Russian Journal of Earth Sciences, Vol. 22, ES01SI02
https://doi.org/10.2205/2022ES01SI02

Russian Journal *of* Earth Sciences

Special Issue: "Fundamental and applied aspects of geology, geophysics and geoecology"

Using Big Data Technology to Protect the Environment

Vitaly A. Dovgal*,[1] and Saida K. Kuizheva[1]

[1] *Maykop State Technological University, Maykop, Russia*

Received 1 July 2021; accepted 10 August 2022; published 27 August 2022.

In the context of digitalization of all aspects of the surrounding world, data is becoming particularly relevant as one of the most valuable resources. The concept of "big data" means a huge amount of information, the size of which is too large, or it is created too quickly or has a structuring that does not allow it to be processed using traditional data management systems. Currently, large amounts of data and analytics are increasingly used by government agencies, non-governmental organizations and private companies in the field of environmental protection. The range of practical use of this technology is quite wide: from improving energy efficiency, tracking climate change over long periods of time, monitoring water quality, and ending with the promotion of environmental justice. This article describes several extremely promising applications of large data sets and their analytics, which can help achieve the goals of environmental protection and sustainable development, provide environmental benefits, help research on the environment, its conservation and protection. The widespread adoption of big data processing solutions allows us to illustrate the range of initiatives and approaches to reduce the environmental burden used by government agencies, non-governmental organizations and private companies.

Keywords: big data, protection of the environment, sustainable development.

Citation: Dovgal, Vitaly A. and Saida K. Kuizheva (2022), Using Big Data Technology to Protect the Environment, *Russian Journal of Earth Sciences*, Vol. 22, ES01SI02, doi: 10.2205/2022ES01SI02.

1 Introduction

According to the law of exponential growth of the volume of knowledge, every year the knowledge that humanity has doubles. And for their processing, more and more high-speed computing devices are used. As a result, a process or phenomenon can increasingly be described, emphasizing aspects of data size, with the help of super-large arrays of information that were not prospectively limited in size or volume by the intention to solve specific research questions [*Ghani et al.*, 2014, p. 976]. These rapidly growing data are of increasing interest to both scientists and practitioners. In databases, information is presented in various forms: text content (i.e. it has the form of a structured, semi-structured and unstructured form), multimedia content (for example, video, images, audio) or integrated platforms (for example, machine-to-machine communications, social networking sites, sensor networks and the Internet of Things) [*Sivarajah et al.*, 2017]. In addition, the data content is constantly changing due to the absorption of additional data collections, the introduction of previously archived data or outdated collections, as well as the addition of streaming data coming from several sources [*Berman*, 2013]. Very often, big data is described by contrasting three main characteristics inherent in it: "Volume", "Velocity" and "Variety" [*Chen et al.*, 2012]. The "Volume" attribute characterizes the size of the data, which is continuously increasing. "Diversity" describes the presence of different types of data collected using a variety of technologies or sources, such as sensors, smartphones or social networks. The "Speed" attribute is a characteristic of the data transfer rate. Recently, some newly emerged concepts have added two more V letters to the characteristics of big data – "Value" and "Veracity". The value of the data is manifested in the detection of hidden values that can provide an increase in the profitability of the company or eliminate losses for the client in the value stream that is relevant in lean manufacturing processes [*Karamushko and Dovgal*, 2020]. "Reliability" emphasizes not only the importance of high-quality data, but also the associated level of security of the information used [*Wamba et al.*, 2017].

*Corresponding author: vdovgal@mkgtu.ru

2 Promising advantages of using big data

Since its appearance, big data has been credited with the ability to revolutionize the art of management [*Wamba et al.*, 2017]. From an organizational point of view, big data can manifest itself in all areas of the company's activity to improve the product and the market, increase operational efficiency, to issue more correct forecasts of market demand, allow making the right business decisions, opening up competitive advantages in the field of strategic cost differentiation [*Dovgal*, 2020]. Environmental protection, as a type of human activity, is also no exception, since it is based on the use of the collected information and the analysis of data, including large sizes.

