Archival and Museum Information as a Component of the Common Digital Space of Scientific Knowledge

N. Kalenov[a], I. Sobolevskaya[b] and A. Sotnikov[c]
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), 119334, Moscow, Leninsky Av., 32 a, Russia

Keywords: Scientific Archive, Science Museum, Knowledge Space, Scientific Heritage, Network Technologies, Digital Libraries, Digital Information Resources, Metadata.

Abstract: The Common Digital Space of Scientific Knowledge (CDSSK), in its modern interpretation, is a fundamentally new information environment that accumulates knowledge from various fields of science and is the basis for solving a wide range of problems: from artificial intelligence to the science popularization. One of the prototypes of the CDSSK model is the digital library "Scientific Heritage of Russia" (DL SHR), within which methods and means of integrating heterogeneous digital information (including archival and museum information) related to Russian scientific achievements are being developed. For several years, the Archive of the Russian Academy of Sciences (ARAN) and the V. I. Vernadsky State Geological Museum Russian Academy of Sciences (GGM RAS) participated in the development and in the DL SHR development and filling the DL SHR with digital content of the DL SHR. The paper discusses the metadata profiles adopted for displaying archival and museum objects in the CDSSK, Provides examples of search and visualization.

1 INTRODUCTION

One of the most important directions in the development of the information society is the creation of the Common Digital Space of Scientific Knowledge (CDSSK) (Antopol'skij A.B., 2019). The implementation of the CDSSK as an information environment that accumulates knowledge from various fields of science will create a basis for solving the problems of artificial intelligence, education, popularization of science, preservation and dissemination of scientific heritage (Savin G.I., 2020).

To provide a multidimensional search for heterogeneous objects in the CDSSK and navigation through related resources, it is necessary to define a metadata set attributes for objects of each type. If we talk about archives and museums then in each of them there are automated systems that ensure the solution of accounting problems, preservation and analysis of their storage facilities. These automatized systems (AS) digitally contain a significant portion of the information required for the CDSSK. Obviously, it is necessary to use this information when forming the content of the CDSSK. However, the problem is that the metadata sets used in professional speakers, on the one hand, contain information that is not of interest from the point of view of the CDSSK, on the other hand, they are presented in special formats, specific for this type of objects.

In this regard, following tasks arise:

- The task of determining the metadata objects profiles of archival and museum storage. These metadata objects profiles should be included in the CDSSK;
- The task of software development that allows loading the necessary metadata attributes from the existing AS of archives and museums into the structure of the CDSSK.

At the same time, it is necessary to take into account, on the one hand, the formats of data presentation adopted in archival and museum practice /studies, and on the other, the need to integrate digital

[a] https://orcid.org/0000-0001-5269-0988
[b] https://orcid.org/0000-0002-9461-3750
[c] https://orcid.org/0000-0002-0137-1255

Kalenov, N., Sobolevskaya, I. and Sotnikov, A.
Archival and Museum Information as a Component of the Common Digital Space of Scientific Knowledge.
DOI: 10.5220/0010512401440149
In *Proceedings of the 10th International Conference on Data Science, Technology and Applications (DATA 2021)*, pages 144-149
ISBN: 978-989-758-521-0

copies of archival objects into a single system, taking into account the requirements of the Semantic WEB (Starkova YU.S., 2017).

The concepts that make up the CDSSK ontology are conventionally divided into concepts for:

- Descriptions of the subject area content;
- Formation of any subject area thesaurus;
- Descriptions of thematic collections;
- Descriptions of the integrating library content with data sources from Linked Open Data. Semantically significant connections are defined between these groups of concepts. In accordance with this ontology proposed by (Ataeva O.M., 2018; Kalenov N.E., 2020), the CDSSK is a collection of thematic subspaces containing scientific knowledge in certain science fields. The connection of the elements of individual subspaces is carried out at the first ontology level. This level includes, in particular, the description of universal objects classes, such as "persons", "organizations", "events" related to many subspaces, regardless of their thematic affiliation. In general, the metadata profiles of spatially mapped objects should include the following minimum set of attributes:

1. Unique identifier of the object in the CDSSK;
2. The name of the object;
3. Object class (person, organization, event, archival document, chemical element, etc.);
4. Object view inside the class (in accordance with custom tables);
5. Dates with which the object is associated;
6. Links to related objects with an indication of the types of links (in accordance with the typology adopted in the CDSSK);
7. References to key terms and classification indices in the system-wide subject ontology (thesaurus) of the CDSSK;
8. Additional text information that characterizes a specific object.

