Scientific article
UDC 330.47
DOI: https://doi.org/10.57809/2024.3.3.10.7

APPLICATION OF DATA MANAGEMENT TOOLS TO IMPROVE EFFICIENCY OF A BIOTECHNOLOGY COMPANY

Valeria Krokhina ✉

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ krohina.vv@edu.spbstu.ru

Abstract. Most companies in today's economy tend to use information technologies both to execute their activities and to support the core business processes of the organization. There are many tools available in the IT market that can meet the needs of organizations and their customers. However, an important requirement for realizing the possibilities of optimal digitalization of production is the high quality of the data used in the organization, as well as the proper data management, which is one of the assets of the organization. A data management approach is important for any company that wants to be competitive in its industry. This paper analyzes the current architecture of data management in a biotechnology company and its disadvantages and proposes a new architecture, taking into account the implementation of integration tools between services to improve data quality and more efficient data management.

Keywords: enterprise service bus, system integration, data management, data architecture, ETL

Citation: Krokhina V. Application of data management tools to improve efficiency of a biotechnology company. Technoeconomics. 2024. 3. 3 (10). 72–82. DOI: https://doi.org/10.57809/2024.3.3.10.7

Научная статья
УДК 330.47
DOI: https://doi.org/10.57809/2024.3.3.10.7

ПРИМЕНЕНИЕ ИНСТРУМЕНТОВ УПРАВЛЕНИЯ ДАННЫМИ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ БИОТЕХНОЛОГИЧЕСКОЙ КОМПАНИИ

Валерия Крохина ✉

Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

✉ krohina.vv@edu.spbstu.ru

Аннотация. Большинство компаний в современной экономике в ходе ведения своей деятельности прибегают к использованию информационных технологий как для реализации своей деятельности, так и для поддержки основных бизнес-процессов организации. На рынке информационных технологий существует масса инструментов, которые могут отвечать на ту или иную потребность организаций и их клиентов. Однако важным условием для реализации возможностей оптимальной цифровизации производства является высокое качество используемых данных в организации, а также грамотное управление данными, которые являются одним из активов организации. Подход к работе с данными важен для любой компании, которая хочет являться конкурентоспособной в своей отрасли. В данной работе путем анализа рассмотрена текущая архитектура работы с данными в биотехнологической компании и ее недостатки, а также предложена новая архитектура с учетом внедрения инструментов интеграции между сервисами для повышения качества данных и работы с ними.

Ключевые слова: сервисная шина, интеграция систем, управление данными, архитектура данных, ETL

Для цитирования: Крохина В. Применение инструментов управления данными для повышения эффективности биотехнологической компании // Техноэкономика. 2023. Т. 3, № 3 (10). С. 72–82. DOI: https://doi.org/10.57809/2024.3.3.10.7

Introduction

Most corporations in various business sectors are actively employing information technology for the efficient operation of their employees and production. The more efficiently an organization works internally and delivers a high-quality product, the higher its position in the marketplace, compared to direct competitors and companies producing substitute goods. A high level of competitiveness allows companies to boost their profits, which is the main goal of any business.

The introduction of digital technology into manufacturing began back in the 20th century and has continued to develop rapidly since then. Thus, almost every business or technological process is accompanied by the use of specific information technology tools, from the use of various devices and sensors to planning and production systems (by levels of computer automation) (Giyosidinov et. al., 2023). Accordingly, the use of information services and tools leads to the generation of constantly growing data.

As previously mentioned, the primary goal of using digital solutions in production is the ability to enhance efficiency and quality of all processes associated with the company. To be more precise, these processes involve:

Process automation: simplification and acceleration of production processes via robotic mechanisms and automated systems.

Data management: collecting, processing, and analyzing quality data to optimize and improve the efficiency of production.

Internet of Things (IoT): connecting equipment and other machines to a single network for real-time monitoring and control from anywhere with the appropriate access.

Quality and production control: implementing quality control systems to reduce errors in production using sensors, data analysis, and artificial intelligence technologies such as machine learning.

Production planning: optimizing logistics and inventory management, adapting quickly to market needs, and planning resource use efficiently using digital platforms.

Cost reduction: optimize resources and reduce operating costs by improving efficiency.

Modelling and simulation: using software to model processes and test changes without the need for actual implementation.

