


<https://doi.org/10.21272/hem.2022.2-04>

JEL Classification: I18, I23

Igor Rekunenko,


Dr.Sc., Professor, Sumy State University, Ukraine

 ORCID ID, 0000-0002-1558-629X

email: i.rekunenکو@biem.sumdu.edu.ua

Anton Boiko,


Dr.Sc., Professor, Sumy State University, Ukraine

 ORCID ID, 0000-0002-1784-9364

email: a.boiko@uabs.sumdu.edu.ua

Olha Kramarenko,


Sumy State University, Ukraine

 ORCID ID, 0000-0002-2616-0147

email: o.kramarenko@ssu.edu.ua

Bhola Khan,

Dr.Sc., Professor, Yobe State University, Nigeria

 ORCID ID, 0000-0002-4479-2762

email: bholakhan1512@gmail.com

Correspondence author: o.kramarenko@ssu.edu.ua

DATA MANAGEMENT IN HEALTHCARE RESEARCH AS A GUARANTEE OF ITS QUALITY

Abstract. Today, in times of outbreaks of epidemics such as the Zika virus and COVID-19, health systems around the globe face an urgent need to respond quickly to overcome their spread and prevent the re-infection of humanity. Among the solutions was the World Health Organization's call for immediate, rapid, and, most importantly, open dissemination of medical research data. Although data sharing benefits science and society, there is a range of ethical, legal, cultural, financial, and technical barriers to the dissemination and reuse of medical research data today. Therefore, the quality of data, namely the possibility of their easy search, availability, compatibility, and reusability, is considered relevant in developing data infrastructure in healthcare. Consequently, there is an urgent need to develop an appropriate research data management system in the healthcare system. The aim is to study the efficiency of data management in the healthcare system. This goal is proposed to be achieved in the following sequence: to consider the essence of the concept of research data management; to conduct a bibliometric study of the concept of data management in the medical fields of knowledge, to analyze the frequency of publications on the issue of data management of medical research, indexed by the Scopus database. The object of research is the healthcare sector. The subject is the determinants of the impact of proper management of medical research data on the healthcare sector. The terminological maps of term relationships were constructed using the VOSviewer visualization tool. The study found that 8% of all data management research was conducted in the medical fields of knowledge. The study found an upward trend in the number of health data management publications with the highest growth rate in 2019 and 2020. Analysis of terminology clusters revealed that the keywords «Big Data», «Machine Learning», «Data Collection», «Data Quality», «Data Sharing», «Data Reuse», «COVID-19», and «Blockchain» have the highest number of mentions in publications and strong connection with other publications. Thus, we justified the importance of developing a coherent program and strategic plans for managing research data in the health care system. The results of the study can be used to support decision-making on future opportunities to effectively influence the development of access to medical research data, as well as to ensure the improvement of the quality and confidentiality of research data in the health system.

Keywords: research data management, open data, data management, data quality, medical data, bibliometric analysis.

Cite as: Recunenko, I., Boiko, A., Kramarenko, O., & Khan, B. (2022). Data Management in Healthcare Research as a Guarantee of its Quality. *Health Economics and Management Review*, 2, 36-43. <http://doi.org/10.21272/hem.2022.2-04>



Introduction. The digitalization of all spheres of human life and the development of technological innovations contribute to the acceleration of scientific discoveries and research, which, according to some estimates, double every ten years.

At the same time, the increase in the volume and diversity of scientific information, as well as the improvement of technological means for carrying out research, directly proportionally contribute to the quantitative increase in research data, their diversity, and complexity. In recent years, it has become increasingly clear that research data is of significant value. The focus of various stakeholders, namely representatives of the scientific community, the real sector of the economy, financial agencies, the public sector, and scientific publishers, is on the importance of transparency, reuse, and verification of research data.

