Open Archives of the SB RAS:
Systems of Historical Factography

Irina Krayneva[1][0000-0002-0601-9795] and Alexander Marchuk[2][0000-0001-8455-725X]

[1.] A.P. Ershov Institute of Informatics Systems, Russian Ac. of Sci., Siberian Branch
6, Lavrentjev pr., 630090, Novosibirsk, Russian Federation;
National Research Tomsk State University
36, Lenin pr., 634050, Tomsk, Russian Federation;
[2] A.P. Ershov Institute of Informatics Systems, Russian Ac. of Sci., Siberian Branch
6, Lavrentjev pr., 630090, Novosibirsk, Russian Federation
Krayneva55@gmail.com

Abstract. Interdisciplinary cooperation between humanities and IT specialists, open scientific communications, high-quality information are the main goals of our academic service projects. This paper presents a brief summary of the twenty years of research carried out at the A.P. Ershov Institute of Informatics Systems SB RAS in the area of developing electronic archives for heterogeneous documents. The phenomenon of electronic archives emerged and has been developing as part of the Novosibirsk school of informatics, which has always been oriented towards the contracting of social services. Over the years, the IIS SB RAS has completed a range of projects on digitizing historical and cultural heritage of the Siberian Branch of the Russian Academy of Sciences. The team created a number of information systems for the support of electronic archives on the history of science in Siberia: the Academician Andrei Ershov Electronic Archive, SB RAS Photoarchive, SB RAS Open Archive, a collection of digitized vintage and old textbooks on mathematics, etc. However, staff cuts in the SB RAS Presidium undercut the ongoing contributions to the SB RAS Scientific Archive. By no means can our projects substitute the function of that archive. We aim at complementing it to preserve valuable historical facts that are often overlooked by the government and institutional archives.

Keywords: high-quality information, open archives, digital archives, interdisciplinarity, history of science, Siberian Branch of RAS.

1 Introduction

The wide use of information technologies in the Russian humanities came along with the mass appearance of personal computers and the Internet in the mid-1990s. The idea of interdisciplinary cooperation of humanities and IT specialists at the dawn of the Internet was based chiefly on the concept of open scientific communication. Es-

sentially this is a cluster of civil society supported by professionals, where necessity and responsibility go side by side. An American computer scientist, recipient of the Turing Award, James Nicolas "Jim" Gray (1940–?) proposed the concept of the fourth paradigm of scientific research with massive amounts of data using grid technology – an archive of science [1]. Gray and his followers stress the need for systematic arrangement and free access to scientific entities, such as experimental data or modeling results in physics. The idea of a virtual scientific archive for humanities research is indisputably just as relevant. In view of the empirically proven positive influence of information and communication technologies (ICT) on the activities of scientific workers, open access to information becomes a priority [2].

Equipping museums, libraries, universities and research institutions with computers and providing access to the Internet resulted in a wider range of user practices in humanities and in the emergence of Internet-oriented resources. They are used to publish museum collections, library catalogues, full texts, research and reference tools for archives as well as complete individual collections. This initiative became deeply rooted in the activities of cultural and academic institutions in Russia. In the Novosibirsk Scientific Center, the implementation of a number of projects on electronic archives dealing with various documents became possible with the deployment of the Novosibirsk Scientific Center Internet Network (1994–1998, Soros Foundation, RFBR, INTAS – The International Association for the Promotion of Cooperation with Scientists from the New Independent States of the Former Soviet Union). It provided organizations and research institutions of the SB RAS with free access to the Internet [3].

The IIS SB RAS team in Novosibirsk has been working on projects on open archives since 1999. The projects are based on the digital historical factography technology. The technology implies publishing historical sources in Internet-oriented information systems according to the rules of publishing archival documents. This includes indicating the origin and source of the documents as well as a number of typological features such as the document type, author, addressee (a person or an organization), date, geographical data, etc. Specialized information systems developed in the IIS SB RAS offer tools for establishing connections between these subject entities. Quoting documents from an electronic archive is supported both as an Internet link and by indicating a specific file and sheet in the archive. These systems are viewed as a viable alternative to the existing brick-and-mortar archives and require a state-supported program on their implementation and support. Apart from concentrating, systemizing, dating and describing sources, they serve as the foundation of a range of research project.

