Complex Information E-Science System Architecture based on Cloud Computing Model

[1]Artem Kazarian[0000-0002-6883-0233], [2]Nataliia Kunanets[0000-0003-3007-2462], [2]Volodymyr Pasichnyk[0000-0003-3007-2462], [2]Nataliia Veretennikova[0000-0001-9564-4084], [2]Antonii Rzheuskyi[0000-0001-8711-4163], [3]Andrii Leheza[0000-0002-8892-2426], [2]Oksana Kunanets[0000-0001-8626-9016]

[1]Computer-Aided Design Department, Lviv Polytechnic National University, Lviv, Ukraine
[2]Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine
[3]Software Supply Systems Department, Uzhhorod National University, Uzhorod, Ukraine
artem.kazarian@gmail.com, nek.lviv@gmail.com, vpasichnyk@gmail.com, nataver19@gmail.com, antonii.v.rzheuskyi@lpnu.ua, andrewlegeza@gmail.com, oksanka.kun@gmail.com

Abstract. The adaptive information system was developed to implement the concept of information and technological support of scientific research in virtual teams that is multicomponent and technologically convenient for making necessary changes, which fulfil procedures for the dynamic development of the electronic science platform. The system is deployed on the SaaS platform as one of the most popular models of cloud computing. The authors present the decomposition of the system, providing a detailed functional description of the system modules.

Keywords: information system, e-Science, cloud computing, virtual team, Saas model, microservices, virtual laboratory.

1. Introduction

In the studies of ten or fifteen years ago, the concept of electronic science was considered as an information technology platform exclusively of two classes of information technologies, namely Grid Computing and Cloud Computing. We distinguish five classes of information technologies that create an information and technology platform of e-Science: Grid technology, Cloud technology, Big data technology, Big data analytics technology and Internet of Things. Each of these classes of technology has its own specifics, features and implementation. The main aim is to integrate the five classes of technologies at the infrastructure level to ensure the proper high-quality, full and effective functioning of virtual research and development teams, which can be geographically distributed in different time zones for solving complex research, scientific and technological, scientific and innovative works.

All five classes of these information technologies form the information technology platform for supporting electronic science. That is, conducting full cycles of research

and development from inception of ideas to the introducing it in practice. The main aim of using an innovative concept of e-Science is an information and technological support of the activities in virtual research and creative teams, which are distributed geographically and temporally, in order to solve general common scientific and research problems.

2. Analysis of recent researches and publications

Nevertheless, the analysis of recent studies due to the definition of the e-science concept was conducted.

Javad Pourabbasi, Abdollah Aghaie, and Mojtaba Hajian Heidary [1] proposed e-Science as a service-oriented architecture. In their opinion, the components of the developed architecture are data, computing, and communication. This architecture aims to generate and share scientific information and knowledge.

Abel Carrión, Miguel Caballer, Ignacio Blanquer, Nelson Kotowski, Rodrigo Jardim, Alberto Martin Rivera Dávila [2] convinced that scientific workflows are widely used for modelling processes in e-Science. Scientific workflows are executed using appropriate management systems that organize the workload at the top of the computing infrastructure. The emergence of cloud computing infrastructures and the use of infrastructure on request made it possible to replace the local infrastructure. They introduce solution to manage workflows that use cloud-based orchestras to deploy on-demand resources to load workloads and combine heterogeneous cloud providers such as satellite clouds, public clouds and traditional infrastructures to minimize time for obtaining scientific information.

According to Roberto C. S. Pacheco, Everton R. Nascimento, Rosina O. Weber [3] e-Science is a new infrastructure not only for computing but also as a virtual laboratory where researchers and professionals in technology can collaborate and share data, information and knowledge. The term electronic infrastructure is used as a synonym for cyberinformation infrastructure or as an infrastructure component of cyber infrastructure. In this case, e-science is seen as a scientific method that involves the absorption of the best digital platforms. Electronic science and electronic infrastructure are considered interchangeable.

The authors of the article [4] emphasize that the work of virtual teams is becoming increasingly important today. Therefore, researchers investigated the effectiveness of teamwork of remote members in virtual teams, using only channels for instant messaging.