It should be noted that the importance of effective data management, ensuring their availability and relevance, is recognized in scientific research in various fields of knowledge. Still, the quality of data collection, analysis, and processing plays a unique role in health research. This is because data-driven medical research is crucial for discovering new methods of treating, diagnosing, preventing, and reducing the risk of repeated outbreaks of human diseases, contributing to new innovative solutions. However, the real challenge for humanity and health data management systems was the Zika virus pandemics in 2016 and COVID-19, which started spreading at the beginning of 2020. Strengthening scientific cooperation and rapid response to ensure equal access to scientific information, empirical data, and real facts have become an urgent need to prevent the spread and subsequently overcome the coronavirus disease COVID-19 (Wellcome, 2020). On January 11, 2020, a team from the Shanghai Public Health Clinical Center shared the first genome sequence of the SARS-CoV-2 virus, a breakthrough in a sharp increase in the volume of medical research data (Zastrow, 2020). Every day, the Nextstrain Clinical Center began receiving between 50 and 200 sequences from research centers worldwide (Zastrow, 2020). Analyzing integrated and augmented data from multiple data sources such as laboratories, hospitals, clinical research centers, and even public government data can generate meaningful information for predictive modeling and further decision-making at the public health administration level. However, researchers and specialists working with data and analyzing it very often face the problem of instant reuse of medical data due to their lack of training, quality, both physical and legal inability to access it, etc. Thus, the main issue of data-driven science is improving the discovery of knowledge that can be achieved through concerted efforts to manage medical research data at all decision-making stages properly. A separate area of improvement in the current situation in the field of data management should be the adoption of agreed standards for working with data throughout its entire life cycle. This will make data globally accessible, interoperable, and reusable for humans and machines.

The article aims to study the efficiency of data management in the healthcare system. It is proposed to achieve this goal in the following sequence:

1. To consider the essence of the concept of research data management.
2. To study the frequency of publications dealing with the issue of managing medical research data that were indexed by the Scopus database.

The object of research is the healthcare sector. The subject is the determinants of the impact of proper management of medical research data on the healthcare sector.

Literature Review. Research on the importance of managing medical research data to ensure qualitative characteristics of data is reflected in many scientific developments of foreign scientists. In their paper, Queralt-Rosinach, N. et al. highlight the importance of developing and adopting the FAIR Research Data Management Plan for hospitals and present their own strategy for managing research data in hospitals with COVID-19 patients (Queralt-Rosinach et al., 2022). Houston et al. conducted a survey on data quality and data management among clinical researchers. Based on the analysis of this survey, scientists have concluded that it was necessary to standardize clinical trials by introducing education and

training for researchers in this area (Houston et al., 2020). In turn, scientists described the FAIR guidelines for research data management in their research (Wilkinson et al., 2016). To be valuable, according to FAIR principles, data must be: findable, accessible, interoperable and reusable (Wilkinson et al. 2016). Thus, the adoption of the FAIR principles for data encourages the development and implementation of systems for the collection and long-term storage of medical research data. This is especially true for extensive medical research data. The advantage is that data warehouses make data easily accessible and reusable, while ensuring that it is retained and available long after the project is completed. Despite numerous benefits of distributing medical research data through data warehouses, the study (Stuart et al., 2018) demonstrates certain gaps in researchers' knowledge of data sharing and working environments, including important issues such as copyright and licensing, privacy, etc. In «Openness in Big Data and Data Repositories», scientists (Xafis and Labude, 2019) provide an overview of the landscape of data exchange through research data repositories, raise ethical issues arising from the dissemination of medical research data such as privacy, restricted access, ownership of the data, issues of who can access medical research data and what levels of restrictions should be applied in which cases. The ethics of big data is currently receiving much attention from researchers. Thus, scientists (Xafis et al., 2019) set out moral principles for big data in the field of healthcare and research, describe a number of procedural values that can be applied in decision-making, as well as a step-by-step process of identifying, considering and addressing ethical issues that may arise in the use of extensive health data. Legal frameworks, protecting the confidentiality of medical information, and the identity of participants in medical research are almost the most important step in the proper management of research data in the healthcare system. However, the sharing of genomic data raises particularly difficult privacy concerns (Heeney et al. 2011). Some considerations for assessing privacy risks in the exchange and dissemination of genomic data, as well as recommendations for protecting their privacy, are investigated by the scientists (Malin and Sweeney, 2004).