With respect to the different types of objects this list is specified and supplemented.

2 THE ARCHIVAL INFORMATION

Archival information is an integral part of the scientific heritage. Without it, it is impossible to conduct research in the historical, political and other humanities. Therefore without archival information, it is impossible to create a full-fledged unified space of scientific knowledge.

The ISAD Basic International Standard regulates archival data for Archival Description (https://clck.ru/RPBZ3).

This standard identifies several basic concepts used in archiving, including:

- Unit of description. It is a document or series of documents in any physical form, perceived as a whole;
- Description level. It is the position of the description unit in the structural hierarchy of the archive fund.

There are 4 main levels of description:

Founds
Inventories
Cases
Documents.

The standard defines the elements that are considered essential in the international exchange of descriptive information:

a) identification code;
b) title;
c) creator;
d) dates;
e) the volume of the unit of description;
f) level of description.

There are 20 more elements related to areas such as identification, context, content and structure, access and use, related materials, notes and description control in addition to the core elements (Sarkar M., 2020; Moscicka A., 2020).

When specifying the metadata profiles of archival objects under the guise of an object within the "archival document" class, the appropriate level specified above must be indicated. In the case of the level "document" it needs to specify an additional document type (for example, letter, protocol, photograph, etc.). The dates the archived object is linked can be the dates of transfer of the document to the archive and the dates of its compilation. An archival document can be linked with persons and organizations by links such as "author", "former owner", "addressee", "actor", etc. The obligatory link with an organization or a person should be the "current owner / custodian" . It is also possible to link an archival document with certain events, but the specification of these links is described in additional text information (Ringel S. 2020).

Some of the listed attributes are explicitly present in the ISAD format (Koch I., 2019). Some of them can be generated automatically from a set of ISAD elements (Haki K., 2019). Some attributes (in particular, the classification indices of the subject ontology) must be formed using special algorithms and elements of artificial intelligence.

Today, digital libraries use the Encoded Archival Description (EAD). EAD is a XML-standard code for encoding descriptive information about archival materials, supporting structural presentation and remote access to detailed hierarchical descriptions of archival holdings based on ISAD (G) principles.

EAD is used not only by archives, libraries, but also by museums, national libraries and historical societies.

When specifying the metadata profiles of archival documents in the CDSSK, it is necessary to take into account the developments in the archival "industry", in particular, the international standard for the creation of archival authority records - ISAAR (CPF) (https://goo.su/2iZ7). This standard describes not only the presentation of archival documents themselves, but also their links with other objects.

3 MUSEUM INFORMATION

Objects of natural science museums are as much an integral part of the CDSSK as archival documents. Collections of natural science orientation contain colossal volumes of information about natural objects; both existing and lost (extinct forms of wildlife, spent mineral deposits, etc.).

Natural science museums, in contrast to art or other museums based on the results of man-made creativity, are based on the use of the natural environment. Therefore, the characteristics of the environment, represented by objects of different scales of various origins, are as important information elements as the object itself.

As a detailed review of the world experience in the application of standards in museum activities shows (Muzejnye standarty: mezhdunarodnyj opyt, 2019), there is no unified approach to this problem in the world. In Russia, the most widespread automated system used in hundreds of museums of various types is KAMIS (https://kamis.ru/). This system is focused on solving a wide range of tasks, from upkeeping of museum objects inventory record accounting for museum items, creating catalogs of museums and ending with restoration processes. The system is flexible enough with a customizable list of data elements, which can reach several dozen for one object. Therefore, there is no need to talk about the use of specific field names for existing descriptions of museum objects in the CDSSK.

The specificity of museum objects is manifested in the types of temporal characteristics and the types of their connections with other objects. In particular, for natural science museums a rather important temporal characteristic is the “collection of samples date ”, and the connection with the person is the “collection author ”.