These options have huge potential and are already used in many modern companies. The market for Industry 4.0 technologies is actively developing, and the growth forecast is almost 20% in the next 10 years (IMARC Group).

An important condition for realizing the possibilities of optimal digitalization is the high quality of data used in an organization, as well as the competent management of data, which is one of the organization's assets. Data management is the process related to collecting, accumulating, organizing, remembering, updating, and storing data and using it to improve a company's performance and increase its profits. Since organizational data quality is one of the fundamental criteria for an organization's high level of data maturity, organizations must ensure that the required level of data quality is supported through activities to automate data quality rules, setting up data quality checks, and so on.

Low-quality data can lead to significant financial losses. Gartner estimates that each year poor data quality costs organizations an average of $12.9 million (Gartner). In 2022, Unity Software reported a loss of $110 million in revenue and $4.2 billion in market capitalization due to poor data quality from a major business partner. Similarly, bad data led Equifax, a publicly traded credit agency, to send lenders inaccurate credit ratings for millions of customers (Equifax). Poor data quality is also found in the medical industry. For example, due to data quality issues during the manufacturing process, Zoll Medical's defibrillators were found to display error messages and even fail during use (Get Right Data). The company had to conduct a Class 1 recall, the most serious type of recall for situations in which there is a reasonable likelihood that the use of these products will result in serious injury or death to a person. The recall resulted in a loss of trust and $5.4 million in fines.

The main purpose of this article is to present an improved data architecture in a biotechnology company to make the data management more efficient and to minimize the occurring data errors. The authors carried out the following tasks in order to achieve this goal:

– analyzed the data processing tools available on the market;
– assessed the ways to improve the data management approach;
– examined the current data architecture of a biotechnology company and suggested an improved structure using integration tools.

Materials and Methods

This research invites the following methods: collection and analysis of information, comparison, description, and modeling. Analytics involves gathering information on data technologies, assessing information on the company and its data architecture, and selecting the best data architecture option that will address the identified specifics of data management in the systems.

Results and Discussion

According to the DIKW pyramid (Gordon, 2024), data makes the basis for the information and knowledge that companies use in their operations. Various data management tools are implemented to ensure higher performance:

Data Integration. Data integration technologies allow data from different sources (databases, applications, IoT devices) to be combined to create a single information model. Using ETL (Extract, Transform, Load) or ELT (Extract, Load, Transform) processes and tools for data integration ensures integrity and relevance of information.

Machine Learning (ML) and Artificial Intelligence (AI). The application of machine learning (ML) and artificial intelligence (AI) algorithms can automate data analysis processes, identify hidden patterns, and build predictions. Machine learning technologies can be used to create personalized recommendations, optimize production processes, and predict demand or any business scenarios.

Blockchain. Blockchain provides secure and transparent storage of data by distributing information across a chain of blocks. Blockchain can be used to ensure data integrity, validate transactions, and manage access to sensitive information.

Data quality management. Using tools to clean data, check for duplicates, and ensure data integrity. Compliance with regulations and standards to maintain high data quality.

Data Analytics Tools. Data analytics platforms (e.g., Tableau, Power BI, Looker) allow visualizing data and creating interactive reports to extract information from the collection of data to make strategic decisions in the organization (Ivanov et. al., 2023). What is more, programming languages (Python and R) are used for statistical analysis and machine learning for the same purpose and even for production, automation, and monitoring.

Not every company is mature enough to utilize all these technologies at once. To use analytics tools, an organization is supposed to possess high-quality data and develop analytics tools gradually to take full advantage of the technological capabilities. The maturity levels of data analytics in an organization are described by Gartner's analytics evolution model (Rowley, 2007). Each of the 4 stages of data analytics evolution in a company answers a specific question: descriptive analytics (What happened?), diagnostic analytics (Why did it happen?), predictive analytics (What will happen?), and (How can we implement it?). It turns out to be impossible to forecast what will happen in the future if the awareness over what is happening now is scarce. Organizations have to meet certain requirements in order to evolve and increase the application of analytics in their operations. Some of them are presented below:

Access to quality data. The availability of reliable and relevant data makes the foundation for the development of data analytics.

Analytical culture. The organization should encourage the use of data in decision-making and support the development of an analytical culture among employees.

Technology infrastructure. Availability of appropriate tools and technologies to collect, store, process, and analyze data.