Real business models can be built based on the following approaches: differentiation of information, mediation in the provision of information and information networks [*Wang*, 2012]. The first approach, using information extracted from big data, allows you to create new diverse offers, diversifying the company's business. Based on the behavior of customers, you can offer new services, increase satisfaction and ensure contextual relevance.

The approach associated with the mediation of information allows you to make a profit through the commercialization of big data owned by the company. Thus, business opportunities are manifested both when selling raw (primary) information, and performing an analysis of market practices, and providing the results of the analysis and conclusions on it. The third approach, related to the creation of networks for data delivery, implements the possibility of information aggregation, its subsequent exchange and transformation, stimulating the emergence of new data. From the point of view of environmental protection, all three approaches to the use of large data sets seem promising.

Thus, companies with big data can benefit from the resulting overabundance of information. Therefore, the emergence of powerful intelligent solutions in the field of big data processing and analytics is currently a very urgent task for the study and analysis of the flow of information, because it can lead to a fundamental change in the way organizations manage their daily activities. After all, the benefits obtained from big data and analytical results are very significant: this is a differentiation from its competitors.

3 Examples of the use of big data in the field of conservation and sustainable development

It is obvious that in the field of environmental research, big data can also be useful, allowing you to identify key conditions and trends in the ecological state of specific regions, countries and the planet as a whole. Big data analysis can bring great benefits, being supported by continuous efforts to ensure sustainable development.

The use of big data based on the methods of collecting, processing, analyzing and visualizing data sets that have appeared and been introduced in recent decades has allowed biogeophysicists to detect, analyze and understand environmental changes on a micro and global scale, as well as to separate those that are caused by human influence. Trends arising from environmental analyses of existing big data sets have been formed, and prospects for using these technologies to mitigate the global environmental downturn are being viewed.

There are many examples of the active use of big data to protect the environment. Let's consider those of them that are most significant and popular on the Internet.

3.1 Preserving the biodiversity of plants, animals and landscapes around the world

One of the examples of international organizations operating in this direction can be called Conservation International [*Conservation International*, 2021]. It is engaged in the protection of nature, drawing attention to this type of activity and providing the most important benefits that nature provides to humanity. At the same time, the best achievements of science, innovative policy and global reach are used (the organization has offices in more than two dozen countries, and the worldwide network covers thousands of partners). The main role in the protection of nature, on which humanity depends for food, fresh water and livelihood, is played by the analysis of the situation with the help of big data. Photo traps installed in various tropical forests of the world and working remotely allow the organization's specialists to receive and save about 500 thousand raw images per year in the database. These images are used to build an annual binary matrix for each species at each site where the camera traps are located. Then scientists create an index of images of wildlife (Wildlife Picture Index, WPI), which can be used to simulate changes in the population of animals in a particular environment, in a particular area. The WPI Analytics system has high data accuracy, timely analysis – updating of fresh data occurs in real time due to the use of powerful servers. The results of processing allow scientists to develop methods that help preserve the population: animal feeding, artificial conditions of detention, fertilization, changing tourist trails, etc.

WPI is used by researchers from Conservation International command centers, scientists from the Smithsonian Tropical Research Institute and the Wildlife Conservation Society, heads of national parks. The site with WPI data [TEAM]. is already being updated with data from other groups that were collected in a similar way.

On the territory of the Russian Federation, the World Wildlife Fund (WWF) is working with a similar project, which installs photo traps in specially protected natural areas throughout the country. Currently, the Foundation's employees have placed more than 650 such devices in nature reserves and national parks of Russia [WWF].

