Hierarchical Representation of Information Objects in a Digital Library Environment

Nikolay Kalenov, Irina Sobolevskaya, and Alexandr Sotnikov(✉)

Joint Supercomputer Center of the Russian Academy of Sciences — Branch of Federal State Institution "Scientific Research Institute for System Analysis of the Russian Academy of Sciences" (JSCC RAS — Branch of SRISA), Leninsky av., 32 a, 119334 Moscow, Russia
asotnikov@jscc.ru

Abstract. This paper studies the problem of building a modern information society as a task of forming a virtual knowledge space. This task can be implemented on the IT platform of a digital library, which ensures the formation and provision of information resources in various areas to general public users. It shows how electronic copies of objects of a library, archive and museum storage represented in the form of texts, graphic images, and audio/video objects, including 3D-models, can be integrated using digital library facilities. The concept of a hierarchical level of electronic objects is introduced. The definitions of objects of various levels are given and the principles of work with objects of each level are formulated. The hierarchical representation of electronic objects in a digital library environment is proposed. A distinction between thematic, theme-specific, and interdisciplinary collections is shown. The information environment for forming subtheme collections is represented as a set of databases created by information fund holders. This environment is distributed and unified in terms of the software and hardware used, as well as in terms of a set of requirements for digital images being formed. The paper formulates the basic principles of forming collections of various levels in a digital library environment and reflects the developed digital solutions that ensure their implementation. For each hierarchical level, it gives examples of objects and collections provided on the portal of the Scientific Heritage of Russia Digital Library (SHR DL) and formed in compliance with the proposed principles and digital solutions. In particular, the paper describes in detail the thematic collection of publications in mathematics being a database formed through LibMeta DBMS; interdisciplinary collection "Garden of Life" created in collaboration with one of the Russian museums, and some other collections.

Keywords: Scientific Heritage of Russia Digital Library · Theme-specific collections · Thematic collections · Interdisciplinary collections · Virtual exhibitions · Digital 3D-models · Museum objects

© Springer Nature Switzerland AG 2019
S. O. Kuznetsov and A. I. Panov (Eds.): RCAI 2019, CCIS 1093, pp. 93–104, 2019.
https://doi.org/10.1007/978-3-030-30763-9_8

1 Introduction

The formed digital knowledge space is one of the most important elements of the modern information society. Access to objects of the knowledge space is provided via the Internet, which offers many opportunities for combining various information sources, extracting knowledge and forming a virtual information space on their basis. An effective means of integrating information resources is a complex of technological, technical, and organizational solutions united by the concept of a digital library that ensures the formation and provision of information resources in various areas to general public users.

There exist several integrators that combine different types of information resources. An example is the European electronic library Europeana [1], which aims at combining digital images of European cultural heritage sites and making their content available to users. At the same time, Europeana provides access to full texts of electronic printed publications. The other example of the pooling of resources is the European library [2], which is a portal that provides a free search for resources available to the leading national and scientific libraries in Europe. From 2011 to 2017, the European library managed the Europeana Libraries project, which was created to include more than 5 million objects from 19 scientific libraries in Europe. In 2017, the European library project was closed, and all the processes associated with the unification of European digital objects of cultural significance is provided by Europeana.

Unlike the European media, most cultural and scientific centers of the United States have long been creating their digital collections and cooperate only for the organization of thematic projects. Since 1990, the Library of Congress has launched an American Memory project called the National Electronic Library. However, this project did not involve pooling the resources of most US public libraries. There were many thematic library projects in the country without centralized management. In 2013, the electronic resources of libraries, universities, museums, and archives were combined and made available through a single portal [3], the electronic public library of America (DPLA), which should provide online access to the cultural heritage of the whole country. DPLA, unlike Europeana, does not contain digital copies of documents but provides metadata about their resources transferred by participating libraries. The DPLA database contains records redirecting the user to the provider library site that contains the corresponding digital document. The digital public library of USA portal serves as a single point of access to millions of objects, providing search across all DPLA-enabled digital libraries [4].

Several large aggregators of information resources described above provide access not only to digital copies of printed publications, but also allow viewing thematic collections or virtual exhibitions provided by participants [5]. However, none of the above resources allows to dynamically form user collections of the presented information objects united by some features [6].