Roberto C. S. Pacheco, Everton R. Nascimento, Rosina O. Weber conceptually represent e-science as a multi-layer system which consists of crowdsourcing, collaboratories, Open access, and Data Science. Interdisciplinary research teams in e-Science use methods (community-based monitoring, scientific workflow systems, semantic e-Science, linked data, network ontology) with the latest information technologies (service-oriented computing, cloud computing, on-demand access grid computing, connectivity technologies).

A. C. Zhou, B. He, S. Ibrahim [5] compared the advantages and disadvantages against e-Science in the Grid and observed the obstacles and opportunities of e-Science services in the cloud. Grid computing has greatly enhanced the development

of e-Science, as many e-Science applications have become guided data, which initiated a free exchange of information and knowledge and consolidation of information arrays.

Cloud computing has the benefits of scalability, high bandwidth and easy accessibility compared to Grid. As noted, e-Science projects in various research areas have migrated from Grid to cloud platforms. Besides, A. C. Zhou, B. He, S. Ibrahim considered e-Science as a component of Big Data. But we completely disagree with this statement. E-Science is also considered as a "big-data science", which includes a wide range of studies in various fields of knowledge. Considering Big data that is inseparable part of e-Science, an attempt has been made to develop and implement four science applications that are life sciences, physics sciences, social and human sciences, and climate and earth sciences on a scalable cloud infrastructure.

3. Innovative platform for conducting research

Conducting modern research in many technologically advanced countries undergoes dramatic changes due to the introduction of an innovative approach called e-Science. This term was first used by John Taylor, chairman of the UK Department of Education and Technology in 1999. By Taylor, the term e-science means global collaboration in key areas of science and innovative infrastructure development that will enable such collaboration [6]. M. Riedel [7] defines e-science as a new field of research based on collaboration in key areas of science and uses innovative infrastructure to expand the capacities of scientific calculations. A. Bosin, N. Dessie, B. Pes understand e-science as a type of large-scale interaction within the typical scenario for conducting an experiment that requires multiple use of data (preliminary processing, calculations, post processing and received data storage) [8]. We consider e-Science as a system-integrated complex of information, telecommunication and social and communication computer technologies, which ensure the fulfillment of the functions and the solution of the actual tasks of science in the information society [9].

The above definitions are based on the assertions that a lot of complicated and topical scientific problems are solved by large groups of globally distributed scholars who need an access to modern computer infrastructure to provide data, computing and network resources. Today, the key areas for the application of e-science, according to leading experts, are physics, astronomy, biology, medicine and computer science.

In the context of the information and communication infrastructure development, the problem of providing open access to information resources, that are created as a result of conducted scientific research, and the expansion of scientific communication becomes more and more relevant. Scientific communication is defined as a complex system whose environment is intended to facilitate the intellectual exchange of information through a wide range of means.

Dissemination of research results, theoretical foundations and concepts, as well as social communication, which provides support for a dialogue aimed at improving knowledge in a certain thematic or subject area, are implemented through its channels. In the conditions of the information society and the development of information technologies, scientific communication passes into the electronic environment due to the reorganization of the information needs of users, because

more and more users need to develop open access to documents. This need generates open access magazines, institutional repositories and funding support for open access initiatives and promotes the formation of an innovation complex called “e-science”.

In the context of modern approaches to the scientific communication processes organizing information and technological support of scientific research, the methods of analysis and synthesis of communicative information are changing. In order to increase the efficiency of scientific communication processes in conducting research on the e-science platform, their information and technology support, the authors developed a project of a complex information system that facilitates the formation and processing of information resources using the Big Data technology, taking into account the factor of their poorly structured nature. The expansion of the system is decided to be carried out on the SaaS platform, which is one of the most popular cloud computing models. The system development pursued the goal of solving key problems for the establishment of effective scientific communication at each stage of scientific research: the formation of a team for scientific research; solution of financial aspects of scientific communication; representation of scientific data, information and knowledge; interaction of scientists; search, association, storage and dissemination of scientific knowledge.

The concept of a virtual team should be backed up by appropriate architectural and technical solutions [9] that contribute to a certain outcome.

