

Заявление о доступности данных: инструмент повышения доверия к исследованию

Для цитирования:

Тихонова, Е. В., Мекеко, Н. М. (2024). Заявление о доступности данных: инструмент повышения доверия к исследованию. *Журнал работа и карьера*, 3(2), 4–9. <https://doi.org/10.56414/jeac.2024.2.72>

Корреспонденция:

Елена Викторовна Тихонова
tikhonova.e.v@inno.mgimo.ru

Заявление о доступности данных:

данные текущего исследования доступны по запросу у корреспондирующего автора.

Поступила: 16.05.2024

Поступила после

рецензирования: 10.06.2024

Принята к публикации:

20.06.2024

© Тихонова Е. В., Мекеко Н. М., 2024

Конфликт интересов:

авторы сообщают об отсутствии конфликта интересов.

Е. В. Тихонова¹, Н. М. Мекеко²

¹ МГИМО Университет, Москва, Российская Федерация

² РУДН, Москва, Российская Федерация

АННОТАЦИЯ

В редакционной статье обсуждается важность заявления о доступности данных исследования (Data Availability Statement, DAS) в научных публикациях. Это заявление предоставляет информацию о местонахождении данных, использованных в исследовании, и условиях доступа к ним, что способствует прозрачности и воспроизводимости научных результатов. В статье представлены шаблоны для составления заявления, которые могут быть адаптированы в зависимости от характера данных и требований научных журналов. Статья подчеркивает значимость корректного оформления заявления для укрепления доверия к результатам исследований. Статья построена в формате ответов на наиболее частотные вопросы авторов научных исследований.

КЛЮЧЕВЫЕ СЛОВА

заявление о доступности данных; цитируемость; доверие к исследованиям; прозрачность исследования; воспроизводимость результатов; открытый доступ к данным; валидация данных; репозитории данных

Data Availability Statement: A Tool for Enhancing Research Trust

Elena V. Tikhonova¹, Natalia M. Mekeko²

¹ MGIMO University, Moscow, Russian Federation

² RUDN University, Moscow, Russian Federation

ABSTRACT

This editorial article discusses the importance of the Data Availability Statement (DAS) in scientific publications. The statement provides information on the location of data used in the research and the conditions for accessing it, which promotes transparency and reproducibility of scientific results. The article presents templates for drafting a DAS, which can be adapted depending on the nature of the data and the requirements of scientific journals. The article emphasizes the significance of correctly preparing the statement to strengthen trust in research outcomes. The article is structured in a Q&A format addressing the most common questions from research authors.

KEYWORDS

data availability statement; citation impact; research trust; research transparency; result reproducibility; open data access; data validation; data repositories

For citation:

Tikhonova, E. V., Mekeko, N. M. (2024). Data Availability Statement: A Tool for Enhancing Research Trust. *Journal of Employment and Career*, 3(2), 4–9. <https://doi.org/10.56414/jeac.2024.2.72>

Correspondence:

Elena V. Tikhonova
etikhonova@gmail.com

Data Availability Statement:

Current study data is available upon request from the corresponding author.

Received: 16.05.2024

Revised: 10.06.2024

Accepted: 20.06.2024

© Tikhonova, E. V., Mekeko, N. M., 2024

Declaration of Competing Interest: none declared.



ВВЕДЕНИЕ

Вопросы воспроизводимости результатов исследований и их прозрачности в науке стали центральными темами, особенно в свете увеличивающегося количества научных публикаций и повышенным требованиям к качеству данных. Так, масштабное исследование, опубликованное в Nature, продемонстрировало, что более 70% ученых столкнулись с проблемами воспроизводимости своих и чужих исследований (Baker, 2016). Таким образом, в очередной раз продемонстрирована необходимость более строгого контроля за доступностью данных.

Заявление о доступности данных (Data Availability Statement, DAS) играет ключевую роль в этом процессе, обеспечивая прозрачность исследования и возможность проверки результатов исследований, параллельно свидетельствуя об ответственности авторов опубликованной статьи перед научным сообществом (Cech, 2003; Wicherts et al., 2021). Публикации, сопровождаемые заявлением о доступности данных, не только оптимизируют доступ к данным (Longo & Drazen, 2016; Wallis et al., 2013), но и повышают шансы рукописи на цитируемость, что подтверждается исследованиями, демонстрирующими связь между открытым доступом к данным и увеличением числа цитирований (Piwowar et al., 2007; Colavizza et al., 2020).