In particular, the task of integrating electronic objects of library storage, archival storage and museum storage presented in the form of text files, graphic images, and multimedia into a unified resource is solved with the help of electronic libraries [7].

We have considered in sufficient detail the approaches to the digital library design and principles of their formation using the example of the Scientific Heritage of Russia Digital Library (SHR DL) in [8–10].

SHR DL provides integration of electronic copies of scientific objects stored in institutions of memory (libraries, archives, museums), presented in the form of text files, graphic images, and multimedia. Objects that are reflected in the SHR DL are accompanied by extensive information about their creators, scientists who have made a significant contribution to the development of Russian science.

SHR DL is based on a combination of the principles of decentralization (preparation of metadata and digitization of materials by resource owners) and centralization (editing, storing information and providing it to users upon their requests).

The main functionalities of SHR DL include:

- formation and storage of content provided by institutions of memory;
- search for metadata objects of electronic funds of institutions of memory;
- search, metadata about scientists and publications;
- thematic search;
- viewing of collections available in the SHR DL;
- forming and editing custom collections;
- searching by collection;
- work with media objects, allowing you to view media objects;
- viewing full texts of digital copies of publications;
- import of metadata from external systems.

User access to the SHR DL is through any browser. The basis of the SHR DL software is the "LibMeta" platform. The core of this platform performs the following functions [11]:

- management of the static content of the SHR DL;
- storage of objects of the SHR DL represented by RDF-triples in relational Database Management System (DBMS);
- batch download;
- indexing;
- full-text search;
- system security;
- news management.

The participants of the project for the creation of the SHR DL (scientific libraries, museums, archives of the Russian Academy of Sciences and several scientific institutes) form and edit the metadata of their resources on-line in the technological database. The ontology of the SHR DL envisages that each object to be included in the SHR DL is characterized by a set of properties fixed for each type of objects (persons, publications, museum objects, etc.). As the formation (editing) of the object metadata is completed, they are loaded in batch mode into the central database with the automatic establishment of links with other objects.

It should be noted that SHR DL is developing as an integral part of the Russian digital knowledge space towards the integrator of scientific information resources of memory institutions.

The formation and maintenance of information funds of digital libraries and the provision of access to them require a specific technological environment that supports relevant procedures. The study of the fund creation process and the nature of material requests has demonstrated that users often need to select information objects from the entire set of interconnected digital library resources united by one feature or some combination of them. This, in turn, provokes a need to develop and analyze the hierarchical representation of digital objects in a digital library environment. To build the above hierarchy, let us introduce the following concepts:

"objects" – images of items of information funds;

"type of an object" – an object property that reflects its membership in one of the resource groups determined in a digital library space (publications, archive materials, museum objects, etc.);

"theme of an object" – an object property that reflects its content identity with the group determined by a certain semantic concept (e.g. a section of science, historical period, attitude to this person, etc.);

"collection" – a set of objects of one or several types and (or) themes.

The proposed hierarchical representation of objects in a DL environment containing 4 levels is given in Fig. 1.