Staff competence. Availability of skilled data analytics professionals, including analysts, data scientists, and developers.

Thus, in order to develop high-quality analytics and obtain reliable and valuable information from data, organizations have to take care of the data quality. It makes the one and only prerequisite for efficient strategic decisions and improved competitiveness.

It is important to bear in mind that manufacturing organizations not only own data on the processes of production but also on entities that support this very production: product, employees who work on the production line, the raw material suppliers, contractors, and so on. The more complex the production line, the more data is generated, stored, and used.

With the current development of technology and economy, the biotechnology industry has become one of the fastest-growing industries in the world. Biotechnology combines the principles of biology, chemistry, and engineering to create innovative products and processes that can solve complex problems in medicine and healthcare. The biotechnology industry is in constant advancement in genetic research technologies, DNA sequencing, and the analysis of proteins and other biomolecules. Studies, discoveries, and innovations in this industry allow scientists and researchers to gain more and more data about living systems and their functions.

Processing data in biotechnology manufacturing can be divided into several aspects based on the specifics of this industry:

Collecting data on production processes: real-time data on temperature, pH, pressure, nutrient concentration, and other parameters allow controlling the necessary conditions for microorganisms or cells to be cultured.

Data analysis: bioinformatics is used to process and analyze genomic data to identify genes responsible for specific functions. Statistical methods are used to analyze experimental data and identify patterns, for example, in tests for the efficacy of new drugs.

Quality control: data on each stage of production helps to monitor compliance with quality and safety standards (e.g., GxP and ISO 9000). Another test is deviation analysis, which consists of collecting and analyzing data on non-conformances to identify potential problems and prevent their recurrence.

Process Optimization: using data to create mathematical models that help to optimize production conditions and improve efficiency. Machine learning algorithms help forecasting the results of experiments based on statistical data (Dubgorn, 2020).

Collecting and analyzing data on the safety and efficacy of new drugs at various stages of clinical trials.

Use of Industry 4.0 technologies: sensors to collect real-time process data and improve monitoring and control; cloud technologies to store and analyze large amounts of data and share them across multiple devices.

Data security: the importance of complying with regulations to protect personal data and intellectual property, especially when dealing with genetic information (Ilin, Iliashenko, 2022).

Data in the biotechnology not only contributes to process efficiency but also opens up new opportunities for research and innovative product development. The large amount of data generated, however, presents a significant challenge. The data resides in various sources, such as laboratory instruments, databases, and others. Integration of these data becomes a necessity to obtain a comprehensive perspective on complex biological processes and to develop new technologies (Iliashenko et. al., 2018).

The use of data integration tools also improves the efficiency of any company. Such technology automates the processes of data collection, storage, and analysis, which reduces the time and effort spent on searching and processing information. Moreover, data integration improves the accuracy of results and reduces the possibility of errors, which is a critical factor in sensitive areas such as drug development or genetic research.

The biotechnology company combines research centres, pharmaceutical and biotechnological production, and preclinical and clinical research systems. It covers the full cycle of drug development: from molecule search and genetic engineering to mass production and marketing support. Thus, the activities in the organization are divided into three components: management, core, and supporting activities.

Supporting activities involve information and technological support, like most modern companies. In this regard, the IT direction is well developed in terms of resources and technologies.

The IT area is structured according to the principle of support provided to the core business: process production automation, infrastructure and systems, and information technology architecture. Each branch is supported by corresponding systems that form the microservice architecture of software. This includes interprise solutions such as 1C: Enterprise, ERP and ECM platforms, automated process control systems, equipment monitoring systems for production and warehouses, tools for BI-analytics, and advanced analytics (e.g., a system for forecasting performance in bioreactors based on real-time data from sensors). Only a small fraction of systems are mentioned, but in total more than 50 systems are collected.

The presence of a large number of services leads to the fact that a huge amount of data is generated as a result of the company's activities. Given that the vector of development of modern companies is aimed at digitalization in order to increase their competitiveness in the biotechnology market, data on business processes and their products is also used for analytics. The organization under consideration claims a high level of analytics development, according to Gartner, as the following advanced analytical tools are used:

– Power BI for reports and dashboards (Iliashenko et. al., 2020);
– Computer vision tools for tracking production and process flows;
– Monitoring platform for instant detection and prevention of IT failures and comprehensive IT infrastructure management;
– Assessment of process data and deviations via identifying anomalies and interrelationships of parameter readings;
– A system for planning clinical trials and control of monitoring activity;
– Monitoring sensors on production equipment to prevent breakdowns and accidents.