Однако, как показывают исследования, практика обмена данными и доступности данных по запросу значительно варьируется в зависимости от научной дисциплины (Tedersoo et al., 2021; Tenopir, 2015; Hardwicke et al., 2018). В некоторых областях, таких как экология, внедрение обязательного архивации данных значительно улучшило доступ к научным материалам (Vines et al., 2013; Vines et al., 2022), в то время как в других областях проблемы с доступностью данных остаются актуаль-

ными (Hardwicke & Ioannidis, 2018; Krawczyk & Reuben, 2012). Эти противоречия подчеркивают необходимость разработки четких и универсальных стандартов для представления заявления о доступности данных, которые обеспечили бы единообразие и доступность данных в различных научных дисциплинах.

Цель данной статьи – рассмотреть значение заявления о доступности данных и сформулировать рекомендации по его формулированию.

В ЧЕМ СУТЬ ЗАЯВЛЕНИЯ О ДОСТУПНОСТИ ДАННЫХ?

Заявление о доступности данных – это краткое пояснение, которое указывает, где именно хранятся данные, использованные в исследовании, и каким образом они могут быть получены. В заявлении также могут быть указаны ссылки на наборы данных, если они существуют в открытом доступе. Цель такого заявления – обеспечить доступность данных для других исследователей, что способствует воспроизводимости и проверяемости научных результатов (Tedersoo, 2021).

КАК КОРРЕКТНО СФОРМУЛИРОВАТЬ ЗАЯВЛЕНИЕ О ДОСТУПНОСТИ ДАННЫХ?

Содержание заявления о доступности данных варьируется в зависимости от характера исследования, типа данных и требований журнала, в котором планируется публикация. Существуют различные шаблоны для составления заявления о доступности данных, и выбор конкретного шаблона зависит от условий, при которых авторы исследования согласны предоставить доступ к данным. Некоторые из наиболее распространенных шаблонов показаны в Таблице 1.

Таблица 1
Распространенные шаблоны для составления заявления о доступности данных

Статус данных и материалов	Пример заявления
Данные открыто доступны в публичном репозитории* с постоянным идентификатором (например, DOI)	Данные, поддерживающие результаты представленного исследования, представлены в открытом доступе в [название репозитория] по адресу http://doi.org/[doi] , номер ссылки [номер ссылки].
Данные открыто доступны в публичном репозитории без постоянного идентификатора	Данные, поддерживающие результаты представленного исследования, представлены в открытом доступе в [название репозитория] по адресу [URL], номер ссылки [номер ссылки].
Данные под эмбарго из-за коммерческих ограничений	Данные, поддерживающие результаты текущего исследования, будут доступны в [название репозитория] по адресу [URL / DOI] после [6-месячного] эмбарго с даты публикации для коммерциализации результатов исследования.

* Ограничений по выбору репозитория нет; авторы могут использовать любой по своему усмотрению, но отдельные журналы могут рекомендовать загрузить данные в определенный репозиторий.

Окончание Таблицы 1

Статус данных и материалов	Пример заявления
Данные доступны по запросу вследствие ограничений конфиденциальности/этики	Данные, поддерживающие результаты текущего исследования, доступны по запросу у соответствующего автора [инициалы]. Данные не доступны для публичного доступа из-за [ограничений, например, содержащейся информации, которая может скомпрометировать конфиденциальность участников исследования].
Данные, сгенерированные на центральном крупномасштабном объекте, доступны по запросу	Исходные данные были сгенерированы на [название объекта]. Производные данные, поддерживающие результаты текущего исследования, доступны у соответствующего автора [инициалы] по запросу.
Данные, подлежащие ограничениям третьих сторон	Данные, поддерживающие результаты текущего исследования, доступны у [третьей стороны]. К данным предоставляется ограниченный доступ, так как они были использованы в данном исследовании по лицензии. Данные доступны у [авторов / по адресу URL] с разрешения [третьей стороны].
Данные доступны в статье или ее дополнительных материалах	Авторы подтверждают, что все данные, поддерживающие результаты текущего исследования, представлены в статье [и / или в дополнительных материалах].
Совместное использование данных не применяется – новые данные не созданы	Доступность данных не применима к представленной статье, так как новые данные не были созданы или проанализированы в этом исследовании.
Данные доступны по запросу у авторов	Данные, поддерживающие результаты данного исследования, доступны у соответствующего автора [инициалы автора] по обоснованному запросу.
Данные, полученные из публичных источников	Данные, поддерживающие результаты текущего исследования, доступны в [название репозитория] по адресу [URL / DOI], номер ссылки [номер ссылки]. Эти данные получены из следующих источников в публичном доступе: [перечислить ресурсы и URL].