2   Academician A.P. Ershov's Digital Archive

In 1999, the IIS SB RAS team began their work on an automated information system for the creation and support of a digital archive of documents – the Academician A.P. Ershov's (1931–1988) Digital Archive http://ershov.iis.nsk.su/. From 1957 through 1988, Andrei Ershov was the head of the Programming department – first in

the Computing Center of the Institute of Mathematics, from 1964on – in the Computing Center SB AS USSR. A modest but challenging position, an outstanding academic career (1962 – Candidate of Sciences, 1967 – Doctor of Physics and Mathematics, 1971 – Corresponding Member of the Academy of Sciences, 1984 – Academician), unconventional research projects and personal charisma – all this contributed to Ershov becoming the universally acknowledged leader of Soviet programming, a respected member of the international computer science community, and founder of informatics in Siberia. He devoted a lot of time to his personal archive, which is now a valuable source of information on the history of programming in the USSR. The archive covers a period from 1949, when the future Academician was still a school student, to 2015.


Fig. 1. Andrei Petrovich Ershov in his office in Computing Center of the SB RAS. A year before the elections to the USSR Academy of Sciences. Novosibirsk, 1969.

The creation of the A.P. Ershov's Archive was supervised by Alexander Marchuk (Doct. Ph.-M. Sci.) and a scientific researcher from his laboratory, Vladimir Filippov. The system was developed by two postgraduates of the Mechanico-Mathematical Department of Novosibirsk State University: Andrey Nemov, Konstsntin Fedorov and a Bachelor student, Sergey Antyufeev. The concept of the system was developed by Mikhail Bulyonkov (Cand. Ph.-M. Sci.), a computer scientist, Natalya Cheremnykh, a mathematical linguist, and Irina Krayneva, a historian. Irina Pavlovskaya, Svetlana Zhukovskaya, Alexander Rar (1928–2011), Liudmila Zmievskaya, Natalia Polyudova, and Anna Bulyonkova all contributed a lot to document description and management of the archive. The development of the digital version of the Archive was sup-

ported by a number of Russian and foreign IT companies, (Microsoft Research, xTech, ATAPY Software, UNIPRO) [4].

In the process of creating the first academic project of the Internet-oriented information system "Academician A.P. Ershov's Digital archive," the team solved the problem of developing original client-server program tools, using predominantly Microsoft instruments and technologies. The archivist's work space was implemented in Perl. Using digitized documents is beneficial not only in terms of communicative convenience, but also ergonomically, since many specialists working with archival sources suffer from a number of ailments caused by prolonged exposure to old paper and dust.
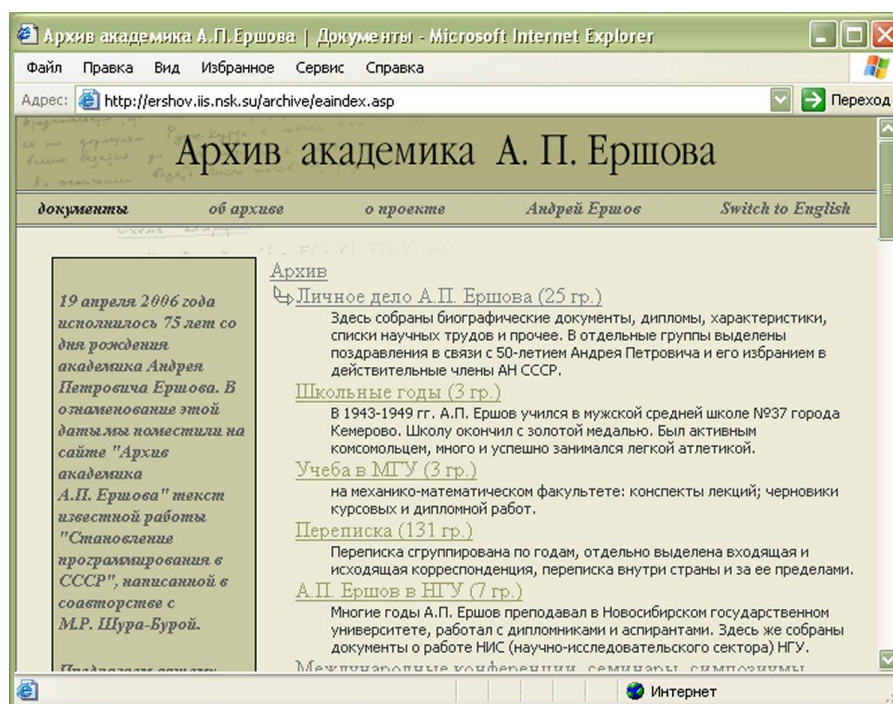
Fig. 2. The list of subject and chronological groups of the Academician A.P. Ershov Digital Archive.