4 IMPLEMENTATION EXAMPLE

The digital library «Scientific Heritage of Russia» (DL SHR) can act as can act as a prototype of the CDSSK (Kalenov N.E., 2012). This project is implemented by a group of scientific organizations of the Russian Academy of Sciences headed by the 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). The DL SHR was created to provide Internet users with multifaceted information about scientists who made a significant contribution to the development of various areas of Russian science in the period from the 18th to the first quarter of the 20th centuries. Biographical information on scientists, museum and archival information (historic/archival documents and museum items) related to them, as well as bibliography and full texts of their main publications were entered (integrated?) into the DL SHR.

During the implementation of the project, research was carried out related to:

- The formation of digital copies of objects of various kinds (books, photographs, archival materials, museum items);
- The definition of the metadata space that unites them;
- The development of the Library architecture;
- Working out technology of its filling and provision to users.

As a result of these studies, the selection of a technical base and technology that meets the quality and safety requirements for objects to be digitized was carried out; special software has been developed for the formation and processing of digital copies, which guarantees the transfer in a digital copy all the nuances of the original objects; a fairly universal ontology of the DL SHR as a whole and its individual components has been developed. During the implementation of the project, in which several dozen organizations participated (including libraries, museums and the RAS Archive), the technology of interaction between the participants in filling the library and monitoring of the technological process of preparing data in a network mode was tested.

Because of those activities the option of the distributed (decentralized) preparation of metadata and digital objects copies, centralized editing and

metadata storage on the serves of the JSCC RAS. A combination of distributed metadata and centralized storage of digital copies of objects on the servers of the JSCC RAS was chosen as the optimal organizational and technological structure that ensures the filling of the DL SHR.

Currently, the DL SHR is available to users at http://e-heritage.1gb.ru.

At the moment, the DL SHR contains information on more than 6400 scientists and more than 25400 digitized books (Fig. 1)

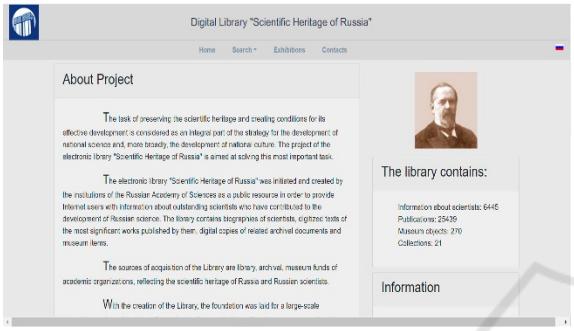

Figure 1: The main page of the DL SHR site.

5 ARCHIVAL INFORMATION IN THE DIGITAL LIBRARY "SCIENTIFIC HERITAGE OF RUSSIA"

The Archives of the Russian Academy of Sciences (ARAN) (Kalenov N.E., 2013) took an active part in provisioning the DL SHR with content. Currently, archival information is available in the DL SHR in the metadata attributes related to scientists and in the form of active links (или просто as links) to the corresponding resources presented on the ARAN website.

The display of archival information about a person in the DL SHR is Information about academician I.I. Schmalhausen is given as an example of archival information displayed in the DL SHR. On the page dedicated to him (Fig. 2), along with general information ("Date of Birth", "Place of Birth", etc.), there are fields:

On the page dedicated to the selected person, along with general information ("Date of Birth", "Place of Birth", etc.), there is a drop-down menu:

1. Curriculum Vitae. This field contains the biographical information about the scientist.

2. Publications. This field represents the scholar's bibliography. Each of the publications is an active link to an electronic copy of the full text of the publication.

3. Archival information. This field contains the drop-down menu of archival documents stored in ARAN. When clicking (expanding the list) on the selected document, the user is redirected to the ARAN page to get acquainted with this archive (Fig. 3).

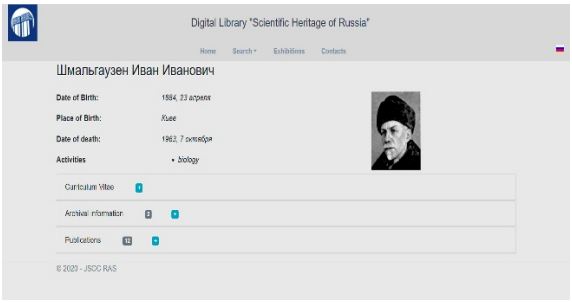

Figure 2: Archival information representation in the metadata about scientist in the DL SHR.