3.2 Solving the world's food problems through a successful combination of digital and physical technologies

The use of big data and advanced analytics allows us to solve the problem of hunger in the world. As an example of such solutions, we can cite the proprietary Cargill Data Platform (CDP) [CIO, a]. It allows you to structure satellite images taken in low Earth orbit and very complex images, recognizing areas with dairy and meat farm animals or row crops. Based on artificial intelligence, it is possible to predict the harvest and provide optimal solutions on the farm. Another example: there are already several hundred shrimp farms in the world on the iQuatic platform, which collects information from shrimp ponds and uses artificial intelligence and a visualization application to grow shrimp as an extremely important source of protein. To ensure transparency and traceability of the entire supply chain, companies' customers can use distributed ledger technologies, commonly known as blockchain.

3.3 Combating climate change

Big data has long been used in climatology to assess the damage caused by pollutants and greenhouse gases. To assess the environmental impact of manufactured products, it is necessary to process a large amount of data that is collected by extensive Internet of Things networks [*Dovgal et al.*, 2021], covering almost every sector of the manufacturing and food industry. There are companies emerging in the world that are already pioneers in using big data to find more environmentally friendly markets. Examples include carbon emission quota registries that have recently appeared in the United States or technologies developed by Oxy Low Carbon ventures that allow trading these quotas in the same way as commodities [Insidebigdata].

In addition, big data-based tools are often used to monitor seasonal fluctuations in climate change [*Manogaran and Lopez*, 2018], manage climate change risks [*Ford et al.*, 2016], etc.

3.4 Monitoring of The Antarctic Ice Sheet

Scientists consider the Antarctic ice sheet to be an important indicator of climate change and a driving force behind sea level rise. To observe the changes in its volume, flow and gravitational attraction, data obtained from satellites are used, on the basis of which the balance of its surface mass is simulated. As an example, a joint study by NASA and the European Space Agency, which showed that Antarctica lost 2720 ± 1390 billion tons of ice between 1992 and 2017 (which corresponds to a sea level rise of 7.6 ± 3.9 mm) [*Shepherd et al.*, 2018].

3.5 Data-driven cities (DDC)

Modern cities use the latest technologies to maintain sustainable development and a high standard of living in conditions of rapid population growth, increasing the territory of the municipality and the complexity of its infrastructure. One of the most urgent tasks in such a city is to reduce the negative impact on the environment. Many projects in this category have already been implemented in the world. As an example, a joint pilot project of the National ICT Center of Australia (NICTA) and the Department of Environmental Protection of the State of New South Wales (NSW EPA), which provides for the collection of data on the state of the atmosphere from installed sensors in the Hunter Valley area [PWC], can be cited. The analysis of the received information, carried out using machine learning methods and forecasting algorithms, allows us to calculate the air quality index AQI (air quality index) for the next day. This indicator allows you to predict the level of air pollution for the next 24 hours in various areas of the state. Based on this index, various enterprises in the field of industrial production can take appropriate measures to maintain air quality at a level corresponding to the norm. In addition to monitoring air pollution, it is possible to monitor urban noise, the environmental impact of construction works, the composition and condition of green spaces in the city in a similar way based on data from installed sensors.

3.6 Urban parks surgery

Pruning and caring for trees in urban squares and parks is a matter of public safety. For example, in New York, there are about 2.5 million trees that require care, without which they can harm the residents of the city – falling tree branches can seriously injure people walking under them or lead to fatal cases. Pruning and other types of tree care are considered a way to increase their resistance to a storm or hurricane wind, which will help reduce the likelihood of material damage, injuries and deaths. For better preventive care of urban trees, New York City parks have created a program for regular pruning and care of large trees. Then, using big data technology, specialists of the non-profit organization DataKind compared the pruning schedule for one year and information about the removal of fallen branches and fallen trees [CIO, b] The data sets used had different levels of detail – the pruning was recorded block-by-block, while the cleaning data was recorded at the address level. Having brought the available data for one city block in line with each other, the analysis of big data based on intensive modeling allowed us to conclude that pruning trees to prevent certain types of hazards led to a 22 percent reduction in the number of calls to the emergency cleaning team. The second phase – intelligent pruning – allows you to determine the prioritization of servicing urban neighborhoods. It is necessary to build a risk profile of the block: the number of trees and their types, the location of the block in the flood or storm zone, and other accompanying adverse factors. Now the city administration can effectively use the investments made in the infrastructure of city parks.