Foreign scholars present several definitions of a virtual team that arose because of long-term discussions of specialists in this area [10]:

1. The virtual team is a temporary set of independent representatives from different institutes that, using information and communication technologies, are joined to conduct research. They integrate vertically to combine the key competencies of individual team members and act as a single organizational unit.
2. A virtual team is an identified group of people or organizations that, through the widespread use of information and communication technologies, are united for business or scientific collaboration.

The virtual team predominantly creates a virtual organization. A virtual organization is a constantly evolving organization, redefining the direction of its work to achieve the practical goals of a particular case, combining geographically distant employees [10].

The main task of the virtual team is purposeful management, whose activities are aimed at achieving the goal. The directions of the activities of virtual organizations can be varied such as production, marketing, distribution, research and development, etc.

Despite all the advantages of a virtual team, during the research there may be some obstacles, first of all communication and information barriers. A number of factors, such as geographical and linguistic, are affected by this.

During an independent scientific search, a user prefers to use a limited number of resources that are presented in open access. The search process involves several steps that are accompanied by a preview of the selected material. Members of virtual teams need to receive information in a convenient form and easy way. This can be achieved by providing on-line access to large amounts of information [11, 12]. However, these information arrays should be processed properly.

4. The need to create an innovation system

The use of modern programs and applications used in the process of scientific research simplifies the scientists' work; at the same time their diversity complicates the process of selection. The processes of performing the same functions differ in the software products of different companies. There is also a problem of observance of uniform standards for storing files and scientists' communication channels. Each scientist prefers to use a usual software product for them to store the results of scientific researches and their design.

While delivering them to their colleagues who use other software products[13,15] for viewing, editing and storing similar types of information, there is a need to reconcile an identical file format and convert the files to the selected format. The variety of programs for communication between scientists generates the need to create a single software product for textual, audio and video communication between researchers, allowing scientists to join the multinational teams and be able to communicate among themselves as a part of a single project. The requirements for the publication of scientific achievements differ in the policy of journals in which they are published. Such differences in standards and requirements using different software products, which in turn perform similar functions in the process of scientific research, create barriers to communication within the scientific community and slow down the work on scientific projects and the speed of dissemination of research results.

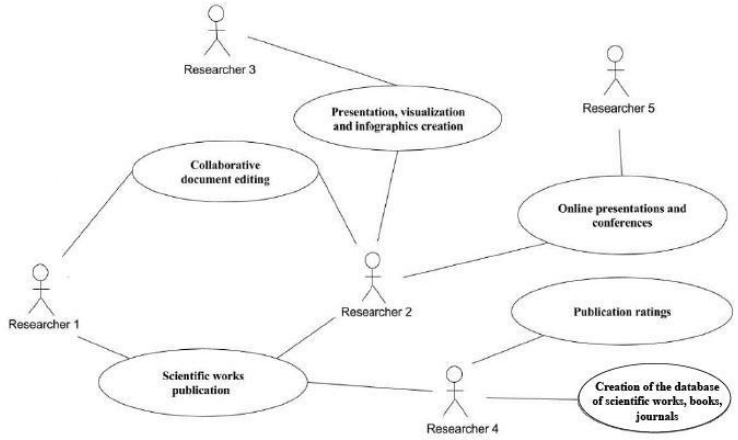

Fig. 1. Case diagram of system users

To eliminate these inconveniences, it is proposed to develop a complex information system focused on being used by scientists at all stages of the research project life cycle. The information system for unifying the processes of work on scientific projects optimizes the time expenditures of scientists during the creation and storage of scientific results and the communication of researchers during the work of distributed teams over one project. The complex scientific system will also be useful for disseminating the obtained scientific results through the creation of scientific publications, presentations and visualizations of scientific research. The processes of

interaction of the scientific team participants are shown in Fig 1.

