

## Роль репозиториев данных и проблемы этики их использования

**Тихонова Елена Викторовна**

*ФГБОУ ВО «Московский государственный университет пищевых производств»*

*Адрес: 125080, г. Москва, Волоколамское шоссе, д. 11*

*E-mail: etihonova@mgupr.ru*

**Шленская Наталия Марковна**

*ФГБОУ ВО «Московский государственный университет пищевых производств»*

*Адрес: 125080, г. Москва, Волоколамское шоссе, д. 11*

*E-mail: n.shlenskaya@mgupr.ru*

Репозитории данных становятся обязательным инструментом для университетов и научных организаций, не просто расширяя функционал электронных библиотек, но и расширяя возможности формирования электронных архивов. Репозитории дают возможности для загрузки, обмена, распространения и использования исследовательских (полевых, сырых) данных исследователей. Публикация исследовательских данных так же важна, как и публикация самих исследований. Возможность получить доступ к таким данным оптимизирует рецензионный цикл и предоставляет исследователям по всему миру возможность синхронизировать и сравнивать данные. Важность публикация исследовательских (сырых) данных также нашла свое отражение и в возникновении самостоятельного вида публикации – публикация данных (data paper). Цель статьи – проанализировать факторы, которые способствуют / препятствуют распространению «сырых» данных исследований. Ограничивают указанный процесс недостаток информации о публикациях такого типа, отсутствие предыдущего опыта обмена или повторного использования данных, опасения присвоения данных другими исследователями или неправильное истолкование данных. Авторами делается вывод о зачаточном формировании культуры распространения и обмена данными и о необходимости выстраивания механизмов формирования указанной культуры.

**Ключевые слова:** репозитории, открытые данные, сырые данные, обмен данными, хранение

В последнее время наблюдается рост числа сетевых архивов открытого доступа (репозиториев) в мире. Онлайн репозитории исследовательских данных – это крупные инфраструктуры баз данных, созданные для управления, совместного использования, доступа и архивирования наборов данных исследователей (Земсков, 2016). Открытые данные – это общедоступная информация, размещенная в интернете для неоднократного, свободного и бесплатного использования в виде машиночитаемых систематизированных данных, в формате, позволяющем их автоматизированную обработку (Draxl & Scheffler, 2019). Активно развиваются два основных типа систем открытых данных – сетевые хранилища (репозитории – «зелёный» вариант открытых данных) и журналы открытого доступа («золотой» вариант). Под открытыми данными понимаются не только опубликованные статьи в журналах открытого доступа, но и публикация необработанных наборов данных (Stieglitz et al., 2020). Очевидно, что традиционные формы научной коммуникации

трансформируются, обуславливая и тенденцию к снижению барьеров на пути распространения результатов научных исследований в формате репозиториев данных (Чадин, 2018; Stieglitz et al, 2020).

Содержание репозиториев шире традиционной библиотеки. Если для библиотек в первую очередь характерны книжные, журнальные и газетные коллекции, то для репозиториев типичен широкий набор данных, включающий «незавершённые комплекты» – части книг, статьи из журналов, отдельные сообщения на конференциях, изображения, наборы данных, в особенности те, где имеются метаданные или документы, достаточные для повторного использования базовых массивов (Земсков, 2016). Среди устоявшихся научных репозиториев различают четыре основных типа: тематические, исследовательские, национальные и институциональные<sup>1</sup>.

Различаются следующие виды данных: первичные / сырые данные всегда получены в результате

<sup>1</sup> Михеенков, А. В. (2018). Самоархивирование и открытые репозитории: методические рекомендации. М.: Ваше цифровое издательство.

исследования. Эти данные не прошли какую-либо обработку и представлены в своем первичном виде. Вторичные данные получаются после анализа первичных данных, в том числе статистического анализа, и эти данные публикуются как результаты исследований. Особый интерес представляют необработанные данные исследователей. Редакторы часто оценивают статистические результаты на основе таблиц и графиков в своих журнальных материалах, не обращаясь к необработанным данным. Поэтому не всегда ошибки в статистических отчетах очевидны, но для интерпретации результатов это всегда имеет значение (Habibzadeh, 2017).