However, even with advanced data tools, the company faces data errors, inappropriate formats, and a lack of digitalization. Unfortunately, such problems in working with data not only affect the competitiveness of the organization but also lead to additional expenditure of IT specialists' resources (Dubgorn et. al., 2020).

No regulations were prescribed for describing the organizational structure of the company, updating this information, and transferring it to other systems. These data on the structure were transferred daily to the ESM platform with some errors, in which they were already used to build the logic of business processes for the company's documentation. Further, reports in Power BI on KPIs of these processes displayed distorted information about the activities of the departments that directly work with documentation and are responsible for it. Thus, data errors in one system caused the following problems:

– Errors in other systems where the data was passed up the chain;
– Burden on technical support services, as these errors were found by the departments whose KPIs are based on documentation processes;
– Developer's labour costs to refine the integration so that data is transmitted without distortions and errors.

This case in the company's practice indicates what problems arise due to the large number of unregulated data flows and why it is necessary to develop/implement a new approach to data transfer between systems and the formation of a common data warehouse that could be used for analytics (Ilin et. al., 2018; 2019; 2020).

Integrating systems separately between each other is technically difficult and does not provide for a common repository of information from the units being integrated. Based on the problems in handling data from different systems, it is suggested to resort to implementing an ESB tool and a common ETL system with which the existing systems will be integrated.

Let's consider the recommended data architecture for the company, taking into account the implementation of ESB and ETL (Fig. 1).

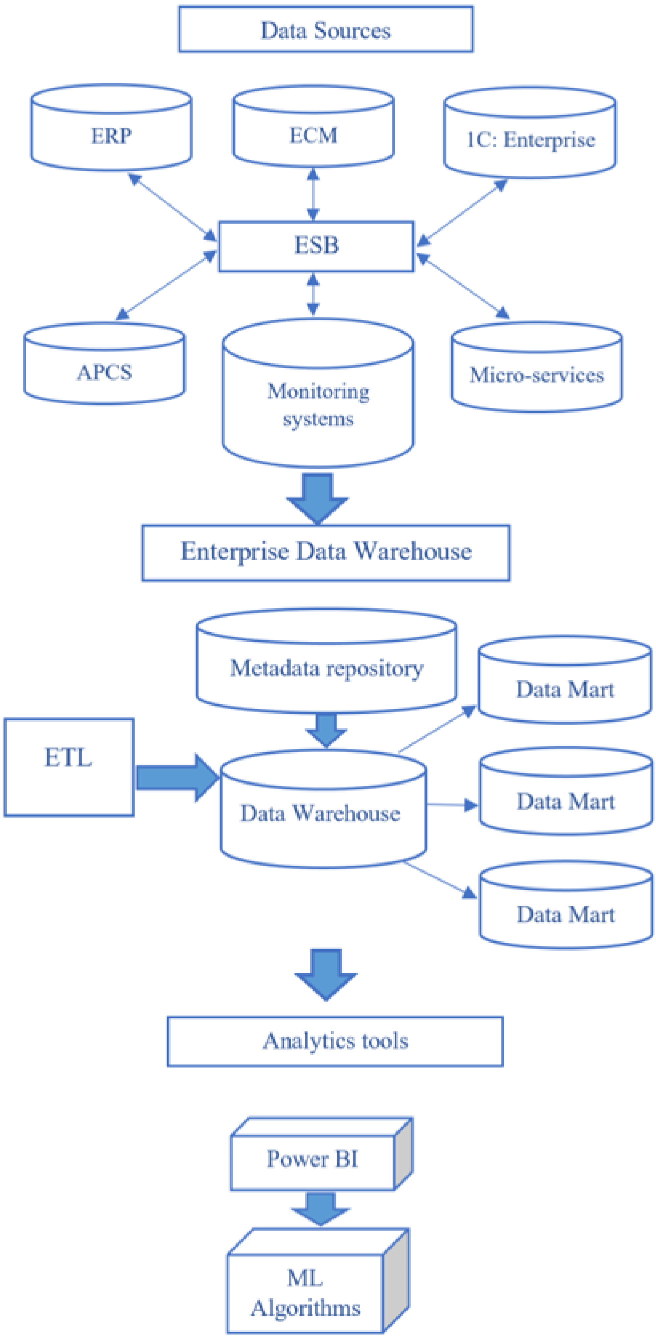

Fig. 1. TO-BE data architecture

As mapped in the TO-BE architecture, it is proposed to integrate the company's services with an ESB tool to exchange data between each other, as well as an ETL system to form a common data warehouse from these services, which in turn will be used for analytics. The following is a closer look at these two technologies.