In order to implement the proposed concept of information and technological support of scientific research, an adaptive information system, that is multicomponent and technologically convenient for making necessary changes that implement procedures for the dynamic development of the electronic science platform, has been developed. The methodology of the system approach was used in the process of its development that allowed considering dependencies existing between the components of the system as subsystems and components, taking into account not only the current requirements and opportunities, but also the prospects for its development. System integration with a cloud-based data warehouse provides convenient twenty-four-hour access for all users of the system to this functionality and avoids the need to store large volumes of data on the client's part of a user. The developed system deploys in 16 modules, which have full code isolation; the only way to execute the code in these services is to have an HTTP request, such as a user request or a call to the RESTful API. The code in one service cannot directly call the code in another service. The software code is deployed for each service independently, and diverse services are written in different languages, such as Python, Java, Go, NodeJS and PHP. Auto-scale and load balancing processes are driven by independent services.

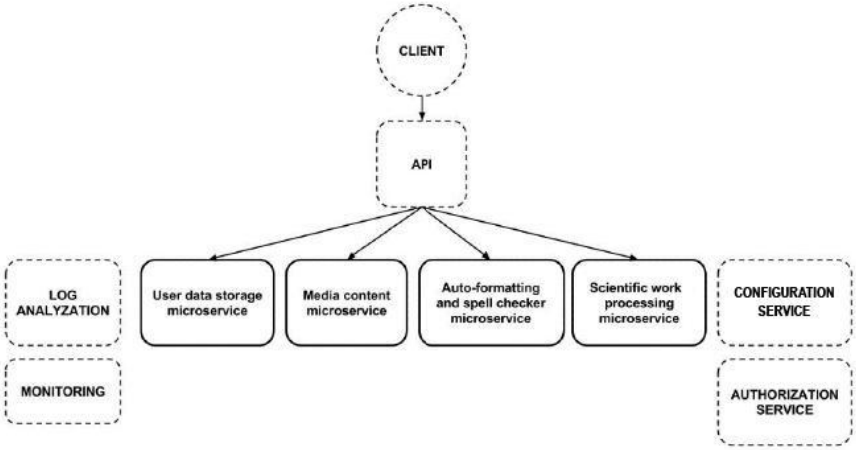

Fig. 2. System microservices

Each module assumes the implementation of a number of microservices. In addition, each service has several deployed versions at a time. For each service, one of these versions is a working version by default, but at the same time it provides direct access to any deployed version of the service, because each version of each service has its own separate address. This structure provides many features, including testing the latest version, testing interdependencies between different versions and simplifying the operations of returning to the previous version.

5. Functional description of system modules

The list of the system functions is divided into two categories, such as social and scientific. The social category includes functions that provide the identification of system users and the establishment of communication methods between them. The scientific category includes functions that are directly used in the implementation of scientific projects.

Social component. The system social component is composed of 5 modules.

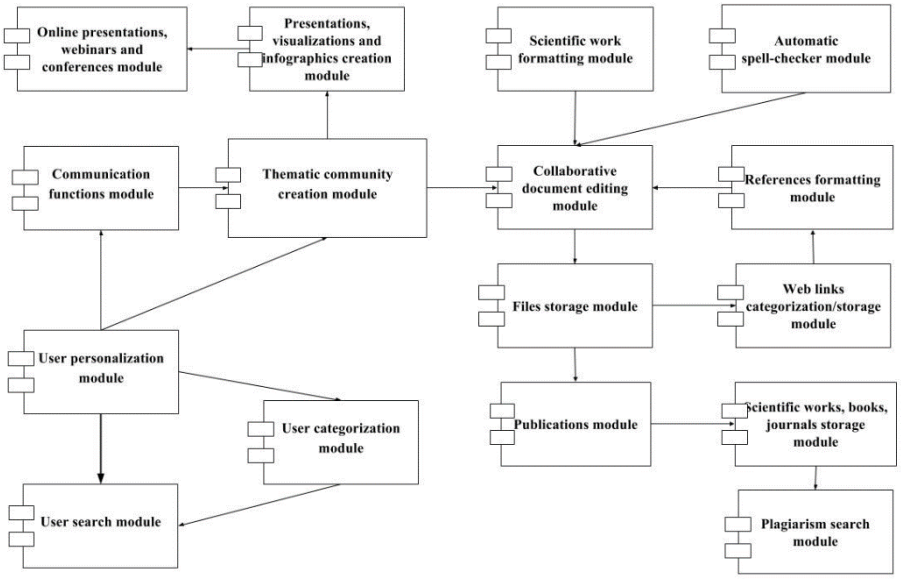

Fig. 3. Component diagram

The **User personalization** module provides identification of users and storage of their personal data. To do this, it is offered an implementation of a new user registration function, which allows you to have a personal profile for each user of the system and to allocate users to separate groups to provide a personalized set of system functions, as well as to save the personal settings of the system for each individual user.