Изначально совместное использование необработанных данных было задумано в основном для облегчения репликации экспериментов и улучшения анализа данных. Однако появление искусственного интеллекта и инструментов машинного обучения делает обмен научными данными еще более важным, поскольку новые (неожиданные для исходных исследовательских групп) корреляции могут возникать в связанных и общих наборах данных (Aranda, 2018). Мотивирующими факторами выступают возможности для совместного использования, получения цельного исследования, возможности воспроизводимости данных и их повторное использование. Первичные (сырые) данные могут быть отправлены издателю журнала одновременно с рукописью статьи, также пакет данных может быть опубликован независимо в хранилище данных - репозитории. Критерии соответствия данных для включения в репозиторий, указывается редакцией журнала. Читатели также могут бесплатно загрузить программное обеспечение и открыть файл с первичными данными, просматривая, обрабатывая и повторно анализируя первичные данные исследования, пользуясь всеми функциями, доступными в программном пакете. Это фактически делает такие данные открытыми для просмотра и воспроизведения независимо от доступа к лицензированным программным приложениям. Влияние такого представления первичных исследовательских данных, связанных со статьями, заключается в том, что они становятся доступными для детального просмотра широкому кругу заинтересованных сторон, таких как рецензенты, редакторы, издатели, коллеги, читатели, создатели технологий искусственного интеллекта и систем машинного обучения (Barba et al., 2019).

Загрузка и обмен первичными данными имеет много преимуществ. Совместное использование и повторное использование первичных наборов данных исследований может повысить эффективность и качество исследований. Необработанные данные могут быть использованы для изучения связанных или новых гипотез, особенно в сочетании с другими доступными наборами данных. Такие данные необходимы для разработки и проверки методов исследования, анализа и работы программного обеспечения. Более широкое научное сообщество также выигрывает: обмен исследовательскими данными поощряет различные точки зрения, помогает выявлять ошибки, препятствовать мошенничеству, полезен для обучения новых исследователей и повышает эффективность использования финансовых и демографических ресурсов, избегая дублирования сбора данных.

Обмен индивидуальными данными позволяет проводить независимую проверку результатов исследований для обеспечения их надежности и воспроизводимости (Bull, Roberts, & Parker, 2015):

- (1) делает исследования более контролируемыми и воспроизводимыми. Существуют примеры, когда исследования были фальсифицированы или просто не проводились. Если бы данные исследований были размещены в репозитории, указанные проблемы обнаружались бы намного раньше и, следовательно, имели бы менее разрушительное воздействие на научное знание<sup>2</sup>.
- (2) ограничивает число опубликованных статей, не вносящих вклад в существующее знание в области. Сравнение данных позволило бы выявить отсутствие научной новизны. А внимательное предварительное знакомство с пакетами данных и вовсе снизило бы вероятность проведения бесполезных исследований в ситуации, когда опубликованные «сырые» данные бы были общедоступными. Последнее предопределило бы возможность сэкономить финансовые ресурсы и использовать их для проведения новых исследований, что является важной мотивацией для фондов, учреждений и исследователей.
- (3) позволяет объединить наборы данных в случаях, когда исследователи получили аналогичные данные. Подобный подход позволил бы сократить статистическую погрешность анализа и, следовательно, повысить шансы обнаружения реальных

<sup>2</sup> Михайлов, А. С., Филатов, М. М. (2021). Нетрадиционные форматы публикаций – Data paper & Preprints и референс-менеджер Mendeley [Презентация]. Elsevier B. V., RELX Group PLC. [https://elsevierscience.ru/files/pdf/Нетрадиционные\\_форматы\\_публикаций\\_и\\_Референс-менеджер\\_Mendeley.pdf](https://elsevierscience.ru/files/pdf/Нетрадиционные_форматы_публикаций_и_Референс-менеджер_Mendeley.pdf) (дата обращения: 10.03.2021).