The ESB (Enterprise Service Bus) is software that provides connectivity and integration between the various applications, systems, and services in the corporate IT landscape. It acts as a central messaging platform that allows applications to communicate with each other without the need for direct integration. The enterprise service bus plays an important role in data integration as it enables applications and systems to communicate with each other in real time (Robin et. al., 2017; Dong et. al., 2011). It provides reliable and efficient messaging, data conversion, and routing.

The advantages of ESB are:

Simplified integration: it allows new applications and systems to be integrated into existing IT infrastructure quickly and easily;

Improved productivity: automates messaging and eliminates the need for manual integration, which increases efficiency;

Flexibility and scalability: provides a flexible and scalable architecture that can adapt to changing business needs;

Improved data quality: provides message conversion and routing services, which improve data quality and consistency between systems;

Improved security: establishes centralized access and security controls, which improves the overall security of IT systems.

ESB functionality implies:

Messaging: provides mechanisms for reliable and efficient messaging between applications and services;

Data conversion: converts data from one format to another, enabling interoperability between systems with different data structures.

Message Routing: routes messages to appropriate recipients based on defined rules;

Process management: coordinates the sequence of tasks and actions required to execute business processes;

Data management: provides data management functions such as data integration, synchronization, and data quality.

ETL (extract, transform, load) is the process of moving data from various sources into a target data system for analysis and reporting. It involves three main steps:

Extraction: extracting data from source systems such as databases, files, or web services.

Conversion: Converting the extracted data into a format compatible with the target system. This may include changing the data structure, data cleansing, and applying business rules.

Loading: Loading the transformed data into the target system, which may be a data warehouse, database, or other repository.

An ETL system is critical to ensuring consistency, accuracy, and completeness of data for analysis. It allows organizations to combine data from different sources, improve data quality, and create a single version of the truth. By implementing a common ETL system, organizations can benefit from centralized data management and simplified integration processes (Mhon, Kham, 2020; Bengeri, Goje, 2022).

The benefits of a common ETL system are:

Centralized data management: provides a single source of consistent data.

Simplified integration processes reduce the time and effort required to integrate new systems.

Improved efficiency: automates data conversion and loading processes, freeing up resources for other tasks.

Data Quality Assurance applies data transformation rules to ensure accuracy and consistency.

Improved analytics: provides quality data for accurate analysis and informed decision-making.

Conclusion

The implementation of ETL in the case study is a solution to the problem of data analysis and other types of analytics available in the company. Since any analytics involves working with a large amount of data obtained from different sources, the proposed solution will make it easier to work with them. The company's employees will be able to compare the received information, analyze it, and make forecasts based on it much easier and more efficiently.

Integration of existing systems with ETL facilitates the formation of MDM—data containing key information about the business and industry, including customers, products, employees, technologies, and materials. Each of these groups can be divided into several subject areas: the people category includes customer, salesperson, and supplier. So we can have a set of validation rules that the data must satisfy.

ETL and ESB are used together to create a comprehensive data management and integration solution. ETL is responsible for extracting, transforming, and loading data into a data warehouse or other target system, while ESB provides real-time data exchange between applications and systems.

Integrating ETL and ESB will allow a company to:

– Automate the exchange of data between different systems;
– Improve data quality and consistency;
– Reduce the time required to deliver data for analysis and reporting;
– Increase the flexibility and scalability of IT systems.

Implementing ETL and ESB requires careful planning and implementation. Organizations must define their integration requirements, select appropriate technologies, and develop an architecture that meets their business needs. All of the recommendations presented for changing a company's data architecture with the addition of ETL and ESB will help to sustain a high level of company analytics development, as keeping data in one environment in a structured manner promotes quality analytics that can continually evolve.