The **User Categorization** module divides users based on criteria like academic degree, research institute, country and city, as well as information needs. Clusterization is used to facilitate the search for researchers, who conduct research in related fields and the search of colleagues for communication and information exchange.

The **User Search** module provides users with a variety of search criteria in the database, such as name, surname, scientific degree, position, country and city of residence, research institute and scientific interests. This information facilitates finding colleagues with common scientific interests for further exchange of knowledge, communication, joint research, writing and publication of joint research papers, and the development of new training courses based on the collected information.

The **Creating thematic communities and teams to work on the project** module provides opportunities for joining into groups of people with similar scientific views and interests. Each community has a specific subject of communication, which allows users to share information with other researchers. Scientists united in the community can easily select team members for collaborative researches, writings and publishing collaborative research papers.

The **User Communication Functions** module provides communication of the system users through text chats, audio / video calls and conferences. This functional, embedded in the system, helps to install and integrate a variety of different software to provide unified communication standards among system users.

Scientific component. System scientific component contains 9 modules.

The **Web link categorization/storage** module, by analogy of CiteULike, has a built-in storage functional for saving online links to documents stored on the Internet, allowing users to collect and store links for research, manage reference data and obtain bibliographic information from the web-pages. This functionality facilitates the creation of relevant information resources and provides colleagues with information on resources for joint research.

The **Database of scientific works, books, magazines** module gives the formation of its own bibliographic and abstract database, which will serve as a tool for tracking the citation of scientific publications of participants in the research team and will act as an information system, that generates statistics, as well as characterizes the state and dynamics of indicators of demand, activity and impact of the activities of individual scholars and research organizations. The system database is synchronized with the leading science-based databases, which will become a complex information system of scientific works.

The **File and Software Saving** module ensures the storage of files, involved in scientific communication, and online access to them. The functionality of the module allows integration with GitHub, Bitbucket, Google Drive and Dropbox web services, and allows you to preserve the access levels to files of various categories of system users, who work on several scientific projects. The system generates a history of changes to each file with details of the specific file change, author, and time of change (Fig. 4).

Most scientific works are co-authored by several scientists, that is why there is a problem of the authors' interaction during the writing of scientific works. The **Collaborative document editing** module allows multiple users to work on text files simultaneously with synchronizing changes of all users, who work on a document using cloud computing and cloud storage technology.

After the completion of scientific researches, the stage of the presentation of scientific results to the scientific community and their approbation comes. With the help of the **Presentation, visualization and infographics creation** module and using integrated presentation editors (for example, Microsoft PowerPoint software), information visualization of the received results is formed. This approach allows us to standardize the format of presentations of scientific works and to reduce the time expenditures for the creation of infographics of scientific research results based on the proposed templates. This functionality is developed as a separate web application, which is an integrated part of the proposed system.

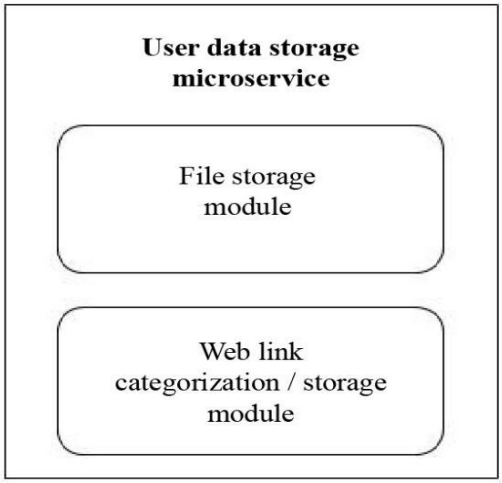

Fig. 4. User data storage microservice

The **Online presentations, webinars and conferences (video communication, screen display and joint chat) module** gives an opportunity to conduct the presentation of the scientific research results to the scientific community of the experts in this field, in which research is conducted, as well as to increase the number of participants in scientific communities that participate in the processes of information and knowledge exchange.