The developers assumed that the public interface of the archive had to be visualized in the same way as it was intended by the creator, i.e. it had to correspond to the physical body of documents from the A.P. Ershov's archive. Ershov formed his folders by the subject-date or subject, and that principle remained unchanged. Folders and sheets were numbered and scanned; corrections were minor and dealt with removing duplicates, chronological arrangement, and recovering authorship and dates. The physical archive formed by Ershov was complemented with a number of documents from government-run archives that came up in the process of studying the scientist's

biography. The digital version supports two types of systematization: folder-based and subject-and-date based, presented as a corresponding catalog.

Apart from documents from Ershov's archive proper, the Digital Archive contains materials on the history of the IIS SB RAS, Start Temporary Research and Technology Team (VNTK «Start», 1985–1988), and the International Andrei Ershov Memorial Conference on the Perspectives of System Informatics (PSI), which has been held and hosted by the IIS SB RAS since 1990. These collections are thematically connected with the main body of documents from Ershov's Archive. There is also a satellite collection of documents of Svyatoslav Sergeevich Lavrov (1923–2004, Saint-Petersburg), a Corresponding Member of the USSR AS, provided to the IIS SB RAS by his followers for the purpose of publishing the collection in the Internet. As of March, 2019, there source contained 44.4 thousand documents.

Work on Academician A.P. Ershov's Digital Archive turned out to be a rather science-intensive project. Apart from a series of research papers published based on the study of the documents from this collection, the members of the team published monographs and defended four theses on the history of science (two by Irina Krayneva in Tomsk, one by Ksenia Tatarchenko in Princeton and one by Margarita Boenig-Liptsin in Harvard) [5, 6]. The study of the archive continues, and its hermeneutical potential is still far from exhaustion.

3   Photographic Archive of the SB RAS

Shortly before the 50th Anniversary of the Siberian Branch of the Russian Academy of Sciences (formerly the Academy of Sciences of the USSR), which was founded in 1957, an initiative group from the IIS SB RAS headed by Alexander Marchuk began to work on a project called "Digital Photographic Archive of the SB RAS" http://www.soran1957.ru (2005–2009). The resource merged various collections and individual photographs dealing with the history of science in Siberia into a single body of documents; the photographs were supplied by photographers, reporters, organizations, such as the Museum, Exhibition Center, Press Center, and various research institutes of the SB RAS, as well as by private collectors.

The creation of Novosibirsk Akademgorodok, a Town of Science, became a landmark event in the history of Novosibirsk in the 20th century. Eventually, scientific centers appeared in other Siberian towns, including Krasnoyarsk, Kemerovo, Omsk. The photographic archive of Akademgorodok was started by its founder, Academician Mikhail Lavrentiev. He invited a photographer, Rashid Akhmerov (1926–2017), who at the time had been photographing the daily life of the institutes of the West Siberian Branch of the USSRAS. Later, other professional photographers joined the initiative, along with many amateurs. Nowadays, new contributions to the Archive come primarily from personal collections. Recently, the Archive received the photographic collection of the Quant («Kvant ») Club of the NSU Physics Department. The effort was beneficial to the historical knowledge of Akademgorodok not only because of the creation of the collection as such, but also because open access publication led to correct dating and description of many of the photos. result of the conducted experiments the obtained fields of the model characteristics have been analyzed and plot-