A methodological framework for data management and processing of clinical trial results was proposed in a scientific paper (Dobrova and Zupanets, 2011). The study of factors that negatively affect the quality of clinical trial data, their classification, and ways of risk management in ensuring the quality of clinical trial data is considered in (Dobrova et al., 2014).

Methodology and research methods. A bibliometric content analysis of these publications was carried out to identify the connection between data management and the healthcare sector, form a terminological basis, and study the dominant trends in the development of the data management category. To identify the relevant literature, we evaluated the content of the issues under consideration in the Scopus and Web of Science databases because the reliability of a scientific work depends on the database since it should cover the area under study well. Thus, we have chosen the Scopus database as one of the most authoritative bibliographic and abstract databases, to which more than 3 million new documents are added annually.

To ensure the integrity of the study, the content analysis of publications was carried out by us in the following sequence: analysis of the sectoral structure of the problem under investigation; study of the total number of publications on data management in the medical fields of knowledge and analysis of the frequency of their appearance by years; study of terminological relationships.

The first stage. To explore the relationship between publications related to data management and medical fields of knowledge, we generated a sample of 43,161 documents for the following key query (TITLE-ABS-KEY(«Data management») AND (LIMIT-TO (LANGUAGE, «English»))) . The phrase «Data Management» was searched in the titles of articles, keywords, and abstracts. At this stage, we did not apply any restrictions regarding the publication period. We applied the «publication language» filter to sort only publications in English.

The second stage. To obtain a relevant search result and ensure the integrity of the study of the dynamics of the publication activity on data management in the medical fields of knowledge, we generated the following search query TITLE-ABS-KEY («Data management»). Thus, the phrase «Data Management» was searched in article titles, keywords, and annotations. To achieve the most relevant search result, the main sample of subjects was limited according to the following criteria: publication period from 2010 to 2021 inclusive; publication language «English»; field of knowledge «Medicine».

The third stage. Content analysis and construction of terminology maps were carried out using the data visualization software VOSviewer.

Results. The very concept of management defines how to manage, monitor, and measure various aspects of an organization. Today, there are various management programs, including the management of information technology, human capital, material resources, etc. With the growth of data, and taking into account its importance in decision-making, there is a need to develop guidelines and principles for data management.

To study and improve good research data management practices, there is a need to define the very concept of research data management as such. Thus, in a management system, data is viewed as a corporate asset that needs to be managed to maximize its value. Thus, the data management process of corporate organizations is primarily designed to ensure the proper use of this asset (Marco, n.d.). According to the Data Governance Institute glossary, data management is the exercise of decision-making and authority on data-related issues (th Data Governance Institute, n.d.). IBM describes the data management process as a set of tools, methods, and architectures for fast, secure, and cost-effective data collection, access, maintenance, and value enhancement (IBM, n.d.). It is worth noting that both definitions of «management» and «data management» primarily emphasize the data control process in the corporate environment. However, the basic principles of the corporate data management process and its ultimate goal are quite consistent with the principles of working with data in the field of science. Therefore it is quite applicable to the construction of the terminological concept of research data management system.

Thus the data management process itself, since its creation in the 1980s, is only rapidly developing every year and today covers many disciplines. The development of specific processes taking place in the world has always found its response in the research of scientists and their scientific publications. That is why we have chosen the method of bibliometric analysis of publications as an effective tool for studying the relationship between the concept of «Data Management» and the healthcare sector, determining trends in the development of the field under study, and analyzing key terminological areas within the framework of the problem under study.

In the context of our research, analyzing the distribution of publications dealing with data management by branches of knowledge is advisable. To ensure the completeness of the study at the stage of distributing all publications by fields of knowledge, we generated a sample of 43,161 documents for the key query "Data Management" by focusing the search on the titles, abstracts, and keywords of all publications indexed by the Scopus database in English without restrictions on industry affiliation.