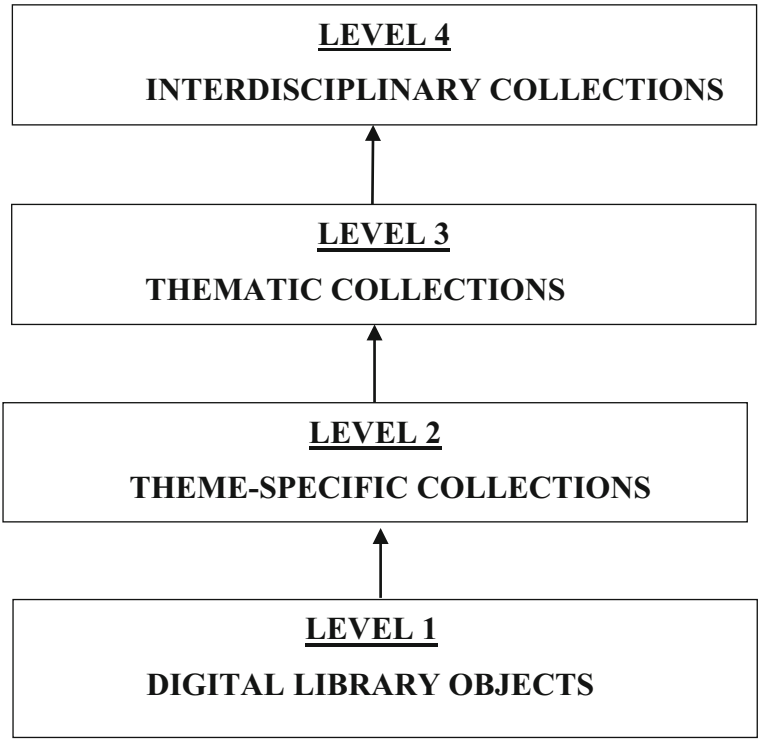

Fig. 1. Hierarchical levels of information objects.

Combining digital objects of various hierarchical levels according to some features allows developing and implementing numerous analysis and processing procedures for them, as well as representations according to the needs of the user. At the same time, it significantly enhances the ability to reflect the results of the search in digital library funds: from exhibiting a separate object to providing collections created by it and the arrangement of virtual exhibitions.

2 Elements of Hierarchical Levels of Information Objects

The elements of **Level 1** are digital library **objects**. The elements of the first level can be presented by digital copies of printed publications, archive materials, museum objects, audio/photo/video items.

Consider collections of objects of the first level relating to the same type and theme as elements of **Level 2**. For example, a collection of e-books in a given scientific area or a collection of museum objects of the same type (collection of minerals, sculptures, etc.). Objects formed in this way are called **theme-specific collections**.

Consider collections of objects of various types relating to a particular topic as elements of **Level 3**. For example, a collection dedicated to a certain scientific area, event or person, which includes at least two types of objects on a given topic: a collection of books, archive materials, museum objects. Such collections are called **thematic**. For example, the thematic collection on the botany of the 1920s–1930s may include a collection of books on this topic, biographies of biologists, as well as multimedia materials of the period, etc. [12, 13].

Objects of **Level 4,** which are formed by combining objects of the previous level, are called **interdisciplinary collections**. Thematic collections (Level 3) may exist by themselves, but may also be a part of interdisciplinary collections. If a collection includes materials relating to various fields of science (knowledge) and intersected by one or several parameters, then such a collection comprises several thematic collections. For example, an interdisciplinary collection dedicated to space exploration may include materials on astronomy, space physics, history of cosmonautics, as well as those related to the problems of space exploration, etc.

Collections of natural history museums are of particular interest for applied and fundamental research. They are used for informative and educational purposes [14]. One of the ways of representing interdisciplinary collections in a distributed digital library environment is to form a virtual exhibition. A virtual exhibition is a multimedia information resource in the Internet environment, demonstrating to the user heterogeneous pieces of information (digital copies of printed materials, archive documents, museum items, etc.), combined according to given characteristics. The chief distinction of a virtual exhibition is the provision of information in an interactive form. Along with materials of various types being represented, multimedia objects, digital 3D-models of museum objects, in particular, are necessary to form digital natural history collections. Virtual exhibitions, in contrast to museum exhibitions, are not limited by the duration of an exhibition [15, 16].

It is important to note that the arrival of virtual exhibitions in a digital library environment may be one of the directions in the integration of heterogeneous resources and representation of the digital museum content.

2.1 Examples of Information Objects of Various Hierarchical Levels in a Digital Library Environment

Consider examples of objects of various levels represented in the Scientific Heritage of Russia Digital Library.

Objects of Level 1 are presented by “scientists”, “publications”, “museum objects”, etc., available in the portal for SHR DL (http://e-heritage.ru) and selected by the user. [9]. A search in this portal can be carried out by several parameters with a fragment of the field value entered for a search indicated. Figure 2 shows the result of the search of D. Bernoulli’s letters to L. Euler published in Latin (Excerpta ex litteris a Daniele Bernoulli ad Leonhardum Euler). The displayed bibliographic description is an active link, following which you can get a complete text of this publication (Fig. 3), is an object of Level 1 in this case.

