

Development of Digital Resources for the Digital Library “Scientific Heritage of Russia”

Anna Lopukhina¹[0000-0002-4952-5671], Anastasiya Rybakova²[0000-0002-8061-8083],
Sergey Kirillov³[0000-0001-7560-0041]

¹⁻²⁻³ Joint Supercomputer Center of RAS – Branch of Federal State Institution “Scientific
Research Institute for System Analysis” of Russian Academy of Sciences, Leninsky pr., 32a,
119334, Moscow, Russia

¹ vespa1477@gmail.com, ² rybakova@jssc.ru, ³ skirillov@jssc.ru

Abstract. The paper describes methods of formation and presentation of digital objects of various types (printed publications, archival materials, museum objects, audio and video materials) for the Digital Library “Scientific Heritage of Russia”. A digital publishing hardware and software system has been created within the framework of the digital library project. The digital publishing hardware and software system consists of five subsystems: dispatching subsystem, content description subsystem, digitization subsystem, generating digital publications subsystem and content control and storage subsystem. Dispatching subsystem organizes the work on content creation. Content description subsystem describes objects of the digital library. Digitization subsystem produces digital copies of library items, archival and museum objects. Digital publications subsystem creates the final products: digital books, 3D models, etc. Content control subsystem controls the quality of the scanned books and metadata.

Keywords¹: databases, research results accounting, information support of science, scientific information, content Web-based system.

¹ CDSSK–2020: International Conference “Common Digital Space of Scientific Knowledge”,
November 10–12, 2020, Moscow, Russia

EMAIL: vespa1477@gmail.com (Anna Lopukhina); rybakova@jssc.ru (Anastasiya Rybakova);
skirillov@jssc.ru (Sergey Kirillov)

ORCID: 0000-0002-4952-5671 (Anna Lopukhina);

0000-0002-8061-8083 (Anastasiya Rybakova); 0000-0001-7560-0041 (Sergey Kirillov)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons
License Attribution 4.0 International (CC BY 4.0)



CEUR Workshop Proceedings (CEUR-WS.org)

1 Introduction

The Digital Library “Scientific Heritage of Russia” is a model digital library solution which allows to integrate in addition to digital versions of printed materials other types of information, in particular: archival materials, museum objects presented in the form of: graphics, photos, 3D models of museum objects, audio and video materials.

The Digital Library “Scientific Heritage of Russia” is a distributed information system uniting several remote participants working according to the unified rules.

The core components of the digital library are model digital content production centers, placed with information fund holders (in archives, libraries, museums and research institutes) in different regions of Russia.

The main task of such centers is to provide the formation of information resources of the following kinds:

- digital copies of printed and handwritten documents from libraries, archives and museums;
- digital models of museum items;
- digital copies of visual and audio-video materials.

The developed architecture of the typical center for digital content formation implies the following technological tasks:

- content selection;
- content description (metadata formation);
- digitization;
- processing of digitized content;
- layout of a digital publication (media-object);
- publication of the digital publication on the Internet portal.

2 Digital content development technology

A model solution for construction of the digital library was developed within the framework of the project for creation of the Digital Library “Scientific Heritage of Russia”.

The collaborators of the JSCC RAS, the CC RAS and the LNS RAS developed special applied software [1–8], based on which the technological hardware and software complex for the preparation of digital publications, consisting of the following subsystems, was implemented.

2.1 Dispatching subsystem

This subsystem is designed to organize the work on creation of metadata and digital publications to form the content for the Digital Library “Scientific Heritage of Russia”. Using a set of values (e.g., “proposed for digitization”, “registered”, “in progress”, “digitized”, “delivered”), the dispatch subsystem records the current state of work at each stage, which makes it possible to monitor the work progress. This makes it possible to obtain different reports about the information being prepared for the digital library (both about scientists and about publications), reports about the current state of processing of

a particular publication, quantitative data about the work done by one or another participant, etc. The subsystem ensures the elimination of the possibility of duplication in the production of digital publications, planning of their preparation, and a number of other service functions related to the organization of work on the formation of digital objects.

