



<https://doi.org/10.20538/1682-0363-2023-2-182-184>

Русскоязычный репозиторий открытых клинических данных SibMED Data Clinical Repository

**Куликов Е.С., Федорова О.С., Толмачев И.В., Рязанцева У.В., Вражнов Д.А.,
Губанов А.В., Нестерович С.В., Шмырина А.А.**

*Сибирский государственный медицинский университет
Россия, 634050, г. Томск, Московский тракт, 2*

РЕЗЮМЕ

В эпоху глобальной цифровизации отрасли здравоохранения, благодаря развитию медицинских информационных систем, все большую актуальность и значимость приобретают открытые медицинские данные. Их использование востребовано для научных исследований и технологических проектов в сфере искусственного интеллекта. ФГБОУ ВО СибГМУ Минздрава России впервые в России инициировал создание первого русскоязычного репозитория клинических данных SibMED Data Clinical Repository (<https://dataset.ssmu.ru/>). В статье описывается структура, функции репозитория, а также перспективы его использования.

Ключевые слова: репозиторий открытых клинических данных, открытые данные, медицинские информационные системы, цифровые технологии в клинической медицине, искусственный интеллект, машинное обучение, открытая наука

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Источник финансирования. Авторы заявляют об отсутствии финансирования при проведении исследования.

Для цитирования: Куликов Е.С., Федорова О.С., Толмачев И.В., Рязанцева У.В., Вражнов Д.А., Губанов А.В., Нестерович С.В., Шмырина А.А. Русскоязычный репозиторий открытых клинических данных SibMED Data Clinical Repository. *Бюллетень сибирской медицины*. 2023;22(2):182–184. <https://doi.org/10.20538/1682-0363-2023-2-182-184>.

Russian-language repository of the open clinical data “SibMED Data Clinical Repository”

**Kulikov E.S., Fedorova O.S., Tolmachev I.V., Ryazantseva U.V., Vrazhnov D.A.,
Gubanov A.V., Nesterovich S.V., Shmyrina A.A.**

*Siberian State Medical University
2, Moscow Trakt, Tomsk, 634050, Russian Federation*

ABSTRACT

The global digitalization has become one of the most significant challenges in the field of medicine and healthcare. The rapid development of digital technologies determines the growing demand for constant access to big data in real time. Their use is in need for research and technological projects in the sphere of artificial intelligence technologies.

✉ Федорова Ольга Сергеевна, fedorova.os@ssmu.ru

development. Siberian State Medical University developed the first Russian-language repository of clinical data “SibMed Data Clinical Repository” (<https://dataset.ssmu.ru/>). The article describes the structure, functions of the repository, and perspectives of its use.

Keywords: open clinical data repository, open data, medical information systems, digital health, artificial intelligence, machine learning, open science

Conflict of interest. The authors declare the absence of obvious or potential conflicts of interest related to the publication of this article.

Source of financing. The authors state that they received no funding for the study.

For citation: Kulikov E.S., Fedorova O.S., Tolmachev I.V., Ryazantseva U.V., Vrazhnov D.A., Gubanov A.V., Nesterovich S.V., Shmyrina A.A. Russian-language repository of the open clinical data “SibMED Data Clinical Repository”. *Bulletin of Siberian Medicine*. 2023;22(2):182–186. <https://doi.org/10.20538/1682-0363-2023-2-182-186>.

Цифровизация стала одним из наиболее значимых вызовов отрасли медицины и здравоохранения в XXI в. Внедрение цифровых решений в современной клинической практике позволяет улучшать качество и эффективность медицинской помощи, снижать затраты и обеспечивать безопасность пациентов. Стремительное развитие телемедицины, технологий на основе искусственного интеллекта определяют растущий спрос на постоянный доступ к большим данным в режиме реального времени [1, 2].

Институт статистических исследований и экономики знаний НИУ ВШЭ с помощью системы анализа больших данных iFORA выявил цифровые технологии, наиболее востребованные в медицине и здравоохранении. Так, среди лидирующих по значимости на рынке технологий и в исследованиях входят базы электронных медицинских записей [3]. Большие данные, накопленные в медицинских информационных базах, активно используются при создании систем поддержки принятия клинических решений, при разработке IT-технологий [4–6].

Одним из масштабных проектов, на основе которых разрабатываются такие алгоритмы, является открытая база медицинских данных MIMIC-IV. База разработана при сотрудничестве Медицинского центра Бет Исраэль Диаконисс и Массачусетского технологического института [7]. Такие данные являются широко востребованными, однако из-за расхождения систем здравоохранения различных стран использование MIMIC довольно ограничено [8].

Для развития направления открытых медицинских данных требуется разработка формализованных алгоритмов агрегации и хранения датасетов, обеспечение информационной безопасности [9, 10].

ФГБОУ ВО СибГМУ Минздрава России впервые в России инициировал создание первого русскоязычного репозитория клинических данных SibMED Data Clinical Repository (<https://dataset.ssmu.ru/>). Проект SibMed Clinical Data Repository включает формирование цифровой инфраструктуры для хранения и быстрого доступа к медицинским данным, а также образовательную программу, направленную на формирование компетенций по работе со структурированными и неструктурированными биомедицинскими данными для решения исследовательских и технологических задач.