эффектов. Более того, совместное использование данных позволяет другим исследователям как из той же, так и из других областей научного знания применять свой опыт и проводить новый анализ, тем самым способствовать многопрофильным исследованиям

Совместное использование тщательно отобранных научных данных, согласно проведенному анализу, будет иметь еще более важные преимущества в среднесрочной перспективе, в разрезе валидации их надежности и воспроизводимости (Aranda, 2018). Отсюда, и интерес научного сообщества не только к их размещению в репозиториях, но и к опубликованию в формате статей-данных (data paper), который уже предлагают крупные научные журналы (Kim, 2020).

Стремясь реализовать преимущества обмена «сырыми» данными, спонсоры, издатели, научные общества и отдельные исследовательские группы разработали инструменты, ресурсы и политику, чтобы побудить исследователей сделать свои данные общедоступными. Например, некоторые журналы требуют представления подробных наборов биомедицинских данных в общедоступные базы данных в качестве условия публикации<sup>3</sup>. С 2003 года Национальные институты здравоохранения в США потребовали гарантировать обмен данными между исполнителями всех крупных грантов (Stieglitz, et. al., 2020). Исследователей призывают к ответственному обмену данными и их повторному использованию.

Основная задача заключается в обеспечении защиты конфиденциальности участников во время вторичного использования данных (Castellani, 2013). Особенно это касается публикаций биомедицинского профиля. Процессы деидентифицирования данных должны быть не только надежными, но и соразмерными для того, чтобы сохранить их полезность (Bull et. al., 2015). Данные, раскрытие которых может нести риски для субъекта, к которому они относятся, должны быть защищены. В отношении физического лица к таким данным можно отнести, например, информацию об этническом или расовом происхождении, состоянии здоровья, политических и религиозных убеждениях, а для коммерческого – информацию о производственных процессах и особенностях сделок. Термин «чувствительные данные» не имеет сегодня четкого определения в российском законода-

тельстве, но уже определен в рамках европейского регламента по защите данных GDPR (General Data Protection Regulation). Существует достаточно обширный пласт информации, раскрытие которой может нести риски для субъекта, к которому она относится.<sup>4</sup>

Хотя репозитории экспериментальных и полевых данных, безусловно, имеют фундаментальное значение для поддержки прогресса исследований и распространения знаний, исследовательское сообщество все еще, кажется, несколько неохотно поддерживает такую инициативу. Анализ публикаций показывают, что многие ученые избегают делиться своими личными данными напрямую с какой-либо третьей стороной для академических исследований или коммерческого анализа, потому что личные данные содержат личную или конфиденциальную информацию (Parker & Bull, 2015; Vidal-Infer, Aleixandre-Benavent, Lucas-Dominguez, & Sixto-Costoya, 2019). Обеспечение баланса между полезностью больших данных и конфиденциальностью пользователей является жизненно важным вопросом для академических кругов и бизнеса. Авторы предлагают схемы сбора необработанных данных, сохраняющую конфиденциальность (Liu, Wang, Wang, Xia, & Xu, 2019).

Признавая преимущества прозрачной отчетности, многие рецензируемые журналы требуют, чтобы их авторы были готовы поделиться своими сырыми, необработанными данными с другими учеными и / или указывать наличие исходных данных в опубликованных статьях. Но пока разработано еще достаточно мало инструкций о том, как данные должны быть подготовлены для публикации или обмена (Hrynaskiewicz, Norton, Vickers, & Altman, 2010). Существуют также опасения, что стимулы к исследовательской деятельности для новых исследований могут ослабнуть, если вторичные пользователи данных смогут свободно пользоваться усилиями тех, кто уже собрал данные (Castellani, 2013). Многие исследователи озвучивают желание получать вознаграждение за размещение данных репозиториях (Aranda, 2018).

Исследователи опасаются, что данные могут быть неправильно истолкованы, или станут предметом предвзятого или некорректно спланированного исследования (Aleixandre-Benavent, Vidal-Infer, Alonso-Arroyo, Peset, & Sapena, 2020). Решением этой проблемы могло бы стать создание коллабо-

<sup>3</sup> ICMJE. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (2019). URL: <http://www.icmje.org/recommendations/> (дата обращения: 12.03.2021).

<sup>4</sup> Аналитический доклад. Работа с открытыми данными: особенности публикации и использования в Российском правовом поле. (2020). URL: <https://www.infoculture.ru/wp-content/uploads/2020/11/OpenDataReview.pdf> (дата обращения: 15.03.2021).

