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

Russian-language open clinical data repository “SibMED Clinical Data 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

Global digitalization has become one of the most significant challenges in the field of medicine and healthcare. Rapid development of digital technologies determines a growing demand for constant access to real-time big data. Their use is in need for research and technological projects in the field of artificial intelligence technologies. Siberian State Medical University developed the first Russian-language clinical data repository “SibMed Clinical Data Repository” in Russia (https://dataset.ssmu.ru/). The article describes the structure and functions of the repository as well as features of its potential 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 open clinical data repository “SibMED Clinical Data Repository”. *Bulletin of Siberian Medicine.* 2023;22(2):176–181. https://doi.org/10.20538/1682-0363-2023-2-176-181.

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

✉ *Fedorova Olga S.,* dmarkych@gmail.com

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):176–181. https://doi.org/10.20538/1682-0363-2023-2-176-181.

Digitalization has become one of the most significant challenges in the field of medicine and healthcare in the XXI century. The introduction of digital solutions in modern clinical practice makes it possible to improve the quality and effectiveness of medical care, reduce healthcare costs, and achieve patient safety. The rapid development of telemedicine and artificial intelligence technologies determine the growing demand for constant access to real-time big data [1, 2].

The Institute for Statistical Studies and Economics of Knowledge of Higher School of Economics, using the iFORA big data analysis system, has identified digital technologies that are the most in demand in medicine and healthcare. Thus, among the leading technologies are open health data / electronic medical records [3]. Big data accumulated in medical databases are actively used for the development of IT technologies [4-6].

One of the large-scale projects that develops such algorithms is the MIMIC-IV open database of clinical data. The database was developed in collaboration with Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology [7]. Such data are in great demand, however, due to the divergence of health systems in different countries, the use of MIMIC is limited [8].

The development of open health data sources requires formalized algorithms for aggregation and storage of datasets, ensuring information security [9, 10].

Siberian State Medical University developed the first Russian-language clinical data repository "SibMed Clinical Data Repository" (https://dataset.ssmu.ru /). The SibMed Clinical Data Repository project includes the formation of a digital infrastructure for storing and quick access to health data and a training program focused on the work with structured and unstructured biomedical data to solve research and technological problems.

The repository combines datasets of outpatient and inpatient services of multidisciplinary SibMed clinics, contains anonymized health information, and is continuously updated. The data are depersonalized in accordance with the legal requirements.

SibMed Clinical Data Repository is recommended for use both for researchers, specialists in the field of data science, machine learning, and healthcare management, entrepreneurs, and for organizations developing digital solutions in healthcare. The tasks that SibMed Clinical Data Repository solves include: development and testing of new digital products for healthcare and business, data analytics in diagnosis and treatment, implementation of artificial intelligence technologies, development of treatment and diagnostic programs.

The openness of health data within this project will contribute to the creation of new solutions for medicine and healthcare, improve the quality of life of the population, and provide more affordable medical care.

REFERENCES

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. Mokhnacheva T., Monogarova Yu., Varakina Zh. Involvement of healthcare organizers in the process of digitalization of healthcare. *Social Aspects of Public Health.* 2023;(1):2–20 (in Russ.). DOI: 10.21045/2071-5021-2023-69-1-11.

3. Grebenyuk A.Yu. Top 10 digital solutions in medicine and healthcare (in Russ.). URL: https://issek.hse.ru/news/691544400.html
4. Karpov O.E., Subbotin S.A., Shishkanov D.V. Using medical data to create medical decision support systems. *Physician and Information Technology.* 2019;(2) (in Russ.).
5. Avetisyan M.S., Egorov K.S., Kokh V.N. et al. Development of an algorithm for searching for clinically homogeneous patients using semi-structured text data of an electronic medical record of cancer patients. *Physician and Information Technology.* 2019;(3) (in Russ.).
6. Gusev A.V., Zingerman B.V., Tyufilin D.S., Zinchenko V.V. Electronic medical records as a source of data from real clinical practice. *MyRWD*. 2022;(2):8–20 (in Russ.).
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. Burov V., Begtin I., Ganeeva E. Working with open data: features of publication and use in the Russian legal field 18.11.2020 [Analytical report]. *Information Culture* (in Russ.). URL: https://www.infoculture.ru/wp-content/uploads/2020/11/OpenDataReview.pdf

Authors' information

Kulikov Evgeniy S. – Dr. Sci. (Med.), Associate Professor, Rector, Siberian State Medical University, Tomsk, kulikov.es@ssmu.ru, https://orcid.org/0000-0002-0088-9204.

Fedorova Olga S. – Dr. Sci. (Med.), Associate Professor, Head of the Intermediate-Level Pediatrics Division, Siberian State Medical University, Tomsk, fedorova.os@ssmu.ru, http://orcid.org/0000-0002-7130-9609.

Tolmachev Ivan V. – Cand. Sci. (Med.), Leading Researcher, Research and Education Laboratory "Bionic Digital Platforms", Siberian State Medical University, Tomsk, ivantolm@mail.ru. http://orcid.org/0000-0002-2888-5539.

Ryazantseva Ulyana V. – Analyst, Scientific Department, Siberian State Medical University, Tomsk, ryazantseva.uv@ssmu.ru, https://orcid.org/0000-0001-9292-3969.

Vrazhnov Denis A. – Researcher, Research and Education Laboratory "Bionic Digital Platforms", Siberian State Medical University, Tomsk, vrazhnov.da@ssmu.ru, https://orcid.org/0000-0002-6915-6156.

Gubanov Alexander V. – Assistant, Research and Education Laboratory "Bionic Digital Platforms", Siberian State Medical University, Tomsk, derzhiabuz@yandex.ru, https://orcid.org/0000-0001-7465-6238.

Nesterovich Sofia V. – Cand. Sci. (Med.), Chief Physician, Clinics of Siberian State Medical University, Tomsk, nesterovich.sv@ssmu.ru, http://orcid.org/0000-0003-2098-2964.

Shmyrina Alexandra A. – Head of the Department of Digital Technologies, Siberian State Medical University, Tomsk, shmyrina.aa@ssmu.ru, http://orcid.org/0009-0002-6549-1608.

(✉) **Fedorova Olga S.,** fedorova.os@ssmu.ru

Received 10.05.2023;
accepted 25.05.2023