Репозиторий объединяет базы данных амбулаторной и стационарной службы многопрофильных клиник Сибирского государственного медицинского университета, содержит анонимизированную медицинскую информацию и непрерывно пополняется. Данные деперсонифицированы в соответствии с требованиями федерального законодательства и содержат информацию текстового, числового характера.

SibMED Data Clinical Repository рекомендован к использованию как для исследователей, специалистов в области data science, машинного обучения, управления здравоохранением, предпринимателей, так и для организаций-разработчиков цифровых решений в сфере здравоохранения. Задачи, которые решает SibMed Clinical Data Repository, включают: разработку и тестирование алгоритмов в сфере здравоохранения, создание продуктов для бизнеса и медицины, аналитику проведенных лечебных и диагностических мероприятий, внедрение технологий искусственного интеллекта, разработку программ лечения.

Открытость медицинских данных, доступных благодаря этому проекту, будет способствовать

созданию новых решений для медицины и здравоохранения, улучшению качества жизни и обеспечению более доступной медицинской помощи.

СПИСОК ИСТОЧНИКОВ

1. Kong H.J. Managing Unstructured Big Data in Healthcare System. *Healthcare Informatics Research*. 2019;25 (1):1–2. DOI: 10.4258/hir. 2019.25.1.1.
2. Мохначева Т., Моногарова Ю., Варакина Ж. Вовлеченность организаторов здравоохранения в процесс цифровизации здравоохранения. *Социальные аспекты здоровья населения*. 2023;(1):2–20. DOI: 10.21045/2071-5021-2023-69-1-11.
3. Гребенюк А.Ю. Топ-10 цифровых решений в медицине и здравоохранении. URL: <https://issek.hse.ru/news/691544400.html>
4. Карпов О.Э., Субботин С.А., Шишканов Д.В. Использование медицинских данных для создания систем поддержки принятия врачебных решений. *Врач и информационные технологии*. 2019;(2).
5. Аветисян М.С., Егоров К.С., Кох В.Н. и др. Разработка алгоритма поиска клинически однородных пациентов по слабоструктурированным текстовым данным электронной медицинской карты онкологического профиля. *Врач и информационные технологии*. 2019;(3).
6. Гусев А.В., Зингерман Б.В., Тюфилин Д.С., Зинченко В.В. Электронные медицинские карты как источник данных реальной клинической практики. *MyRWD*. 2022;(2):8–20.
7. Johnson A.E.W., Bulgarelli L., Shen L. et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci. Data*. 2023;10(1):1. DOI: 10.1038/s41597-022-01899-x.
8. Giesa N., Heeren P., Klopfenstein S., Flint A., Agha-Mir-Salim L., Poncette A. et al. MIMIC-IV as a clinical data schema. *Stud. Health Technol. Inf*. 2022;294:559–560. DOI: 10.3233/SHTI220522.
9. Zhang J., Symons J., Agapow P., Teo J.T., Paxton C.A., Abdi J. et al. Best practices in the real-world data life cycle. *PLOS Digit. Health*. 2022;1(1). DOI: 10.1371/journal.pdig.0000003.
10. Буров В., Бегтин И., Ганеева Э. Работа с открытыми данными: особенности публикации и использования в российском правовом поле 18.11.2020 [Аналитический доклад]. Информационная культура. URL: <https://www.infoculture.ru/wp-content/uploads/2020/11/OpenDataReview.pdf>

Информация об авторах

Куликов Евгений Сергеевич – д-р мед. наук, доцент, ректор, СибГМУ, г. Томск, kulikov.es@ssmu.ru, <https://orcid.org/0000-0002-0088-9204>

Федорова Ольга Сергеевна – д-р мед. наук, доцент, зав. кафедрой факультетской педиатрии с курсом детских болезней лечебного факультета, СибГМУ, г. Томск, fedorova.os@ssmu.ru, <http://orcid.org/0000-0002-7130-9609>

Толмачев Иван Владиславович – канд. мед. наук, вед. науч. сотрудник научно-образовательной лаборатории «Бионические цифровые платформы», СибГМУ, г. Томск, ivantolm@mail.ru, <http://orcid.org/0000-0002-2888-5539>

Рязанцева Ульяна Вячеславовна – аналитик, научное управление, СибГМУ, г. Томск, ryazantseva.uv@ssmu.ru, <https://orcid.org/0000-0001-9292-3969>

Вражнов Денис Александрович – научный сотрудник научно-образовательной лаборатории «Бионические цифровые платформы», СибГМУ, г. Томск, vrazhnov.da@ssmu.ru, <https://orcid.org/0000-0002-6915-6156>

Губанов Александр Валерьевич – ассистент научно-образовательной лаборатории «Бионические цифровые платформы», СибГМУ, г. Томск, derzhiabuz@yandex.ru, <https://orcid.org/0000-0001-7465-6238>

Нестерович Софья Владимировна – канд. мед. наук, гл. врач, клиники СибГМУ, г. Томск, nesterovich.sv@ssmu.ru, <http://orcid.org/0000-0003-2098-2964>

Шмырина Александра Андреевна – начальник управления цифровых технологий, СибГМУ, г. Томск, shmyrina.aa@ssmu.ru, <http://orcid.org/0009-0002-6549-1608>

✉ **Федорова Ольга Сергеевна**, fedorova.os@ssmu.ru

Поступила в редакцию 10.05.2023;
принята к публикации 25.05.2023