2.2 Content description subsystem

This subsystem includes a special software shell and a database into which the metadata describing the information objects of the Digital Library “Scientific Heritage of Russia” are entered. The subsystem is located on a special technological server; access to its contents is available only to the authorized users who have the rights to enter and/or correct the data.

Metadata creation and input begins with entering author information. The information about the scientist is entered into the appropriate fields of the biographical information database: full name, possible pseudonyms, variants of name and surname spelling, occupation, years of life, etc. The author's biography, bibliography of works, as well as other necessary information is compiled.

Work is being done to find external sources of information on the author. In the database of biographical information, links to the websites of museums, various digital pages on the Internet, which contain meaningful information about the life and work of the scientist, are entered.

Then, a set of metadata describing the digital resources is formed. For library items and archival storage objects, classification and service information such as title of publication in the original language, translation of the title into Russian, type of publication, year of publication, place of publication, publishing house, series titles and volume numbers in the series, indications of possible differences from previous editions both in the content and in the title of the book, indications of illustrations and maps, etc. are entered into appropriate fields of the bibliographic information database.

For museum storage objects, classification and service information is entered into the appropriate fields of the museum information database, such as collection, described in the publication, keywords, preservation, physical size, age, date of receipt, method of receipt, author of collection, place of collection.

2.3 Digitization subsystem

The subsystem is intended for making digital copies of museum storage objects, library and archive items, which have been selected as content for the Digital Library “Scientific Heritage of Russia”. To create the digital library information (e.g., book pages) is converted into digital form using different types of scanners with different resolution, possibility of scanning different types of images and other properties. In this connection, the following basic types of digitization complexes have been formed in the project:

Black-and-white scanning complex based on the MINOLTA PS7000 planetary scanner. The Minolta PS7000 scanner digitizes A5-A3 books with a resolution of 600 dpi;

Scanning complex based on the PlanScan C2-CA-600 color planetary scanner. This scanning complex replaces the outdated Minolta PS7000 based complex. The PlanScan S2-CA-600 scanner operator can scan any type of publications up to A2 size (in color, gray and black and white mode) at a resolution of 600 dpi;

Planetary color scanning complex PlanScan A2-VC-B. The Plan Scan A2-VC-B is similar to Plan Scan C2-CA, but is more compact and equipped with a built-in LED illumination system. The main feature of this complex is the ability to scan old and tightly bound publications;

Scanning complex based on the Powerscan D14000 A0-20/25 scanner. The main feature of this complex is the ability to scan materials of particularly large format (up to the size of A0), with high resolution, such as newspapers, maps, posters, pictures;

Scanning complex based on Kodak ABR 3000 DSV microfiche scanner. This complex is used to digitize materials stored on microfiche and microfilm;

3D digitizing complex based on Recam T-50 rotary platform, 3D-Maker control program and Canon EOS600D digital camera. The complex allows you to perform in automatic mode 3D-shooting of museum objects up to 150 cm high and weighing up to 50 kg. To create a digital 3D-model of a museum object a round sectional survey is taken. As a result, a sequential set of images of a museum object from all its sides is obtained.

The use of the specified digitization complexes allows to perform the most accurate digitization of publications.

2.4 Generating digital publications subsystem

This subsystem consists of a set of software products designed to create the end product – digital copies of library, archive and museum objects. Several service programs are used within the subsystem.

The NNMetaDis program is designed for the formation of digital books. It allows assembling books from digitized pages, inputting bibliographic descriptions of editions and their table of contents.

Autodesk 3dsMax, 3DMaker, and a set of in-house developed Java scripts are used to create 3D models of items in museum storage.

Photos, portraits, maps, paintings, etc. are presented using a set of proprietary macro commands in Adobe PhotoShop.

To represent audio and video resources a set of editors, codecs and converters of free and open-source software is used.