Thus, it is quite expected that the largest number of scientific developments (20,458 publications) is concentrated in the area of «Computer Science», which is 27% of the total number of publications, and «Engineering», which is 18% (13,410 publications). It should be noted that the top three branches of knowledge with an indicator of 8% of all studies, which is equivalent to 5,683 publications, are in the medical plane (Figure 1).

Thus, a bibliometric study of publications by branches of knowledge demonstrates that this problem is represented in almost all disciplines and branches of knowledge. However, the focus of a significant proportion of research on medical, chemical, and biochemical fields of knowledge is important. It demonstrates the interconnectedness and importance of the medical data management process in the healthcare sector.

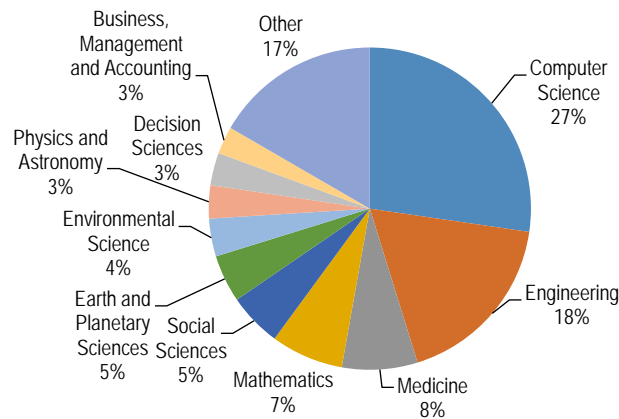


Figure 1. Structure of publications by branches of knowledge of the research field «Data Management»

Sources: developed by the authors based on Scopus data.

An in-depth analysis of trends in the development of scientific research on the topic under study is proposed to begin with an analysis of the total number of studies conducted by scientists in the field of «Data Management» in medical fields of knowledge for the period from 2010 to 2021 (Figure 2).

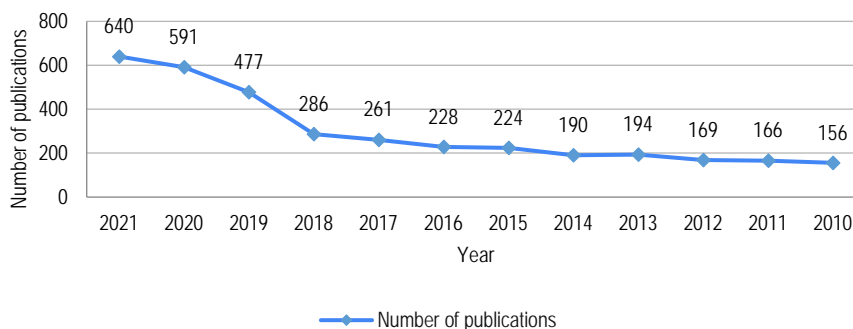


Figure 2. Dynamics of publications for the search query «Data Management» in medical fields of knowledge

Sources: developed by the authors based on Scopus data.

Thus, according to the Scopus database, there is a steady trend towards an increase in the number of publications on the issues under study. The relevance of research in this area in the healthcare sector is confirmed by their increase annually by an average of 15% (Figure 2).

Continuing to study the data in Figure 2, it is worth noting a significant increase in the number of publications in 2019 and 2020, as evidenced by their 45% annual growth. Thus, the number of publications in 2019 increased by 67%, 191 units more than in 2018. In 2020, there is a 24% annual increase compared to 2019. Thus, there is a certain correlation between the focus of researchers on the importance of medical data management and a number of events that have emerged in healthcare, namely the spread of Zika