ПРИМЕРЫ ПРЕДСТАВЛЕНИЯ
ЗАЯВЛЕНИЯ О ДОСТУПНОСТИ ДАННЫХ
В СОЦИОГУМАНИТАРНЫХ ИССЛЕДОВАНИЯХ

Данные доступны в публичном репозитории:

- «Наборы данных, собранные и/или проанализированные в ходе данного исследования, доступны в репозитории Zenodo по адресу <https://doi.org/10.5281/zenodo.1234567>.»
- «Данные, использованные в этом исследовании, доступны для общего доступа под лицензией Creative Commons в репозитории OpenNeuro: <https://openneuro.org/datasets/ds002345>.»

Данные доступны в репозитории с ограниченным доступом:

- «Данные, собранные исследовательской группой Smith et al., доступны в репозитории UK Data Service по адресу <https://ukdataservice.ac.uk>, но доступ к ним ограничен в соответствии с условиями использования данных.»
- «Исходные данные этого исследования защищены и недоступны из-за законов о конфиденциальности данных. Однако обработанные наборы данных доступны в репозитории ICPSR под кодом досту-

па 98765432 (<https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/98765432>).»

Данные не могут быть представлены в открытом доступе, но предоставляются по запросу у авторов:

- «Данные, поддерживающие выводы данного исследования, доступны от Национального института демографических исследований (INED), но публичный доступ к ним ограничен. Однако данные доступны у авторов по обоснованному запросу и с разрешения INED.»
- «Данные, использованные в этом исследовании, были получены у сторонней организации согласно лицензии и, следовательно, недоступны для публичного доступа. Тем не менее, они могут быть предоставлены по запросу к соответствующему автору.»

Данные представлены в статье или дополнительных материалах:

- «Автор подтверждает, что все данные, сгенерированные или проанализированные в ходе данного исследования, включены в эту статью и ее дополнительные материалы.»
- «Все данные, необходимые для воспроизведения результатов данного исследования, включены в основной текст статьи и приложения.»

Совместное использование данных не применяется:

- «В рамках данного исследования не создавались новые наборы данных, поскольку оно основано на теоретическом и аналитическом подходе.»
- «Данное исследование не связано с генерацией новых данных, поэтому совместное использование данных не применимо.»

КАКАЯ ИНФОРМАЦИЯ СЧИТАЕТСЯ «ДАННЫМИ»?

«Данные» в широком смысле понимаются как любые доказательства или ресурсы, которые необходимы другим для полной оценки обоснованности ваших выводов, а также для проверки или воспроизведения исследования. Это определение включает в себя как необработанные, так и обработанные наборы данных, код и протоколы, а также качественные ресурсы, такие как изображения, аудио, видео, карты, расшифровки интервью, полевые заметки и публичные отчеты. Информация, необходимая для доступа к этим ресурсам, их интерпретации и обработки, также является обязательной.

Заявление о доступности данных должно всесторонне охватывать все типы данных, связанные с представленной в редакцию статьей, а также предоставлять информацию о любом программном обеспечении, использованном для обработки результатов. В зависимости от характера исследования, заявление о доступности данных может включать до четырех разделов:

(1) Исходные данные: данные, полученные от третьих сторон, которые авторы исследования не собирали самостоятельно в рамках исследования, но использовали для анализа. Если исследование опирается на такие данные, необходимо предоставить информацию о том, откуда эти данные были получены и на каких условиях они могут быть использованы.

(2) Основные данные: первичные данные, которые авторы исследования собрали или создали в процессе исследования. Эти данные важны для воспроизведения полученных результатов, поскольку именно они определяют его выводы, и должны быть загружены в онлайн-репозиторий.