The module functional creates an opportunity to join the discussions (events) with the help of the developed system functional, regardless of its residence place.

The system will allow participants to present the video stream to the participants, display the online presentation slideshow screen in real time, and make discussions in common text and video chats.

With the help of the module functional, it is provided the video stream of a presenter, the image of the screen with slides of the presentation; organizing text and video chat for further discussion (Fig.5).

The Scientific work formatting module provides a reduction of the time costs of scientific staff for the scientific papers designing in accordance with the requirements of journals and established standards by being integrated into the system of the most common scientific paper formatting templates from the relevant professional journals in the research subject field.

This functionality is developed as a separate web application, which is a part of the proposed system.

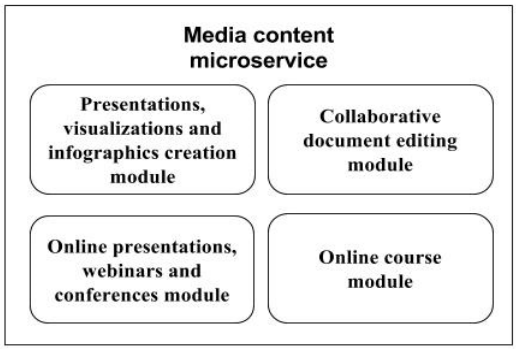

Fig. 5. Media content microservice

Scientists who draw up the results of scientific research and submit the state of the research problem based on publication analysis, according to the results of predecessor researches, are required to indicate the metadata of each information source that has been processed. The formats used to register the sources vary, depending on the type of an article, the type of publication, and the requirements of the publisher. The reference formatting module provides users with an opportunity to use the functionality to automate the formatting of metadata sources according to the user-specified format, which will speed up the process of formatting a scientific paper. The module integrates with leading bibliographic managers such as Mendeley, EndNote, Zenodo. Most scientific papers are published in English, regardless of the author's residence country that in turn ensures the dissemination of research results among scholars from different countries. The Automated spell and punctuation checking module facilitates the process of writing scientific work, providing not only text testing, but also giving tips on using synonyms, which improves the quality of the text presented by scientists who are not native English speakers (Fig.6).

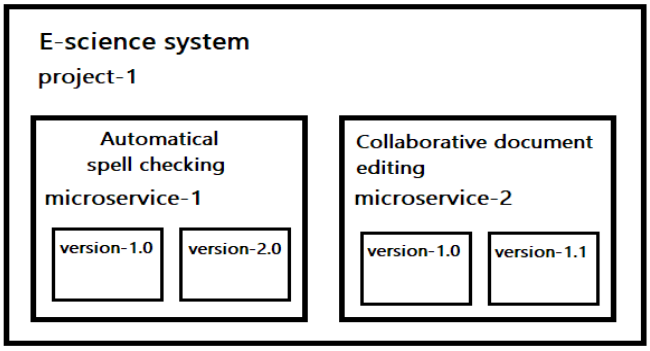

Fig. 6. E-science system microservice versioning

The Plagiarism search module provides an opportunity to test the research text adherence to the principles of academic integrity, originality of the text and rules of citation. The system will display the result as general text uniqueness in percentage,

as well as uniqueness of each found matching page. This module uses the access to scientific texts depositaries to find the probability of a coincidence of published scientific works both by an author and colleagues. The functionality of the module allows the implementation of plagiarized testing systems, presented on the IT market of software products. The Publication module ensures the publication of the scientific research results on the concept of open data for the scientific community, which includes such activities as public research publications, stimulation of scientists to use "science with open notes", and, in general, facilitates the scientific knowledge dissemination.

6. Conclusions

Thus, the presented architecture of the information system consists of variety of modules. Each module performs a certain set of functions, providing information and technological support for all stages of the research life cycle, such as data collection, analysis, presentation of the results, registration of scientific publications, archivation, distribution and design of the publication text and patches. The proposed system is designed primarily for information support of virtual teams, especially:

- providing unification of information storage and scientific communication channels among distributed members of virtual teams;
- improving the efficiency of scientific communication;
- the model of SaaS and data storage in the «cloud» allows to distribute optimally the computing load of the system;
- members of virtual teams have an opportunity to access the system 24/7.