2.5 Content control subsystem

The subsystem includes methods of control operations: checking the quality of book scanning, availability of pages, order of their placement, correctness, and completeness of metadata, etc. Content control subsystem controls the quality of the scanned books and metadata.

3 Web-portal

Functioning of the Digital Library “Scientific Heritage of Russia” as a public search system is ensured by the server of the web-portal. User interface of the digital library provides information search, visualization and navigation through the related objects of the Digital Library. A number of services, providing integration with the information resources of other organizations, is envisaged within the Digital Library.

4 Conclusion

The methods of generating and presenting digital images of various types of objects (printed editions, archival materials, museum holdings, audio and video materials) elaborated within the framework of the Digital Library “Scientific Heritage of Russia” project contribute to the development of the digital information and library environment: they help to improve digital description of the documents, digitization techniques and search capabilities of the system. The created fund of digitized documents and the reference and bibliographic apparatus make a significant contribution to the formation of the Common Digital Space of Scientific Knowledge.

The work was carried out at the Joint Supercomputer Center of Russian Academy of Sciences – Branch of Scientific Research Institute for System Analysis of Russian Academy of Sciences, within the framework of state assignment No. 0580-2021-0014.

References

1. Kalenov, N.E., Savin, G.I., Serebryakov, V.A., Sotnikov, A.N.: Printsipy postroeniya i formirovaniya elektronnoy biblioteki “Nauchnoe nasledie Rossii”. Programmnye Produkty i Sistemy (100). S. 30–40 (2012).
2. Pogorelko, K.P.: Evolyutsiya programmnogo obespecheniya sistemy podgotovki materialov dlya elektronnoy biblioteki “Nauchnoe nasledie Rossii”. In: Informatsionnoe obespechenie nauki: novye tekhnologii: sbornik nauchnykh trudov. S. 260–263. Moskva (2011).
3. Yakshin, M.M.: Platforma SciRus – osnova tekhnologicheskogo kompleksa elektronnoy biblioteki “Nauchnoye nasledie Rossii”. In: Elektronnyye biblioteki: perspektivnye metody i tekhnologii, elektronnyye kolleksii: XVI Vserossiyskaya nauchnaya konferentsiya RCDL 2014. Trudy konferentsii. S. 362–368. Moskva (2014).
4. Yakshin, M.M.: Razvitie platformy SciRus. In: Informatsionnoe obespechenie nauki: novye tekhnologii: Sbornik nauchnykh trudov. S. 203–207. Moskva (2015).
5. Pogorelko, K.P.: Sistema eksporta dannykh elektronnoy biblioteki “Nauchnoe nasledie Rossii”. In: Informatsionnoe obespechenie nauki: novye tekhnologii: Sbornik nauchnykh trudov. S. 200–202. Moskva (2015).
6. Sotnikov, A.N., Kirillov, S.A.: Tekhnologiya podgotovki elektronnykh izdaniy dlya elektronnoy biblioteki “Nauchnoe nasledie Rossii”. In: Informatsionnoe obespechenie nauki: novye tekhnologii: Sbornik nauchnykh trudov. Moskva. S. 178–190 (2015).

7. Zablovskaya, I.E., Kirillov, S.A., Kondrat'eva, E.A., Pruglo, O.A., Sotnikov, A.N.: Voprosy formirovaniya fondov elektronnoy biblioteki "Nauchnoe nasledie Rossii". In: Informatsionnoe obespechenie nauki: novye tekhnologii: Sbornik nauchnykh trudov. Moskva. S. 184–191. (2017).
8. Kalenov, N.E., Kirillov, S.A., Sobolevskaya, I.N., Sotnikov, A.N.: Sovremennoe sostoyanie elektronnoy biblioteki "Nauchnoe nasledie Rossii". Trudy NIISI RAN. Matematicheskoe i komp'yuternoe modelirovanie slozhnykh sistem: teoreticheskie i prikladnye aspekty (8). S. 166–169 (2018).