(3) Дополнительные данные: включают материалы, которые поддерживают ключевые утверждения статьи, но не являются обязательными для воспроизведения дизайна и анализа исследования. Например, анкеты, дополнительные изображения или таблицы. Как и основные данные, дополнительные данные должны быть загружены в онлайн-репозиторий.

(4) Руководства по отчетности: если для исследования обязательны руководства по отчетности, необходимо включить копию этих руководств в репозиторий. Это гарантирует, что исследование соответствует необходимым стандартам прозрачности и воспроизводимости данных.

Пример Заявления о доступности данных, охватывающего четыре блока:

(1) Исходные данные: «В данном исследовании использовались данные, предоставленные Национальным институтом статистики (National Institute of Statistics), которые не были собраны авторами самостоятельно. Эти данные можно получить по запросу у Национального института статистики по адресу [URL]. Доступ к данным ограничен условиями лицензии, предоставляемой институтом.»

(2) Основные данные: «Первичные данные, собранные в ходе данного исследования, доступны в репозитории Zenodo по адресу <https://doi.org/10.5281/zenodo.1234567>. Эти данные включают все сырые данные, которые использовались для анализа и получения результатов, представленных в статье.»

(3) Дополнительные данные: «Дополнительные материалы, поддерживающие основные выводы данной статьи, включают анкеты, использованные для сбора данных, и дополнительные графики. Эти данные загружены в репозиторий Zenodo и доступны по адресу <https://doi.org/10.5281/zenodo.7654321>.»

(4) Руководства по отчетности: «Для данного исследования использовались стандарты отчетности STROBE (Strengthening the Reporting of Observational Studies in Epidemiology). Копия этих руководств загружена в репозиторий Zenodo и доступна по адресу <https://doi.org/10.5281/zenodo.9876543>.»

ГДЕ В РУКОПИСИ СЛЕДУЕТ РАЗМЕСТИТЬ ЗАЯВЛЕНИЕ О ДОСТУПНОСТИ ДАННЫХ?

Заявление о доступности данных обычно размещается в конце рукописи, перед списком литературы, если выбранный для публикации журнал не указывает иное.

Однако, если в журнале используется процесс двойного слепого рецензирования, важно убедиться, что информация в заявлении о доступности данных не нарушает анонимность авторов. Если в заявлении о доступности данных содержится информация, которая может использоваться для идентификации авторов рукописи, необходимо представить заявление о доступности данных так же, как и любую другую идентифицирующую информацию, связанную с рукописью.

(согласно инструкциям вашего журнала). Например, такую информацию следует указать на титульной странице, представленной как отдельный от рукописи документ.

ЗАЧЕМ ПИСАТЬ ЗАЯВЛЕНИЕ О ДОСТУПНОСТИ ДАННЫХ?

Фактически, заявления о доступности данных занимает всего одно или два предложения, поэтому нет причины, по которой его не следует включать в рукопись. Более того, большинство журналов и финансирующих организаций требуют их для подачи статьи.

Заявление о доступности данных поддерживает (Fisher, 210; Open Science Collaboration, 2015):

(1) Валидацию данных: значимые результаты строятся на надежных данных. Валидация данных – это часто упускаемый из виду шаг, являющийся ключевым для достижения точных результатов. Предоставляя доступ к данным для читателей, автор дает им возможность самостоятельно проверить качество и релевантность исходных данных.

(2) Повторное использование информации: прогресс в науке в конечном итоге основывается на использовании и повторном использовании информации. Эти данные как кирпичи для бесконечно растущего здания, упомянутого во введении. Заявления о доступности данных позволяют легко находить исследовательские данные и проводить их полноценную проверку и валидацию учеными, которым они могут понадобиться для дальнейших исследований. В сочетании с индексацией баз данных, заявления о доступности данных являются очень эффективным инструментом для общего доступа к научным информационным сетям, что облегчает ученым разработку научных проектов и продвижение знаний.

