPUB.
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. <https://doi.org/10.7717/peerj.175>.
- Resnik, D. B., Morales, M., Landrum, R., Shi, M., Minnier, J., Vasilevsky, N. A., et al. (2019). Effect of impact factor and discipline on journal data sharing policies. *Accountability in Research*, 26(3), 139–156. <https://doi.org/10.1080/08989621.2019.1591277>.
- Rousi, A.M., & Laakso, M. (2019). Data of the submitted article “Journal research data sharing policies: A study of highly-cited journals in neuroscience, physics, and operations research”. *Zenodo*. <https://doi.org/10.5281/zenodo.3268351>.
- Silverman, D. (2005). *Doing qualitative research* (2nd ed.). London: Sage.
- Springer Nature. (2018). *Availability of data, materials, code and protocols*. Resource document. <https://www.nature.com/authors/policies/availability.html>. Accessed 9 May 2018.
- Springer Nature. (2019). *Recommended repositories*. Resource document. <https://www.springernature.com/gp/authors/research-data-policy/repositories/12327124>. Accessed 9 January 2020.
- Stodden, V., Guo, P., & Ma, Z. (2013). Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals. *PLoS ONE*, 8(6), e67111–e67118. <https://doi.org/10.1371/journal.pone.0067111>.
- Sturges, P., Bamkin, M., Anders, J. H., Hubbard, B., Hussain, A., & Heeley, M. (2015). Research data sharing: Developing a stake-holder driven model for journal policies. *Journal of the Association of Information Science and Technology*, 66(12), 2445–2455. <https://doi.org/10.1002/asi.23336>.
- Vasilevsky, N. A., Minnier, J., Haendel, M. A., & Champieux, R. E. (2017). Reproducible and reusable research: Are journal data sharing policies meeting the mark? *PeerJ*, 5, e3208. <https://doi.org/10.7717/peerj.3208>.
- Weber, N. M., Piwowar, H. A., & Vision, T. J. (2010). Evaluating data citation and sharing policies in the environmental sciences. *Proceedings of the ASIST annual meeting*, 47, 1–2. <https://doi.org/10.1002/meet.14504701445>.
- Wilkinson, M., Dumontier, M., Aalbersberg, I., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Womack, R. P. (2015). Research data in core journals in biology, chemistry, mathematics, and physics. *PLoS ONE*, 10(12), e0143460. <https://doi.org/10.1371/journal.pone.0143460>.
- Zenk-Möltgen, W., & Lepthien, G. (2014). Data sharing in sociology journals. *Online Information Review*, 38(6), 709–722. <https://doi.org/10.1108/OIR-05-2014-0119>.