3.7 The appearance of updated maps with locate data.

Increasing the objectivity of the collected data is provided by a sufficiently high intellectual level of sensors, which are becoming more and more thin, flexible, economical and environmentally friendly, using nanotechnological solutions. The miniaturization of sensors contributes to their active implementation in mobile devices, and, therefore, they can now be used everywhere in an increasing number of applications for collecting environmental data.

The sensors built into the smartphone, regardless of the location of its owner, allow you to create a network of personal environmental sensors, and a significant number of people involved in such projects can ensure the provision of accurate environmental data through crowdsourcing to the interested party [*Dovgal and Dovgal*, 2019]. Collected in this way, the big data array will allow you to create maps of highly localized data on environmental parameters in real time (for example, temperature, NO_2 and levels of solid particles in the air) and even detect toxic chemical leaks. This information will be relevant for people living near federal highways, multi-storey parking lots or near industrial facilities. The environmental situation at such facilities is usually monitored by environmental sensors under the jurisdiction of state bodies or monitoring stations, and does not give a complete picture.

The availability of specific information in real time can not only allow sick people to avoid health problems and know the areas that should be avoided on any given day, but also gives scientists a more correct map of pollution sites, their causes and distribution over time. Based on the information received, it is possible to develop a plan for improving the environmental situation.

In addition, we can mention studies in which the processing of huge temporary data on night lights, forest fires, news databases and Flickr photos made it possible to identify World Heritage Sites affected by the conflict almost in real time [*Levin et al.*, 2019]. A non-standard example of big data analysis is an automated tracking and monitoring system for marine vessels that predicts illegal fishing in real time [*Ford et al.*, 2018]. Using a constellation of satellites and ground-based receivers, the model can detect suspicious behavior of vessels, which allows governments to conduct targeted investigations against specific vessels carrying out illegal activities in their waters. In addition to environmental justice, such a system allows preserving the potential of marine resources by preventing unregulated fishing activities.

Thus, we are on the threshold of the widespread introduction of big data technologies into the world around us and its protection. Currently, big data can be used for environmental forecasting of decision-making processes in both the public and private sectors. It should be emphasized that, in order to solve the problems of sustainable development, it is important to maintain close ties between decision makers and those affected by them. Big data analysis only allows us to provide environmental authorities with detailed evidence of rapid changes in nature: from documenting the ecology of our planet to identifying places where resources are illegally extracted. Ultimately, there must be a close link between big data analysis and a sustainable development program to ensure that humanity still has the time and space to save not only the environment around us, but also humanity itself.

References

Berman, J. J., *Principles of Big Data: preparing, sharing, and Analyzing Complex Information Waltham*, Elsevier, Morgan Kaufmann, Amsterdam, 2013.

Chen, H., R. H. L. Chiang, and V. C. Storey, Business Intelligence and Analytics: From Big Data to Big Impact, *MIS Quarterly*, *36*(4), 1165–1188, doi:10.2307/41703503, 2012.

CIO, How Cargill IT is helping to solve the world's food problems, https://www.cio.com/, Accessed 12.04.2021, 2021a.

CIO, New York Turns to Big Data to Solve Big Tree Problem, https://www.cio.com/, Accessed 12.04.2021, 2021b.

Conservation International, https://www.conservation.org/, Acessed 11.04.2021, 2021.

Dovgal, V. A., Decision-Making for Placing Unmanned Aerial Vehicles to Implementation of Analyzing Cloud Computing Cooperation Applied to Information Processing, in *Proceedings – 2020 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, pp. 1–5, IEEE, Sochi, doi:10.1109/ICIEAM48468.2020.9111975, 2020.