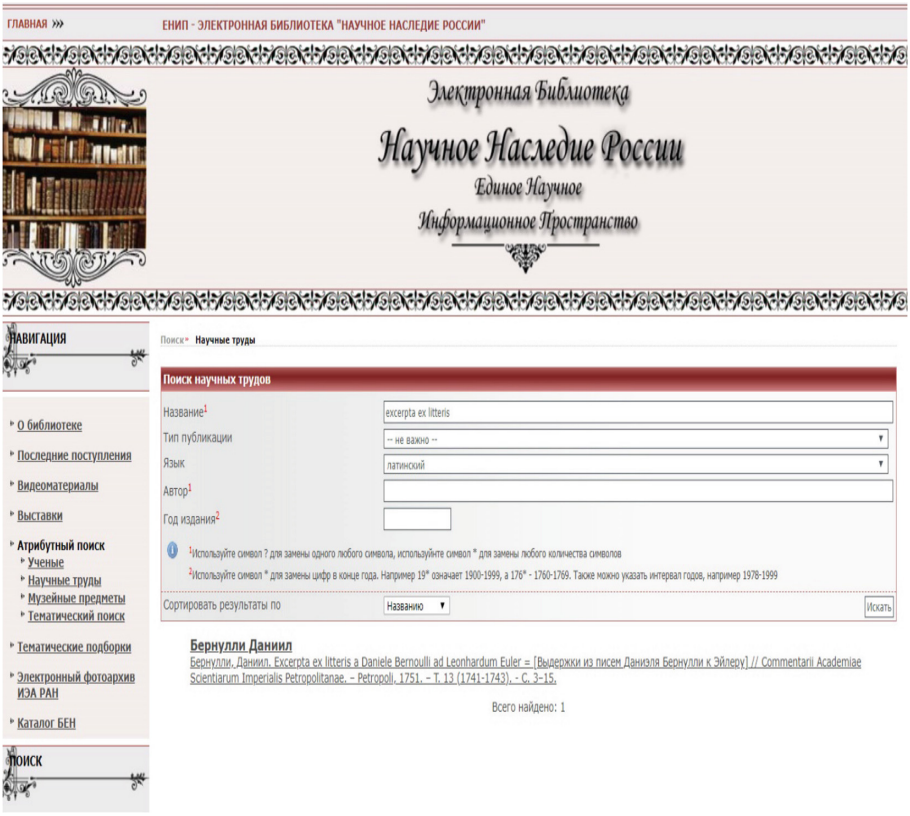

Fig. 2. Search result of a particular publication.