The practical approbation of the presented system took place in two virtual research laboratories: in the Ternopil intelligent city laboratory at Ternopil Ivan Poluj National Technical University and in the Modeling of social and communication processes and systems laboratory at Rivne State Humanitarian University.

References

1. Pourabbasi, Javad, Aghaie, Abdollah, Hajian Heidary, Mojtaba: A service-oriented architecture for generating and sharing science and knowledge in the e-science environment. Journal of Information Processing and Management, 31 (183), 269–297 (2015).
2. Abel, Carrión, Miguel, Caballer, Ignacio, Blanquer, Nelson, Kotowski, Rodrigo, Jardim, Alberto Martin, Rivera Dávila: Managing workflows on top of a cloud computing orchestrator for using heterogeneous environments on e-Science. International Journal of Web and Grid Services, 13 (4), 375–402 (2017).
3. Roberto, C. S. Pacheco, Everton R., Nascimento, Rosina O., Weber: Digital Science: Cyberinfrastructure, e-Science and Citizen Science. Knowledge Management in Digital Change, 377–388 (2018).
4. Carlson, J. R., Carlson, D. S., Hunter, E. M., Vaughn, R. L., George, J. F.: Virtual team effectiveness: investigating the moderating role of experience with computer-mediated communication on the impact of team cohesion and openness. Journal of Organizational and End User Computing, 25(2), 1–18. (2013).

5. Zhou, A.C., He, B., Ibrahim, S.: Chapter 18 – eScience and Big Data Workflows in Clouds: A Taxonomy and Survey. Big Data, 431–455 (2016).
6. Taylor, J. M.: Talk given at UK e-Science Town Meeting in London. (2001).
7. Riedel, M.: Classification of different approaches for e-science applications in next generation computing infrastructures. In: Proceedings of the e-science Conference, Indianapolis, Indiana, USA, 2008, 198–205 (2008).
8. Bosin, A., Dessi, N., Pes, B.: Extending the SOA paradigm to e-Science environments. Future Generation Computer Systems, № 27, 20–31 (2011).
9. Lypak, H., Rzheuskyi, A., Kunanets, N., Pasichnyk, V.: Formation of a consolidated information resource by means of cloud technologies". In: 2018 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T, 157-160 (2018).
10. Luis M. Camarinha-Matos, Ivan Silveri, Hamideh Afsarmanesh, Ana Ines Oliveira: Towards a framework for creation of dynamic virtual organizations. Collaborative Networks and Their Breeding Environments: IFIP TC5 WG 5.5 Sixth IFIP Working Conference on VIRTUAL ENTERPRISES, Spain. Valencia, pp. 69–81 (2005).
11. Rzheuskyi, A., Matsuik, H., Veretennikova, N., Vaskiv, R.: Selective Dissemination of Information – Technology of Information Support of Scientific Research. Advances in Intelligent Systems and Computing 871, pp. 235-245 (2019).
12. O. Grigorov, E. Druzhynin, G. Anishchenko, M. Strizhak, V. Strizhak, “Analysis of various approaches to modeling of dynamics of lifting-transport vehicles," in International Journal of Engineering and Technology(UAE), № 7 (4), 64-70 (2018).
13. D. Peleshko, T. Rak, I. Izonin, “Image superresolution via divergence matrix and automatic detection of crossover," in International Journal of Intelligent Systems and Applications, vol. 8, №.12, 2016, 1-8 (2016).
14. D. Peleshko, T. Rak, M. Peleshko, I. Izonin, and D. Batyuk, "Two-frames image superresolution based on the aggregate divergence matrix," 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, pp. 235-238 (2016).
15. Tkachenko, R., Doroshenko, A., Izonin, I., Tsymbal, Y., Havrysh, B. Imbalance data classification via neural-like structures of geometric transformations model: Local and global approaches. Advances in Intelligent Systems and Computing 754, 112-122 (2019).