

От редакции

Публикации о базах данных в сфере образования

Журнал «Вопросы образования» приглашает авторов исследований в сфере образования к публикации на своих страницах статей и обзоров о собранных ими базах данных.

Публикации нацелены на распространение уникальных данных для использования их в новых исследованиях и развития кооперации ученых вокруг уже собранных эмпирических материалов. Приветствуется описание опыта сбора и организации данных, характерных для этого этапа работы трудностей и способов их преодоления. Мы надеемся, что такие материалы привлекут внимание широкого круга исследователей и будут способствовать генерации новых знаний об образовании.

Исследовательские данные для открытой науки

Расширение возможностей повторного использования исследовательских данных — одно из важнейших условий получения новых знаний. Все больше становится открытых репозиторий и платформ для их хранения, например *European Open Science Cloud*¹, *FIVE project*², запускаются системы поиска баз данных: *Google Dataset*³ и *DataSearch*⁴. Многие журналы просят авторов привязывать к статье исходные датасеты для повторного их использования [Hrynaszkiwicz et al., 2020] — и такая практика способствует широкому распространению результатов оригинального исследования и повышает его цитируемость примерно на четверть [Colavizza, 2020].

Основные барьеры на пути открытия и распространения исследовательских данных порождены нормами и практиками, сложившимися в сообществе ученых. В самом деле, чтобы скрупулезно обработать данные и привести их в вид, доступный для других пользователей, требуется немало усилий. С другой стороны, найти доступные уже собранные данные тоже бывает не просто. Ученые, которые прибегают к использованию готовых

¹ <https://eosc-portal.eu/>

² <http://five.dartmouth.edu/>

³ <https://datasetsearch.research.google.com/>

⁴ <https://datasearch.elsevier.com>

данных, отмечают, что чаще всего узнают про интересные для них материалы из журнальных публикаций [Gregory, 2020]. Они напрямую делают запросы авторам исследований, но в 40% случаев ответы на такие запросы не приходят, а в 20% случаев ответ бывает отрицательный — даже от авторов, публикующихся в *Nature* и *Science* [Tedersoo et al., 2021].

Зачем ученому обрабатывать и выкладывать данные в открытый доступ просто как бонус к опубликованному исследованию? Это требует времени, да и всегда есть опасность, что данные будут использованы в коммерческих целях или потенциальные пользователи не смогут в них разобраться [Walters, 2020]. Работа на общее благо и вклад в развитие науки — вот та мотивация, которая движет авторами, предоставляющими коллегам добытые ими данные.

Публикации о базах данных — шаг к созданию культуры открытой науки. Во-первых, они позволяют авторам подробно описать массив полученных данных, презентовать возможности для их дальнейшего анализа и потенциальные темы для исследований, в том числе совместных. Во-вторых, статья о данных имеет тот же статус, что и любая другая публикация в рецензируемом журнале. В-третьих, такие публикации интенсифицируют коммуникацию внутри сообщества, так как сразу дают сигнал о желании авторов развивать кооперацию вокруг уникального массива данных.

Практика публикации статьей о базах данных получает распространение [Павлов, 2019]. В 2020 г. У. Уолтерсу удалось идентифицировать 13 научных журналов, публикующих статьи с отчетами о данных, соответствующими принципам открытой науки, — в основном это периодика в области биологии, экологии, химии, медицины и здравоохранения [Walters, 2020]. Появляются и журналы, которые открывают специальные рубрики для публикаций о базах данных, например *GigaScience*, *International Journal of Food Contamination*, *Journal of Economics and Statistics* [Walters, 2020], *Strategic Management Journal* [Ethiraj, Gambardella, Helfat, 2017].

Так, *Journal of Economics and Statistics* с 2016 по 2021 г. опубликовал 57 статей о базах данных. Есть среди них и связанные со сферой образования — данные о вкладе университетов в инновационное развитие [Warnecke, 2018], об аспирантах и кандидатах наук в Германии [Lange et al., 2017], описание немецкого национального панельного исследования образования [Fuß, von Maurice, Roßbach, 2016].