REFERENCES

Bengeri A., Goje A. 2022. Technological and scientific developments towards use of big data in health data management–an overview. International journal of scientific research in engineering and management 06 (02). doi:10.55041/IJSREM11727

Dong G., Xu H., Shi W. 2011. Healthy Record and Hospital Medical Record System Intercommunication Based on ESB. Advances in Multimedia, Software Engineering and Computing 2, 129. doi:10.1007/978-3-642-25986-9_54

Dubgorn A., Ilin I., Levina A., Borremans A. 2020. Reference model of healthcare company functional structure. Proceedings of IBIMA 2019, 5129-5137.

Dubgorn A., Svetunkov S., Borremans A. 2020. Features of the functioning of a geographically distributed medical organization in Russia. E3S Web Conf., 217 06014. doi:10.1051/e3sconf/202021706014

Giyosidinov B., Fedorchuk V., Voronova O. 2023. Digital transformation of trade: trends, stages and factors of digitalization at the sectoral level. Technoeconomics 2.4 (7), 38–45. doi:10.57809/2023.2.4.7.4

Gordon I., Thompson N. 2024. Radical Technologies. Data and the Built Environment. Digital Innovations in Architecture, Engineering and Construction. doi:10.1007/978-3-031-51008-3_6

Iliashenko O., Bikkulova Z., Dubgorn A. 2018. Opportunities and challenges of artificial intelligence in healthcare. SPbWOSCE, 02028. doi:10.1051/e3sconf/201911002028

Iliashenko O., Iliashenko V., Dubgorn A. 2020. IT-architecture development approach in implementing BI-systems in medicine. Lecture Notes in Networks and Systems 95, 692-700. 10.1007/978-3-030-34983-7_68

Iliashenko V., Ilin I., Iliashenko O. 2022. Medical organization data management: challenges and opportunities. Science and business: ways of development 10 (136), 257-262.

Ilin I. Lepekhin A., Levina A. Iliashenko O. 2018. Analysis of Factors, Defining Software Development Approach. Advances in Intelligent Systems and Computing, 1306-1314. doi:10.1007/978-3-319-70987-1_138

Ilin I., Iliashenko O., Iliashenko V. 2019. Architectural approach to the development of medical organization in the conditions of healthcare digitalization. Journal of Management Research 5 (1), 37-47.

Ilin I., Iliashenko O., Iliashenko V. 2019. Architectural approach to the digital transformation of the modern medical organization. IBIMA 2019, 5058-5067.

Ivanov D., Pelipenko E., Dubgorn A. 2023. Improving the efficiency of drug supply management of a medical organization through the use of a business intelligence system. Technoeconomics 2.3 (6), 79–89. doi:10.57809/2023.2.3.6.7

Mhon G., Kham N. 2020. ETL Preprocessing with Multiple Data Sources for Academic Data Analysis. ICCA, 1 (5). doi:10.1109/ICCA49400.2020.9022824

Robin S.B, Narendra S. et. al. 2017. The Performance Metric for Enterprise Service Bus (ESB) in SOA system: Theoretical underpinnings and empirical illustrations for information processing. Information Systems 65, 158-171. doi:10.1016/j.is.2016.12.005

Rowley J. 2007. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science, 33 (2), 163-180. doi:0.1177/0165551506070706

Equifax. (n.d.). Equifax statement on recent coding issue. URL: https://www.equifax.com/newsroom/all-news/-/story/equifax-statement-on-recent-coding-issue/ (accessed 16.08.2024).

Gartner. (n.d.). How to improve your data quality. URL: https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality (accessed 15.08.2024).

Get Right Data. (n.d.). Why quality matters: The 10 biggest data quality disasters. URL: https://www.getrightdata.com/resources/why-quality-matters-the-10-biggest-data-quality-disasters (accessed 16.08.2024).

IMARC Group. (n.d.). United States Industry 4.0 market. URL: https://www.imarcgroup.com/united-states-industry-4-0-market (accessed 15.08.2024).

СПИСОК ИСТОЧНИКОВ

Bengeri A., Goje A. 2022. Technological and scientific developments towards use of big data in health data management—an overview. International journal of scientific research in engineering and management 06 (02). doi:10.55041/IJSREM11727

Dong G., Xu H., Shi W. 2011. Healthy Record and Hospital Medical Record System Intercommunication Based on ESB. Advances in Multimedia, Software Engineering and Computing 2, 129. doi:10.1007/978-3-642-25986-9_54

Dubgorn A., Ilin I., Levina A., Borremans A. 2020. Reference model of healthcare company functional structure. Proceedings of IBIMA 2019, 5129-5137.