Dovgal, V. A., and D. V. Dovgal, Using the Internet of Things for Environmental Protection, in *Fundamental and Applied Aspects of Geology, Geophysics and Geoecology Using Modern Information Technologies: Proceedings of the 5th Intern. scient. and pract. conf. Pt. 1*, pp. 152–157, Publishing House of Kucherenko V. O., Maykop, 2019.

Dovgal, V. A., P. Y. Buchatskiy, V. V. Buchatskaya, and S. V. Teploukhov, Analysis of impact of the Internet of things on the prospects for development of e-learning, in *CEUR Workshop Proceedings. SLET 2021 – Proceedings of the International Scientific Conference Innovative Approaches to the Application of Digital Technologies in Education and Research*, 2021.

Ford, J. D., S. E. Tilleard, L. Berrang-Ford, M. Araos, R. Biesbroek, A. C. Lesnikowski, G. K. MacDonald, A. Hsu, C. Chen, and L. Bizikova, Big data has big potential for applications to climate change adaptation, *Proceedings of the National Academy of Sciences*, *113*(39), 10729–10732, doi:10.1073/pnas.1614023113, 2016.

Ford, J. H., D. Peel, D. Kroodsma, B. D. Hardesty, U. Rosebrock, and C. Wilcox, Detecting suspicious activities at sea based on anomalies in automatic identification systems transmissions, *PLOS ONE*, *13*(8), e0201640, doi:10.1371/journal.pone.0201640, 2018.

Ghani, K. R., K. Zheng, J. T. Wei, and C. P. Friedman, Harnessing Big Data for Health Care and Research: Are Urologists Ready?, *European Urology*, *66*(6), 975–977, doi:10.1016/j.eururo.2014.07.032, 2014.

Insidebigdata, How Big Data Helps in the Fight Against Climate Change, https://insidebigdata.com/, Accessed 10.04.2021.

Karamushko, G. V., and V. A. Dovgal, Materials on the organization and conduct of training case at the Factory of processes on the basics of lean production "Value stream mappin", Certificate of registration of the database 2020622312, 18.11.2020, 2020.

Levin, N., S. Ali, D. Crandall, and S. Kark, World Heritage in danger: Big data and remote sensing can help protect sites in conflict zones, *Global Environmental Change*, *55*, 97–104, doi:10.1016/j.gloenvcha.2019.02.001, 2019.

Manogaran, G., and D. Lopez, Spatial cumulative sum algorithm with big data analytics for climate change detection, *Computers and Electrical Engineering*, *65*, 207–221, doi:10.1016/j.compeleceng.2017.04.006, 2018.

PWC, Data-driven cities, https://www.pwc.ru/ru/government-and-public-sector/assets/ddc_rus.pdf, Accessed 10.04.2021.

Shepherd, A., E. Ivins, E. Rignot, The IMBIE team, et al., Mass balance of the antarctic ice sheet from 1992 to 2017, *Nature*, *558*(7709), 219–222, doi:10.1038/s41586-018-0179-y, 2018.

Sivarajah, U., M. M. Kamal, Z. Irani, and V. Weerakkody, Critical analysis of Big Data challenges and analytical methods, *Journal of Business Research*, *70*, 263–286, doi:https://doi.org/10.1016/j.jbusres.2016.08.001, 2017.

TEAM, The TEAM Network is moving to Wildlife Insights, https://www.wildlifeinsights.org/team-network, Accessed 11.04.2021.

Wamba, S. F., A. Gunasekaran, S. Akter, S. J. fan Ren, R. Dubey, and S. J. Childe, Big data analytics and firm performance: Effects of dynamic capabilities, *Journal of Business Research*, *70*, 356–365, doi:10.1016/j.jbusres.2016.08.009, 2017.

Wang, R., What a Big-Data Business Model Looks Like, https://hbr.org/2012/12/what-a-big-data-business-model, Harvard Business Review, Accessed 11.04.2021, 2012.

WWF, Keep an eye on nature with WWF camera traps, https://wwf.ru/smotri, Accessed 14.04.2021.