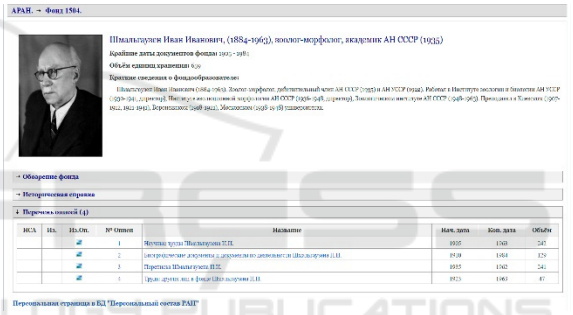

Figure 3: ARAN page containing the archive of I.I. Schmalhausen.

Table 1: The distribution of scientists by scientific fields. Archival information about scientist is presented as links to ARAN resources.

Science Abstracting	Total number of personalities with links to ARAN	Number of personalities with links to ARAN and date of birth before 1900
Mathematics, mechanics	39	21
Physics	53	26
Chemistry	62	47
Biology, medical sciences	75	56
Earth sciences	69	52
Technical science	42	29
Humanitarian sciences	46	31
Social Sciences	104	70

Archival material related to 490 Russian scientists and representing all thematic areas of science is presented in the DL SHR. In the Table 1 is presented the distribution of scientists by thematic areas of science, archival information about which is presented in the form of links to ARAN resources.

The social sciences in this context include history, philosophy, legal and political sciences.

As can be seen from the table, archival materials are fairly evenly distributed among the main scientific fields.

6 MUSEUM INFORMATION IN THE DL SHR

More than 260 digital images of objects of natural science museums are presented in the DL SHR. Of these, about 100 are digital 3D models (Kalenov N.E., Kirillov S.A., Sobolevskaya I.N., Sotnikov A.N. 2020; Sobolevskaya I. N., 2019).

The V. I. Vernadsky State Geological Museum Russian Academy of Sciences (GGM RAS) bears part in filling with content the DL SHR.

The set of metadata fields for representing a museum object was specified in relation to geological artefacts.

Figure 4 shows an example of the representation of a geological museum object in the DL SHR.

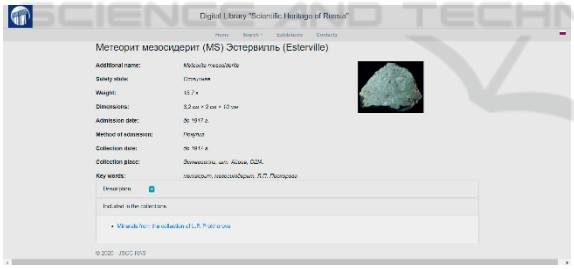

Figure 4: An example of a geological museum object representation in DL SHR.

The inclusion of museum objects in the CDSSK offers a way of using complex links, including indirect links of these objects with…, in exhibitions and scientific work.

Figures 5 and 6 show an example of the links between a geological object and a scientist who has studied it. Figure 5 shows the shell core of Perisphinctes stschurowskii. The "Collection Author" (CA) is an active drop-down manu on A.O. Mikhalskiy. He is a scientist who was engaged in the sciences of the Earth and, in particular, ammonites. Figure 6 shows the page dedicated to information about A.O. Mikhalsky to which the user gets using (CA). On the same DL SHR page a user can get acquainted with biographical information about the scientist, his works, as well as with museum objects related to him (to the scientist).

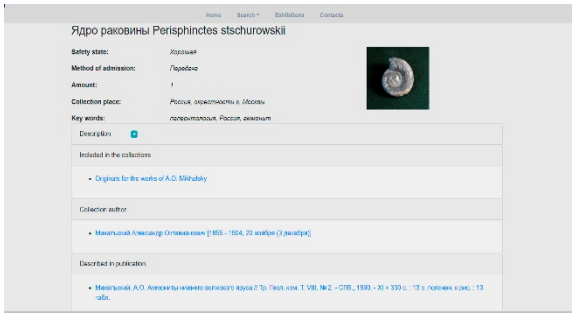

Figure 5: Shell core Perisphinctes stschurowskii.

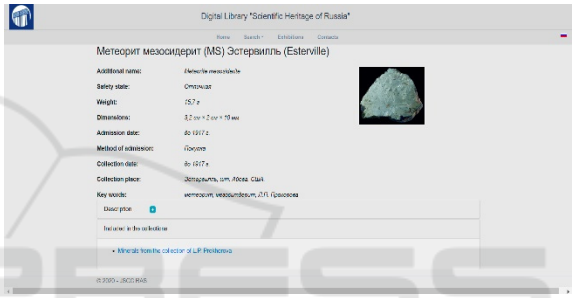

Figure 6: Page dedicated to information about A.O. Mikhalsky.