рации соответствующих заинтересованных сторон для проведения консультации «первичными» собственниками данных (Manju & Buckley, 2012). Также одним из препятствий на пути публикации данных является дополнительная работа по подготовке таких данных к публикации. Наиболее предпочтительным способом обмена исследовательскими данными на сегодняшний день считается загрузка наборов данных в онлайн-репозитории данных и размещение ссылки на репозиторий в авторской статье (Зельдина, 2019). Данные публикуются в репозиториях данных как дополнительный материал к исходной статье. Цитирование таких наборов данных является инструментом соблюдения этических норм и авторских прав. Во многих журналах Elsevier можно загружать и хранить данные в хранилище Mendeley Data как часть процесса подачи рукописи. Можно также загрузить данные непосредственно в репозиторий. В каждом случае такие данные получают идентификатор цифрового объекта (DOI), что делает их независимо цитируемыми, и их можно связать с любой связанной статьей на ScienceDirect, что облегчит читателям их поиск и повторное использование<sup>5</sup>. Публикация исследовательских данных увеличивает количество цитирований автора (Lichtfouse, 2018).

Пока еще не сложилось четкого представления о том, какие категории авторов открыто делятся необработанными исследовательскими данными, а какие нет, и как такие инициативы коррелируют с высокими темпами обмена данными. Для выявления закономерностей, в соответствии с которыми исследователи архивируют в репозиториях свои необработанные наборы данных после публикации исследования были использованы библиометрические методы (Piwowar, 2011). Было обнаружено, что число связанных с публикациями наборов данных в репозиториях увеличилось. Примерно 45% последних исследователей в области генетики сделали свои сырые данные общедоступными. Факторный анализ выявил 15 факторов, описывающих авторство, финансирование, учреждение, публикацию и доменную среду размещения данных. Авторы более склонны обмениваться исходными данными, если у них уже был предыдущий опыт обмена или повторного использования данных, а также если их исследование было опубликовано в журнале открытого доступа или журнале с относительно сильной политикой обмена данными, или если исследование финансировалось за счет большого количества грантов. Также были обнаружены другие законо-

мерности. Уровень обмена данными относительно низок для исследований с низким цитированием и коэффициентом сотрудничества учреждения, на базе которого проведено исследования, с другими исследовательскими организациями. Более высокий уровень характерен для авторов исследований из учреждений, имеющих высокий уровень цитирования и активно участвующих в научных коллаборациях. Большинство людей в академических кругах готово к обмену данными, когда предполагаемые преимущества перевешивают недостатки (Stieglitz et al., 2020).

Культура обмена исследовательскими данными в России и мире пока находится на этапе своего формирования, а данные наименее доступны в тех областях, где они могут оказать наибольшее влияние на развитие современного научного знания. Чтобы обеспечить широкую доступность исследовательских данных, журналам необходимо настойчиво запрашивать их публикацию, а не делать это в рекомендательном порядке (Gabelica, Cavar, Puljak, 2019). Интересно, что многие авторы опасаются публиковать свои исследовательские данные, чтобы не потерять собственные права на них или же «не подарить» свое открытие недобросовестным конкурентам (Aleixandre-Benavent, et.al., 2020). Подобные заблуждения скорее свидетельствуют о недостаточной осведомленности автора о сути публикационной и научно-исследовательской этики. Ибо именно публикация данных способна обеспечить однозначное авторство и запатентировать первенство исследования в той или иной области. Представляется, что роль отчуждающих факторов играет не только недостаточная информированность исследователей, но и несформированность компетенций, необходимых для эффективного обмена данными и их эффективного использования.

В России процесс «открытия» национальных данных и их интеграция в международные проекты находится на начальном этапе и активно развивается. Все больше исследователей видят преимущества и перспективы в размещении данных исследований в открытом доступе. Вместе с тем, очевидна необходимость целенаправленного развития компетенций исследователей, которые необходимы для эффективного обмена данными и совместного их использования. Речь идет не только о процессе обмена данными, но и о навыках унификации подходов к кодированию данных, их обработке, верификации, объединения с новыми наборами данных и др.