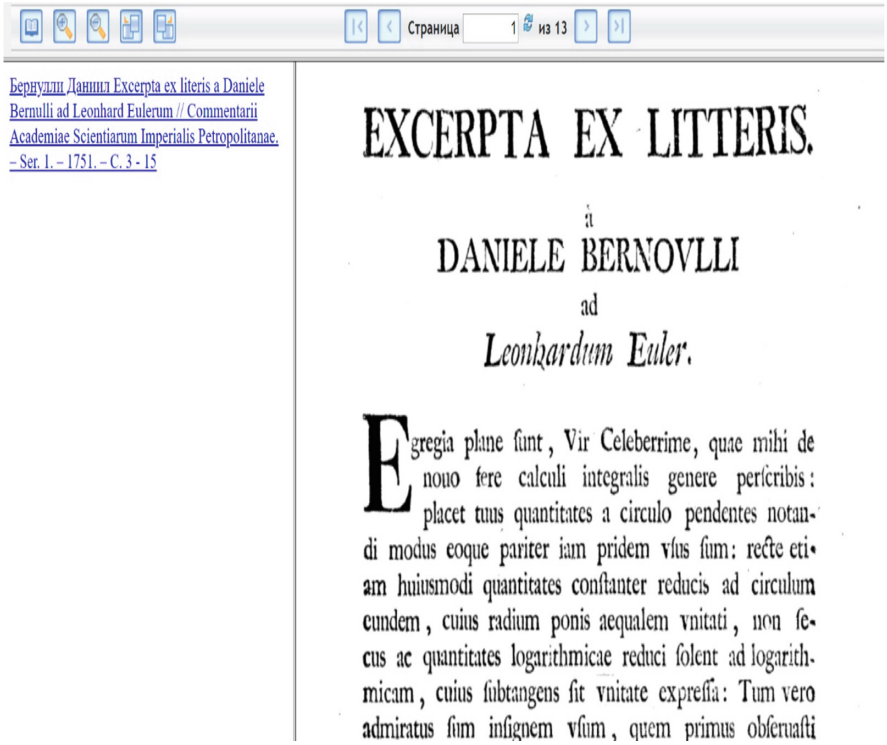

Fig. 3. Fragment of the e-book "Excerpts from Daniel Bernoulli's Letters to Euler".

Theme-specific collections are most popular among the users of the Scientific Heritage of Russia Digital Library. They allow, for example, to see the overall picture of the development of a specific scientific area or to form an updated bibliographic collection of scientific works. Some theme-specific collections formed in the information space of the Scientific Heritage of Russia Digital Library are given below.

One of the examples of a theme-specific collection is publications in mathematics available in the digital library. This collection contains bibliographic data and complete texts of 969 publications in mathematics.

The majority of the collection publications are in Russian (70,07%). In addition to publications in Russian, the collection includes books in Latin (21,67%) and French (5,06%). Publications in English, German and other languages make up about 3,2%.

The percentage distribution of available publications into periods is given in Table 1.

A dramatic reduction in the number of available works since the second half of the 20th century is primarily due to the copyright law, following which SHR DL members prepare materials to be included in the Library.

Another example of a theme-specific collection is "Scientific publications of XIX century in the Scientific Heritage of Russia Digital Library". The main distinction of this collection from the previous one is that (a) in this case, the selection of publications

Table 1. Time distribution of publications

Period	%
1700–1799	20
1800–1899	14
1900–1949	54
1950–1999	10
2000–2018	6

is limited to a specific historical period – XIX century, (b) this collection includes works in various scientific areas. This collection includes 4848 scientific publications published in 1800–1899 in Russian, English, German, French, Latin and other languages.

An example of a **thematic** digital collection formed by SHR DL is "Minerals in P.V. Yeremeyev's publications". This thematic collection offers the user to get acquainted with digital images of minerals, which are studied in P.V. Yeremeyev's works, as well as his published works, in which these minerals are described (Fig. 4).

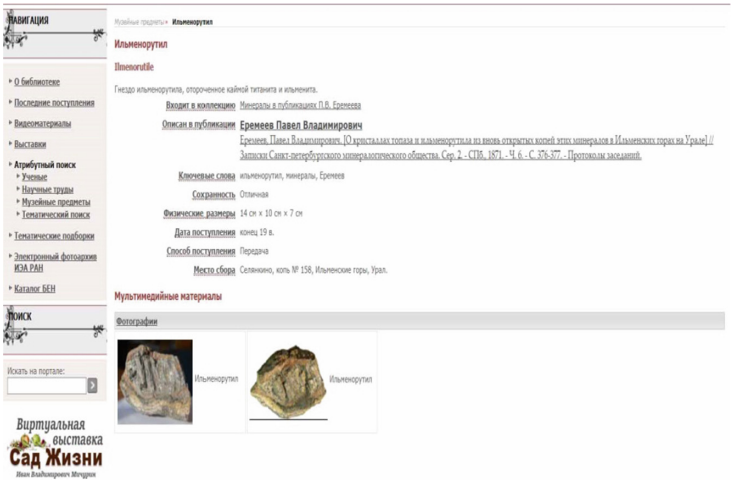

Fig. 4. Thematic collection "Minerals in P.V. Yeremeyev's publications".

The virtual exhibition "Garden of Life", dedicated to the 160th anniversary of the birth of I.V. Michurin (Fig. 5), and the exhibition "Portraits from Skeletons", dedicated to M.M. Gerasimov's scientific work (Fig. 6), are examples of **interdisciplinary** collections created in the form of virtual exhibitions.