Dubgorn A., Svetunkov S., Borremans A. 2020. Features of the functioning of a geographically distributed medical organization in Russia. E3S Web Conf., 217 06014. doi:10.1051/e3sconf/202021706014

Giyosidinov B., Fedorchuk V., Voronova O. 2023. Digital transformation of trade: trends, stages and factors of digitalization at the sectoral level. Technoeconomics 2.4 (7), 38–45. doi:10.57809/2023.2.4.7.4

Gordon I., Thompson N. 2024. Radical Technologies. Data and the Built Environment. Digital Innovations in Architecture, Engineering and Construction. doi:10.1007/978-3-031-51008-3_6

Iliashenko O., Bikkulova Z., Dubgorn A. 2018. Opportunities and challenges of artificial intelligence in healthcare. SPbWOSCE, 02028. doi:10.1051/e3sconf/201911002028

Iliashenko O., Iliashenko V., Dubgorn A. 2020. IT-architecture development approach in implementing BI-systems in medicine. Lecture Notes in Networks and Systems 95, 692-700. 10.1007/978-3-030-34983-7_68

Iliashenko V., Ilin I., Iliashenko O. 2022. Medical organization data management: challenges and opportunities. Science and business: ways of development 10 (136), 257-262.

Ilin I. Lepekhin A., Levina A. Iliashenko O. 2018. Analysis of Factors, Defining Software Development Approach. Advances in Intelligent Systems and Computing, 1306-1314. doi:10.1007/978-3-319-70987-1_138

Ilin I., Iliashenko O., Iliashenko V. 2019. Architectural approach to the development of medical organization in the conditions of healthcare digitalization. Journal of Management Research 5 (1), 37-47.

Ilin I., Iliashenko O., Iliashenko V. 2019. Architectural approach to the digital transformation of the modern medical organization. IBIMA 2019, 5058-5067.

Ivanov D., Pelipenko E., Dubgorn A. 2023. Improving the efficiency of drug supply management of a medical organization through the use of a business intelligence system. Technoeconomics 2.3 (6), 79–89. doi:10.57809/2023.2.3.6.7

Mhon G., Kham N. 2020. ETL Preprocessing with Multiple Data Sources for Academic Data Analysis. ICCA, 1 (5). doi:10.1109/ICCA49400.2020.9022824

Robin S.B, Narendra S. et. al. 2017. The Performance Metric for Enterprise Service Bus (ESB) in SOA system: Theoretical underpinnings and empirical illustrations for information processing. Information Systems 65, 158-171. doi:10.1016/j.is.2016.12.005

Rowley J. 2007. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science, 33 (2), 163-180. doi:0.1177/0165551506070706

Equifax. (n.d.). Equifax statement on recent coding issue. URL: https://www.equifax.com/newsroom/all-news/-/story/equifax-statement-on-recent-coding-issue/ (accessed 16.08.2024).

Gartner. (n.d.). How to improve your data quality. URL: https://www.gartner.com/smarter-withgartner/how-to-improve-your-data-quality (accessed 15.08.2024).

Get Right Data. (n.d.). Why quality matters: The 10 biggest data quality disasters. URL: https://www.getrightdata.com/resources/why-quality-matters-the-10-biggest-data-quality-disasters (accessed 16.08.2024).

IMARC Group. (n.d.). United States Industry 4.0 market. URL: https://www.imarcgroup.com/united-states-industry-4-0-market (accessed 15.08.2024).

INFORMATION ABOUT AUTHOR / ИНФОРМАЦИЯ ОБ АВТОРЕ

KROKHINA Valeria V. – student.
E-mail: krohina.vv@edu.spbstu.ru
КРОХИНА Валерия Витальевна – студент.
E-mail: krohina.vv@edu.spbstu.ru

Статья поступила в редакцию 28.08.2024; одобрена после рецензирования 04.09.2024; принята к публикации 09.09.2024.

The article was submitted 27.05.2024; approved after reviewing 04.09.2024; accepted for publication 09.09.2024.