- Houston, L., Yu, P., Martin, A., & Probst, Y. (2020). Heterogeneity in clinical research data quality monitoring: a national survey. *Journal of Biomedical Informatics*, 108, 103491. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- IBM (n.d.). Data Management Software & Solutions. Retrieved from [\[Link\]](#)
- Malin, B., & Sweeney, L. (2004). How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems. *Journal of biomedical informatics*, 37(3), 179-192. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Marco, D. P. (n.d.). Foundations of Enterprise Data Management. Retrieved from [\[Link\]](#)
- Queralt-Rosinach, N., Kaliyaperumal, R., Bernabé, C. H., Long, Q., Joosten, S. A., van der Wijk, H. J., ... & Roos, M. (2022). Applying the FAIR principles to data in a hospital: challenges and opportunities in a pandemic. *Journal of Biomedical Semantics*, 13(1), 1-19. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Stuart, D., Baynes, G., Hrynaszkiewicz, I., Allin, K., Penny, D., Lucraft, M., & Astell, M. (2018). Whitepaper: Practical challenges for researchers in data sharing. *LONDON: Figshare*. [\[CrossRef\]](#)
- The Data Governance Institute. (n.d.). Definitions of Data Governance. Retrieved from [\[Link\]](#)
- Wellcome. (2020). Sharing Research Data and Findings Relevant to the Novel Coronavirus (COVID-19) Outbreak. Retrieved from [\[Link\]](#).
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Xafis, V., & Labude, M. K. (2019). Openness in big data and data repositories. *Asian Bioethics Review*, 11(3), 255-273. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Xafis, V., Schaefer, G. O., Labude, M. K., Brassington, I., Ballantyne, A., Lim, H. Y., ... & Tai, E. S. (2019). An ethics framework for big data in health and research. *Asian Bioethics Review*, 11(3), 227-254. [\[Google Scholar\]](#) [\[CrossRef\]](#)
- Zastrow, M. (2020). Open science takes on the coronavirus pandemic. *Nature*, 581(7806), 109-110. [\[Google Scholar\]](#)

Ігор Рекуненко, д.е.н., професор, Сумський державний університет, Україна

Антон Бойко, д.е.н., професор, Сумський державний університет, Україна

Ольга Крамаренко, Сумський державний університет, Україна

Бхола Хан, д.е.н., професор, Державний університет Йобе, Нігерія

Менеджмент даних у дослідженнях сфери охорони здоров'я як запорука їх якості

Сьогодні, у часи спалаху епідемій, таких як вірус Зіка та COVID-19, перед системи охорони здоров'я усього світу постала нагальна потреба у швидкому реагуванні на подолання їх поширення та недопущення повторного зараження людства. Одним із способів вирішення став заклик Всесвітньої організації охорони здоров'я до негайного, швидкого, а що найголовніше, відкритого поширення даних медичних досліджень. Саме тому якість даних, а саме можливість їх легкого пошуку, доступність до даних, сумісність та можливість повторного використання, вважається нагальною у розбудові інфраструктури даних галузі охорони здоров'я. Відтак, виникає нагальна потреба у розбудові належної системи управління даними досліджень в системі охорони здоров'я. Метою є дослідження рівня ефективності управління даними в системі охорони здоров'я. Досягнення поставленої мети запропоновано здійснити в наступній послідовності: розглянути сутність поняття управління даними досліджень; провести бібліометричне дослідження поняття управління даними в медичних галузях знань, здійснити аналіз частоти публікацій присвячених питанню управління даними медичних досліджень, що були проіндексовані базою даних Scopus. Об'єктом дослідження є сфера охорони здоров'я. Предметом дослідження є детермінанти впливу належного управління даними медичних досліджень на сферу охорони здоров'я. Побудова термінологічних карт взаємозв'язків термінів була здійснена із застосуванням інструменту візуалізації VOSviewer. У ході дослідження визначено, що 8% усіх досліджень з питань управління даними, були здійснені в медичних галузях знань. Дослідження виявило тенденцію до зростання кількості публікацій з управління даними в сфері охорони здоров'я із найбільшим показником їх зростання у 2019 та 2020 роках. Аналіз термінологічних кластерів виявив, що ключові слова «Big Data», «Machine Learning», «Data Collection», «Data Quality», «Data Sharing», «Data Reuse», «COVID-19», «Blockchain», - мають найбільшу кількість згадувань в публікаціях та їх силу зв'язку з іншими публікаціями.

Ключові слова: управління даними досліджень, відкриті дані, управління даними, якість даних, медичні дані, бібліометричний аналіз.