Ожидаемый формат статьи о базах данных

Публикации о базах данных по жанру и стилю являются академическими текстами. Их цель не представить ответ на исследовательский вопрос или результаты проверки гипотез, а макси-

мально подробно описать собранную авторами базу данных, так чтобы в дальнейшем ее могли использовать другие исследователи, в том числе в кооперации с авторами базы. Мы ожидаем, что статьи будут включать следующие содержательные блоки:

- краткую справку о контексте создания базы данных, например об исследовательском проекте;
- подробное описание данных, их базовых характеристик и возможностей, которые они предоставляют для анализа;
- сведения о процессе сбора данных и их организации, включая указание на трудности, возникшие при сборе данных, и способы их преодоления, обсуждение любых ранее опубликованных исследований, в которых использовались данные;
- краткий эмпирический анализ с использованием данных для иллюстрации потенциала их применения;
- дискуссию об исследовательских вопросах и задачах, которые могут быть решены с использованием представленной базы данных.

В зависимости от глубины анализа эмпирического материала и обоснованности представленных в дискуссии потенциальных исследовательских задач редакция имеет право дать итоговой публикации статус исследовательской статьи (*research article*) или обзора (*review*).

Под базой данных мы подразумеваем организованный набор эмпирического материала, который может содержать не только количественную информацию, но и текстовую:

- массив уникальных количественных данных, полученных в ходе опросов;
- массивы статистической информации;
- массивы «больших данных», созданных в результате выгрузки из общедоступных ресурсов, в том числе из социальных сетей;
- переформатированный набор существующих данных, который облегчает работу исследователей, например слитые массивы разных баз данных;
- организованную базу транскриптов интервью;
- организованную базу текстовых документов для контент-анализа.

Мы надеемся, что этот первый шаг к публикации информации о базах данных станет вкладом в сплочение сообщества исследователей образования и будет способствовать продвижению в поиске знаний об образовании, основанных на эмпирическом материале.

Литература

1. Павлов Л.П. (2019) Статьи о данных как вид вторичных источников научной информации // Информатизация и связь. № 1. С. 92–95. doi:10.34219/2078-8320-2019-10-1-92-95
2. Colavizza G., Hrynaszkiewicz I., Staden I., Whitaker K., McGillivray B. (2020) The Citation Advantage of Linking Publications to Research Data // PLoS One. Vol. 15. No 4. Art. No e0230416. doi:10.1371/journal.pone.0230416
3. Ethiraj S.K., Gambardella A., Helfat C.E. (2017) Improving Data Availability: A New SMJ Initiative // Strategic Management Journal. Vol. 38. No 11. P. 2145–2146. doi:10.1002/smj.2690
4. Gregory K., Groth P., Scharnhorst A., Wyatt S. (2020) Lost or Found? Discovering Data Needed for Research // Harvard Data Science Review. doi:10.1162/99608F92.E38165EB
5. Hrynaszkiewicz I., Simons N., Hussain A., Grant R., Goudie S. (2020) Developing a Research Data Policy Framework for All Journals and Publishers // Data Science Journal. Vol. 19. No 5. P. 1–15. doi:10.5334/dsj-2020-005
6. Fuß D., von Maurice J., Roßbach H.G. (2016) A Unique Research Data Infrastructure for Educational Research and Beyond: The National Educational Panel Study // Jahrbücher für Nationalökonomie und Statistik. Vol. 236. No 4. P. 517–528. doi:10.1515/jbnst-2015-1021
7. Lange J., Lietz A., Ambrasat J., Tesch J., Wegner A. (2017) The German Doctoral Candidates and Doctorate Holders Study ProFile // Jahrbücher für Nationalökonomie und Statistik. Vol. 237. No 4. P. 349–363. doi:10.1515/jbnst-2015-1037
8. Tedersoo L., Küngas R., Oras E., Köster K., Eenmaa H., Leijen Ä. ... & Sepp T. (2021) Data Sharing Practices and Data Availability upon Request Differ across Scientific Disciplines // Scientific Data. Vol. 8. No 1. Art. No 192. doi:10.1038/s41597-021-00981-0
9. Walters W.H. (2020) Data Journals: Incentivizing Data Access and Documentation within the Scholarly Communication System // Insights. Vol. 33. Art. No 18. doi:10.1629/uksg.510
10. Warnecke C. (2018) New Survey Data on the Role of Universities in the German Regional Innovation System // Jahrbücher für Nationalökonomie und Statistik. Vol. 238. No 6. P. 601–608. doi:10.1515/jbnst-2017-0159