The first virtual exhibition includes archive materials dedicated to I.V. Michurin and his scientific school (including text documents, digitized newsreel, and photos) and two theme-specific collections on this topic – digitized publications from library funds and 3D-model fruit of I.V. Michurin, stored in K.A. Timiryazev State Biological

Fig. 5. The main page of the virtual exhibition “Garden of Life”.

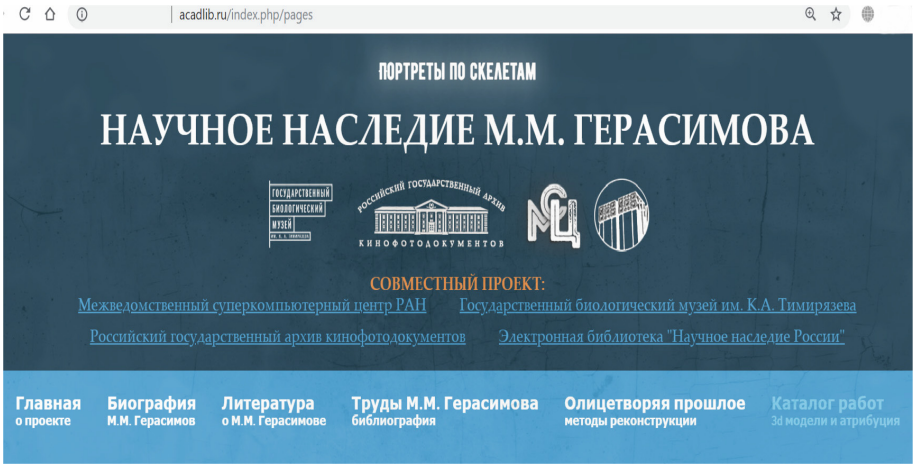

Fig. 6. The main page of the virtual exhibition “Portraits from Skeletons”.

Museum. The exhibition is freely available at http://vim.benran.ru/ and referred to in the portal of SHR DL.

The exhibition “Portraits from Skeletons” dwells on M.M. Gerasimov’s scientific heritage, his school, and anthropology development. This exhibition has the same

structure as the above one and includes, among other things, digital 3D-models of M. M. Gerasimov's anthropological reconstructions and his students, interactive component allowing the visitor to get acquainted with M.M. Gerasimov's work in a playful form (Fig. 7).

Fig. 7. The page of joint virtual exhibition "Portraits from Skeletons" describing M.M. Gerasimov's work.

Both exhibitions created on the platform of the Scientific Heritage of Russia Digital Library in association with K.A. Timiryazev State Biological Museum and Russian State Documentary Film and Photo Archive.

3 Conclusion

In the course of studying the problems of organizing various collections, the following basic principles of their formation have been formulated:

- The IT platform of the digital library should support the ability to form and providing collections of various levels.
- The information environment for forming collections is a set of distributed databases created by information fund holders. This principle presupposes the existence of a single space of information support ensuring the creation, dynamic updating, and accumulation of information resources by organizations participating in the formation of DL funds.

- The technological environment for the formation of digital objects is distributed and unified in terms of the software and hardware used and a set of requirements for digital images being formed. The conceptual basis for the organization of such environments is a distributed formation and storage of large data arrays. At the same time, a centralized metadatabase equipped with common search and navigation services is also created in the distributed information environment. Along with it, metadata are also stored on the servers of participants of the formation of DL funds, which enables their use in local systems of information search and provision [15].
- Provision of complete and sufficient digital object (metadata set) description means for their inclusion in a collection and their representation in information funds of the digital library.
- Independent collection formation by information fund holders combined with their availability.

The developed hierarchical approach to the formation of collections within the digital library makes it possible to more effectively and dynamically meet the information needs of users of different categories.

According to these principles, the dispatch system for stages of integration of interdisciplinary collections has been developed. It allows immersing various digital objects (images of museum objects, printed publications, archive materials, multimedia objects) into the distributed SHR DL environment online. The subsystem of data packet exchange allows exchanging data in the RDF/XML format according to the ontology metadata model. In its turn, the architecture of formation of the information support system for interdisciplinary collections and the application profile of the extended data storage support allow forming such collections in the distributed SHR DL environment.

Thematic collections presented in the SHR DL environment have been created in compliance with the available technology of forming and representing thematic collections in a distributed information environment of the digital library [6, 17, 18].