<sup>5</sup> Sharing research data. URL: <https://www.elsevier.com/authors/tools-and-resources/research-data> (дата обращения: 14.03.2021).

## Литература

- Зельдина, М. М. (2019). *Политика обмена исследовательскими данными*. М.: ООО «Ваше цифровое издательство». <https://doi.org/10.24108/978-5-6040408-7-4>
- Земсков, А. И. (2016). Открытый доступ и открытые архивы информации. *Научные и технические библиотеки*, 6, 41-61. [https://www.gpntb.ru/ntb/ntb/2016/6/NTB6\\_2016\\_%D0%905\\_3.pdf](https://www.gpntb.ru/ntb/ntb/2016/6/NTB6_2016_%D0%905_3.pdf)
- Чадин, И. (2018). Zenodo и GBIF: инструменты для публикации наборов первичных данных. *Вестник Института биологии Коми НЦ УрО РАН*, 3, 34-36. [https://doi.org/10.31140/j.vestnikib.2018.3\(205\).5](https://doi.org/10.31140/j.vestnikib.2018.3(205).5)
- Aleixandre-Benavent, R., Vidal-Infer, A., Alonso-Arroyo, A., Peset, F., & Sapena, A. F. (2020). Research data sharing in Spain: Exploring determinants, practices, and perceptions. *MDPI*, 5(2), 1-14. <https://doi.org/10.3390/data5020029>
- Aranda, M. A. G. (2018). Sharing powder diffraction raw data: Challenges and benefits. *Journal of Applied Crystallography*, 51(6), 1739-1744. <https://doi.org/10.1107/S160057671801556X>
- Barba, A., Dominguez, S., Cobas, C., Martinsen, D. P., Romain, C., Rzepa, H. S., & Seoane, F. (2019). Workflows Allowing Creation of Journal Article Supporting Information and Findable, Accessible, Interoperable, and Reusable (FAIR)-enabled publication of spectroscopic data. *ACS Omega*, 4(2), 3280-3286. <https://doi.org/10.1021/acsomega.8b03005>
- Bull, S., Roberts, N., & Parker, M. (2015). Views of ethical best practices in sharing individual-level data from medical and public health research: A systematic scoping review. *Journal of Empirical Research on Human Research Ethics: JERHRE*, 10(3), 225-238. <https://doi.org/10.1177/1556264615594767>
- Castellani, J. (2013). Are clinical trial data shared sufficiently today? Yes. *British Medical Journal*, 347(7916), f1881. <https://doi.org/10.1136/bmj.f1881>
- Draxl, C., & Scheffler, M. (2019). The NOMAD laboratory: From data sharing to artificial intelligence. *Journal of Physics: Materials*, 2(3), 036001. <https://doi.org/10.1088/2515-7639/ab13bb>
- Gabelica, M., Cavar, J., & Puljak L. (2019) Authors of trials from high-ranking anesthesiology journals were not willing to share raw data. *Journal of Clinical Epidemiology*, 109, 111-116. <https://doi.org/10.1016/j.jclinepi.2019.01.012>
- Habibzadeh, F. (2017). Statistical data editing in scientific articles. *Journal of Korean medical science*, 32(7), 1072-1076. <https://doi.org/10.3346/jkms.2017.32.7.1072>
- Hrynaskiewicz, I., Norton, M. L., Vickers, A. J., & Altman, D. G. (2010). Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. *British Medical Journal (Clinical Research Ed.)*, 340, c181. <https://doi.org/10.1136/bmj.c181>
- Kim, J. (2020). An analysis of data paper templates and guidelines: types of contextual information described by data journals. *Science Editing*, 7(1), 16-23. <https://doi.org/10.6087/kcse.185>
- Lichtfouse, E. (2018) Publishing science without results and recycling research. *Environmental Chemistry Letters*, 16, 1-4. <https://doi.org/10.1007/s10311-017-0690-7>
- Liu, Y.-N., Wang, Y.-P., Wang, X.-F., Xia, Z., & Xu, J. F. (2019) Privacy-preserving raw data collection without a trusted authority for IoT. *Computer Networks*, 148, 340-348. <https://doi.org/10.1016/j.comnet.2018.11.028>
- Manju, R., & Buckley, B. S. (2012). Systematic archiving and access to health research data: Rationale, current status and way forward. *Bulletin of the World Health Organization*, 90, 932-939. <https://doi.org/10.2471/BLT.12.105908>
- Parker, M., & Bull, S. (2015). Sharing public health research data: Towards the development of ethical data-sharing practice in low- and middle-income settings. *Journal of Empirical Research on Human Research Ethics*, 10(3), 217-224. <https://doi.org/10.1177/1556264615593494>
- Piwowar, H. A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PloS One*, 6(7), e18657. <https://doi.org/10.1371/journal.pone.0018657>
- Stieglitz, S., Wilms, K., Mirbabaie, M., Hofeditz, L., Brenger, B., Lopez, A., & Rehwald, S. (2020). When are researchers willing to share their data? – Impacts of values and uncertainty on open data in academia. *PLoS One*, 15(7), e0234172. <https://doi.org/10.1371/journal.pone.0234172>
- Vidal-Infer, A., Aleixandre-Benavent, R., Lucas-Dominguez, R., & Sixto-Costoya, A. (2019). The availability of raw data in substance abuse scientific journals. *Journal of Substance*, 24(1), 36-402. <https://doi.org/10.1080/14659891.2018.1489905>