7 CONCLUSION

The experience of operating the DL SHR has confirmed the effectiveness of organizational, technological and software solutions, which form its basis, and the expediency of transforming the DL SHR into a Single Digital Space of Knowledge.

The importance and relevance of work on including the resources of archives and museums in the CDSSK (at the first stage - in the DL SHR as a prototype of the CDSSK model) is beyond doubt.

The research is carried out by 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" within the Russian Foundation for Basic Research RFBR (project 20-07-00773). The research used the MVS-10P supercomputer installed at the MSC RAS.

REFERENCES

Antopol'skij A.B., Kalenov N.E., Serebryakov V.A., Sotnikov A.N. 2019. O edinom cifrovom prostranstve nauchnyh znanij. In *Vestnik Rossijskoj akademii*, Vol. 89 (7). pages 728-735.

Ataeva O.M., Serebryakov V.A. 2018. Ontologiya cifrovoj semanticheskoj biblioteki LIBMETA. In Informatika i ee primeneniya, Vol. 12 (1). pages 2-10.

Haki K., Blaschke M., Aier S., Winter R. 2019. A Value Co-creation Perspective on Information Systems Analysis and Design. In *Business & information systems engineering*. Vol. 61(4). pages 487-502.

http://e-heritage.1gb.ru/Catalog/IndexL (last access 24.01.2021).

https://clck.ru/RPBZ3 (last access 24.01.2021).

https://goo.su/2iZ7 (last access 24.01.2021).

https://kamis.ru/ (last access 24.01.2021).

Kalenov N.E., Kirillov S.A., Sobolevskaya I.N., Sotnikov A.N. 2020. Vizualizaciya cifrovyh 3d- ob"ektov pri formirovanii virtual'nyh vystavok. In *Elektronnye biblioteki*. Vol. 23 (4). pages 418-432.

Kalenov N.E., Savin G.I., Serebryakov V.A., Sotnikov A.N. 2012. Principy postroeniya i formirovaniya elektronnoj biblioteki "Nauchnoe nasledie Rossii". In *Programmnye produkty i sistemy*, 2012. Vol. 4 (100). pages 30-40.

Kalenov N.E., Serebryakov V.A. 2020. Ob ontologii Edinogo cifrovogo prostranstva nauchnyh znanij. In *Informacionnye resursy Rossii*, № 5. pages 10-12.

Koch I., Freitas N., Ribeiro C., Lopes CT., da Silva JR. 2019. Knowledge Graph Implementation of Archival Descriptions Through CIDOC-CRM. In *Digital libraries for open knowledge, tpdl* 2019. Vol. 11799. pages 99-106.

Moscicka A., Zwirowicz-Rutkowska A. 2020. Description of old maps in the Europeana Data Model. In *Journal of cultural heritage*. Vol. 45. pages 315-326.

Muzejnye standarty: mezhdunarodnyj opyt. 2019. pod red. I.A.Grin'ko; M.B. Gnedovskogo. In *Perspektiva*. pages 97.

Ringel S. 2020. Interfacing with the past: Archival digitization and the construction of digital depository. In *Convergence-the international journal of research into new media technologies*. № 1354856520972997.

Sarkar M., Biswas,S. 2020. Exploring Archives Space: An Open Source Solution for Digital Archiving. In *Desidoc journal of library & information technology*. Vol 40(5). psges 272-276.

Savin G.I. 2020. Edinoe cifrovoe prostranstvo nauchnyh znanij: celi i zadachi. In *Informacionnye resursy Rossii*. № 5. pages 3-5.

Sobolevskaya I. N., Sotnikov A. N. 2019. Principles of 3D Web-collections Visualization. In *Proceedings of the 3rd International Conference on Computer-Human Interaction Research and Applications*. pages 145-151.

Starkova YU.S., Cyrul'nikova E.S., Kulieva N.V., Samojlov A.N. 2017. Poiskovye sistemy Semantic WEB. In *Informatizaciya i svyaz'*. № 3. pages. 70-74.