The research is carried out by JSCC RAS — branch of SRISA within supported by Russian Foundation for Basic Research (projects 17-07-00400 and 18-07-00893).

The computational capacity of JSCC RAS, MVS-100 K cluster, in particular, was used to build 3D-models.

References

1. LNCS Homepage. https://www.europeana.eu/portal/ru. Accessed 13 Nov 2018
2. LNCS Homepage. http://www.theeuropeanlibrary.org/tel4/. Accessed 13 Nov 2018
3. LNCS Homepage. https://dp.la/exhibitions. Accessed 13 Nov 2018
4. Marttila, S., Botero, A.: Infrastructuring for cultural commons. Comput. Support. Coop. Work (CSCW) **26**(1–2), 97–133 (2017)
5. Díaz-Corona, D., Lacasta, J., Nogueras-Iso, J.: Barriers for the access to knowledge models in linked data cultural heritage collections. In: ACM International Conference Proceeding Series Vol. Part F137707. 5th Spanish Conference on Information Retrieval, CERI, Zaragoza (2018)

6. Shahzad, F., Alwosaibi, F.M.: Development of an e-Library Web application. In: Proceedings of the 11th International Multi-Conference on Society, Cybernetics and Informatics, IMSCI Orlando, pp. 153–158 (2017)
7. Anagnostopoulos, I., Zeadally, S., Exposito, E.: Handling big data: directions and future directions. J. Supercomputing **72**(4), 1494–1516 (2016)
8. Kalenov, N., Sobolevskaya, I., Sotnikov, A.: Hierarchical levels of representation of information objects in a digital library environment. Inf. Innov. **13**(2), 25–31 (2018)
9. Kalenov, N., Savin, G., Serebryakov, V., Sotnikov, A.: Principles of building and formation of the Scientific Heritage of Russia Digital Library. Softw. Prod. Syst. **4**(100), 30–40 (2012)
10. Kalenov, N., Sobolevskaya, I., Sotnikov, A.: Digital museum collections and representation of objects of natural history museum storage in the Scientific Heritage of Russia Digital Library. Sci. Tech. Inf. Serv. **1**(10), 33–38 (2016)
11. Antopolsky, A., Atayeva, O., Serebryakov, V.: Environment of integration of data of scientific libraries, archives, and museums "LibMeta". Inf. Resour. Russia **5**(129), 8–12 (2012)
12. Kalenov, N., Sobolevskaya, I., Sotnikov, A.: On the interaction of the Scientific Heritage of Russia Digital Library with natural history museums. Inf. Resour. Russia № 148, 2–6 (2015)
13. Ivanov, V.M., Strelkov, S.V., Kholina, A.A., Avtyushenko, A.L.: Virtual reconstructions in multimedia exhibitions of objects of cultural heritage. Virtual archaeology. Collection/Hermitage, pp. 41–49 (2015). http://www.virtualarchaeology.ru/pdf/281_va_book2015.pdf. Accessed 13 Nov 2018
14. Barutkina, L.P.: Multimedia in a modern museum exhibition. Bulletin of St. Petersburg State University of Culture and Arts. SPbSUCA, pp. 106–108 (2011)
15. Vassileva, S., Kovatcheva, E.: The innovative model for interactivity in Bulgarian museums. In: 10th Annual International Conference of Education, Research and Innovation (ICERI), pp. 5407–5412. CERI Proceedings (2017)
16. Maggio, A., Kuffer, J., Lazzari, M.: Advances and trends in bibliographic research: examples of new technological applications for the cataloging of the georeferenced library heritage. J. Librarianship Inf. Sci. **49**(3), 299–312 (2017)
17. Frandsen, T.F., Tibyampansha, D., Ibrahim, G.R., von Isenburg, M.: Library training to promote electronic resource usage: a case study in information literacy assessment. Inf. Learn. Sci. **118**(11–12), 618–628 (2017)
18. Mi, X.Y., Pollock, B.M.: Metadata schema to facilitate linked data for 3D digital models of cultural heritage collections: a university of South Florida libraries case study. Cataloging Classif. Q. **56**(2–3), 273–286 (2018)