# The Role of Data Repositories and Issues of Ethical Use

**Elena V. Tikhonova**

*Moscow State University of Food Production  
11, Volokolamskoe highway, Moscow, 125080, Russian Federation  
E-mail: etihonova@mgupp.ru*

**Nataliya M. Shlenskaya**

*Moscow State University of Food Production  
11, Volokolamskoe highway, Moscow, 125080, Russian Federation  
E-mail: n.shlenskaya@mgupp.ru*

Data repositories are becoming an indispensable tool for universities and scientific organizations, not only expanding the functionality of electronic libraries, but also expanding the possibilities of forming electronic archives. Repositories provide opportunities for downloading, sharing, distributing and using research (field, raw) data from researchers. The publication of research data is just as important as the publication of the research itself. The ability to access such data streamlines the review cycle and provides researchers around the world with the ability to synchronize and compare data. The importance of the publication of research (raw) data is also reflected in the emergence of an independent type of publication - the publication of data (data paper). The purpose of the article is to analyze the factors that facilitate / hinder the dissemination of "raw" research data. This process is limited by a lack of information about this type of publication, lack of previous experience in sharing or reusing data, fears of misappropriation of data by other researchers, or misinterpretation of data. The authors conclude about the embryonic formation of a culture of dissemination and exchange of data and the need to build mechanisms for the formation of this culture.