(3) Цитирование: заявления о доступности данных повышают шансы на то, чтобы статья, основанная на конкретных данных, была корректно процитирована другими

ми исследователями. Включая в свою статью заявление о доступности данных, авторы указывают, где и как можно получить доступ к данным, использованным в исследовании, и это увеличивает вероятность того, что другие исследователи смогут найти и использовать эти данные в своих собственных исследованиях. Когда другие ученые используют данные из статьи, они, как правило, ссылаются на эту статью, чтобы подтвердить источник данных. Таким образом, наличие заявления способствует тому, что статья будет чаще цитироваться в других научных работах, что положительно влияет на научную репутацию автора и повышает его индекс цитирования. Корректное цитирование также означает, что другие исследователи могут легко найти оригинальные данные и убедиться в их качестве, что важно для воспроизводимости и проверки результатов, что, в свою очередь, укрепляет доверие к исследованию и его автору.

ЗАКЛЮЧЕНИЕ

Заявление о доступности данных – это важный компонент научной статьи, который обеспечивает прозрачность и воспроизводимость исследований. Независимо от того, доступны ли данные в открытом репозитории или требуют специальных условий для доступа, корректно составленное заявление помогает укрепить доверие к исследованию и его результатам. Следуя предложенным шаблонам и учитывая требования конкретного журнала, авторы могут эффективно и четко донести информацию о доступности своих данных.

Вместе с тем, обязательное внедрение заявления о доступности данных в процесс направления рукописи оригинального исследования в редакцию журнала может столкнуться с определенными препятствиями, такими как этические или юридические ограничения, особенно в случаях, когда данные содержат конфиденциальную информацию. Отсюда, важно разработать гибкие модели доступа к данным исследования, которые учитывали бы эти ограничения, но при этом сохраняли бы максимальную прозрачность исследования. Этой проблематике будут посвящены наши дальнейшие исследования по теме.

REFERENCES

- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533, 452–454. <https://doi.org/10.1038/533452a>
- Cech, T.R., Eddy, S.R., Eisenberg, D., Hersey, K., Holtzman, S.H., Poste, G.H., Raikhel, N.V., Scheller, R.H., Singer, D.B., & Waltham, M.C. (2003). Sharing publication-related data and materials: Responsibilities of authorship in the life sciences. *Plant Physiology*, 132(1), 19–24. <https://doi.org/10.1104/pp.900068>
- Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLOS ONE*, 15(4), e0230416. <https://doi.org/10.1371/journal.pone.0230416>
- Fischer, B.A., & Zigmond, M.J. (2010). The essential nature of sharing in science. *Science and Engineering Ethics*, 16(4), 783–799. <https://doi.org/10.1007/s11948-010-9239-x>

- Hardwicke, T. E., & Ioannidis, J. P. A. (2018). Populating the Data Ark: An attempt to retrieve, preserve, and liberate data from the most highly-cited psychology and psychiatry articles. *PLOS ONE*, 13(8), e0201856. <https://doi.org/10.1371/journal.pone.0201856>
- Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsson, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal *Cognition*. *Royal Society Open Science*, 5(8), 180448. <https://doi.org/10.1098/rsos.180448>
- Krawczyk, M., & Reuben, E. (2012). (Un)Available upon request: Field experiment on researchers' willingness to share supplementary materials. *Accountability in Research*, 19(3), 175–186. <https://doi.org/10.1080/08989621.2012.678688>
- Longo, D. L., & Drazen, J. M. (2016). Data sharing. *The New England Journal of Medicine*, 374(3), 276–277. <https://doi.org/10.1056/NEJMe1516564>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2(3), e308. <https://doi.org/10.1371/journal.pone.0000308>
- Roche, D. G., Kruuk, L. E. B., Lanfear, R., & Binning, S. A. (2015). Public data archiving in ecology and evolution: How well are we doing? *PLOS Biology*, 13(11), e1002295. <https://doi.org/10.1371/journal.pbio.1002295>
- Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8(1), 192. <https://doi.org/10.1038/s41597-021-00981-0>
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLOS ONE*, 10(8), e0134826. <https://doi.org/10.1371/journal.pone.0134826>
- Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., Moore, J., Moyers, B. T., Renaut, S., Rennison, D. J., Veen, T., & Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *The FASEB Journal*, 27(4), 1304–1308. <https://doi.org/10.1096/fj.12-218164>
- Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE*, 8(7), e67332. <https://doi.org/10.1371/journal.pone.0067332>
- Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, 6(11), e26828. <https://doi.org/10.1371/journal.pone.0026828>