**Keywords:** repositories, open data, raw data, data exchange, storage

## References

- Chadin, I. (2018). Zenodo i GBIF: instrumenty dlya publikatsii naborov pervichnykh dannykh [Zenodo and GBIF: Tools for Publishing Primary Datasets]. *Vestnik Instituta biologii Komi NTs UrO RAN [Bulletin of the Institute of Biology of the Komi Scientific Center of the Ural Branch of the Russian Academy of Sciences]*, 3, 34-36. [https://doi.org/10.31140/j.vestnikib.2018.3\(205\).5](https://doi.org/10.31140/j.vestnikib.2018.3(205).5)
- Zel'dina, M. M. (2019). *Politika obmena issledovatel'skimi dannymi [Data Sharing Policy]*. Moscow: OOO «Vashe tsifrovoye izdatel'stvo». <https://doi.org/10.24108/978-5-6040408-7-4>
- Zemskov, A. I. (2016). Otkrytyi dostup i otkrytye arkhivy informatsii [Open access and open archives of information]. *Nauchnye i tekhnicheskie biblioteki [Scientific and technical libraries]*, 6, 41-61.
- Aleixandre-Benavent, R., Vidal-Infer, A., Alonso-Arroyo, A., Peset, F., & Sapena, A. F. (2020). Research data sharing in Spain: Exploring determinants, practices, and perceptions. *MDPI*, 5(2), 29. <https://doi.org/10.3390/data5020029>
- Aranda, M. A. G. (2018). Sharing powder diffraction raw data: Challenges and benefits. *Journal of Applied Crystallography*, 51(6), 1739-1744. <https://doi.org/10.1107/S160057671801556X>
- Barba, A., Dominguez, S., Cobas, C., Martinsen, D. P., Romain, C., Rzepa, H. S., & Seoane, F. (2019). Workflows Allowing Creation of Journal Article Supporting Information and Findable, Accessible, Interoperable, and Reusable (FAIR)-enabled publication of spectroscopic data. *ACS Omega*, 4(2), 3280-3286. <https://doi.org/10.1021/acsomega.8b03005>
- Bull, S., Roberts, N., & Parker, M. (2015). Views of ethical best practices in sharing individual-level data from medical and public health research: A systematic scoping review. *Journal of Empirical Research on Human Research Ethics: JERHRE*, 10(3), 225-238. <https://doi.org/10.1177/1556264615594767>
- Castellani, J. (2013). Are clinical trial data shared sufficiently today? Yes. *British Medical Journal*, 347(7916), f1881. <https://doi.org/10.1136/bmj.f1881>
- Draxl, C., & Scheffler, M. (2019). The NOMAD laboratory: From data sharing to artificial intelligence. *JPhys Materials*, 2(3), 036001. <https://doi.org/10.1088/2515-7639/ab13bb>
- Gabelica, M., Cavar, J., & Puljak L. (2019) Authors of trials from high-ranking anesthesiology jour-

- nals were not willing to share raw data. *Journal of Clinical Epidemiology*, 109, 111-116. <https://doi.org/10.1016/j.jclinepi.2019.01.012>
- Habibzadeh, F. (2017). Statistical data editing in scientific articles. *Journal of Korean medical science*, 32(7), 1072-1076. <https://doi.org/10.3346/jkms.2017.32.7.1072>
- Hrynaszkiwicz, I., Norton, M. L., Vickers, A. J., & Altman, D. G. (2010). Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. *British Medical Journal (Clinical Research Ed.)*, 340, 1-5. <https://doi.org/10.1136/bmj.c181>
- Kim, J. (2020). An analysis of data paper templates and guidelines: types of contextual information described by data journals. *Science Editing*, 7(1), 16-23. <https://doi.org/10.6087/kcse.185>
- Lichtfouse, E. (2018) Publishing science without results and recycling research. *Environmental Chemistry Letters*, 16, 1-4. <https://doi.org/10.1007/s10311-017-0690-7>
- Liu, Y.-N., Wang, Y.-P., Wang, X.-F., Xia, Z., & Xu, J. F. (2019) Privacy-preserving raw data collection without a trusted authority for IoT. *Computer Networks*, 148, 340-348. <https://doi.org/10.1016/j.comnet.2018.11.028>
- Manju, R., & Buckley, B. S. (2012). Systematic archiving and access to health research data: Rationale, current status and way forward. *Bulletin of the World Health Organization*, 90, 932-939. <https://doi.org/10.2471/BLT.12.105908>
- Parker, M., & Bull, S. (2015). Sharing public health research data: Towards the development of ethical data-sharing practice in low- and middle-income settings. *Journal of Empirical Research on Human Research Ethics*, 10(3), 217-224. <https://doi.org/10.1177/1556264615593494>
- Piwowar, H. A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PloS One*, 6(7), e18657. <https://doi.org/10.1371/journal.pone.0018657>
- Stieglitz, S., Wilms, K., Mirbabaie, M., Hofeditz, L., Brenger, B., Lopez, A., & Rehwald, S. (2020). When are researchers willing to share their data? – Impacts of values and uncertainty on open data in academia. *PLoS One*, 15(7), e0234172. <https://doi.org/10.1371/journal.pone.0234172>
- Vidal-Infer, A., Aleixandre-Benavent, R., Lucas-Dominguez, R., & Sixto-Costoya, A. (2019). The availability of raw data in substance abuse scientific journals. *Journal of Substance*, 24(1), 36-402. <https://doi.org/10.1080/14659891.2018.1489905>