ted. Several fields such as sea level field, temperature field and ice coverage in the Polar Ocean are presented below.
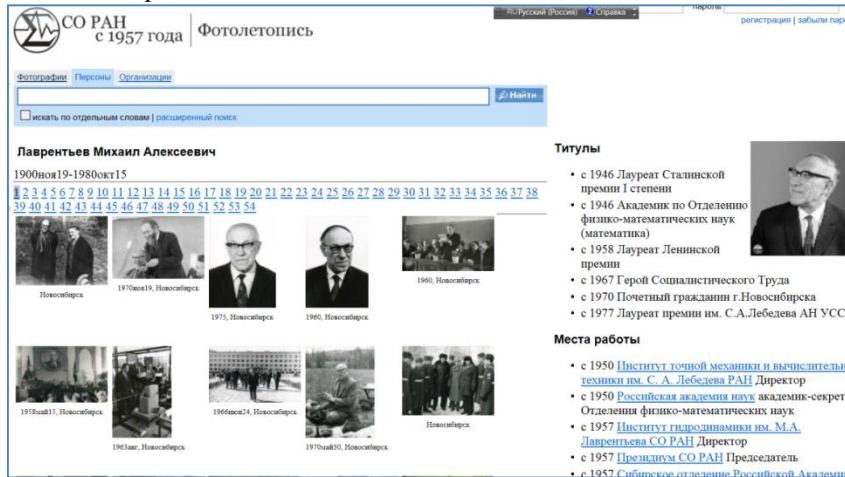
Fig.3. The collection of photos of Academician Mikhail A. Lavrentyev in the SB RAS Photographic Archive

Fig. 4. Nikita Sergeevich Khrushchev, the First Secretary of the CPSU Central Committee, gets acquainted with the development plan of the Academgorodok. He sharply criticized it, after which high-rise buildings disappeared from the project. Photo by Rashid Akhmerov. 1959, October.

A specialized information system – SORAN 1957 – was developed for the SB RAS Photographic Archive. It is a structure designed for collecting, structuring and digital publishing of historical data and documents, which supports program and organizational mechanisms key to the achievement of these purposes. SORAN 1957 includes a system of structured data that represent entities of the real world and relationships between then. The structuring system is based on the Semantic Web ideology. This approach consists in structuring data according to anontology. Ontology is a formal specification of a shared conceptual model – an abstract model of a subject area describing a system of concepts of the subject area. The shared model is a conventional understanding of the conceptual model by a specific community (a group of people). "Specification" here presumes describing the conceptual system explicitly, and "formal" presumes that the conceptual model is machine-readable. An ontology consists of classes of the subject area, properties of these classes, and connections between them. To solve a broad range of information problems, the IIS SB RAS built a basic ontology. The created software tools enable input and editing of data and using data from other sources (newspapers in particular).

Currently, the Photographic Archive database contains information on approximately 7,000 persons, 2,000 organizations and events, along with 24,000 scans of photographic documents. Before their submission to the database, photographic scans are repaired, both automatically and manually, using graphic software tools including color, brightness and contrast correction, noise, dust and damage removal, etc., in a way that does not affect the contents of the document. Documents are scanned in a resolution sufficient for consequent reprinting, from 300 to 1200 dpi, as uncompressed tif files in RGB color model. Documents in jpg format created using modern digital devices are also included in the archive.

The SB RAS Digital Photographic Archive platform also hosts the archive of the weekly newspaper of the Siberian Branch, "Nauka v Sibiri" (*Science in Siberia*), which was named "Za Nauku v Sibiri" (*For Science in Siberia*) until 1983; the newspaper has been published since 1961 and has its own website with an archive page (http://www.nsc.ru/HBC/). The newspaper archive is thematically linked with the photographic archive and was systematized based on the entities existing in the photographic archive (persons, organizations, events, etc.). To retain the quality and minimize the volume of transmitted and processed information on the client's side, Deep Zoom technology was used – a solution for Web-publishing high-resolution images from Microsoft. Silverlight, a browser technology allowing to view an image in general and zoom into its specific part, linking it to some existing entities, was also used. Unfortunately, Microsoft Research terminated the support of this tool, and we are currently searching for an alternative solution.

4   Open Archive of the SB RAS

The experience gained in the projects described above has allowed us to expand the subject coverage range of historical sources. Beginning from 2012, the "SB RAS Open Archive as a system of presentation, accumulation and systematization of scientific heritage" project has been implemented, with financial support in 2012-2014 (http://odasib.ru/). Apart from the IIS SB RAS, a number of humanities institutes of

the SB RAS participated in this project: The Institute of History, Institute of Archeology and Ethnography, Russian National Public Library of Science and Technology; Institute of Mongolian Studies, Buddhology and Tibetology, and museum departments of these institutes. Each of the participants presented their own specific collection accumulated in the course of their professional activity. Currently, the SB RAS Open Archive contains 24 individual collections, with approximately 90,000 document scans as of November 1, 2020.
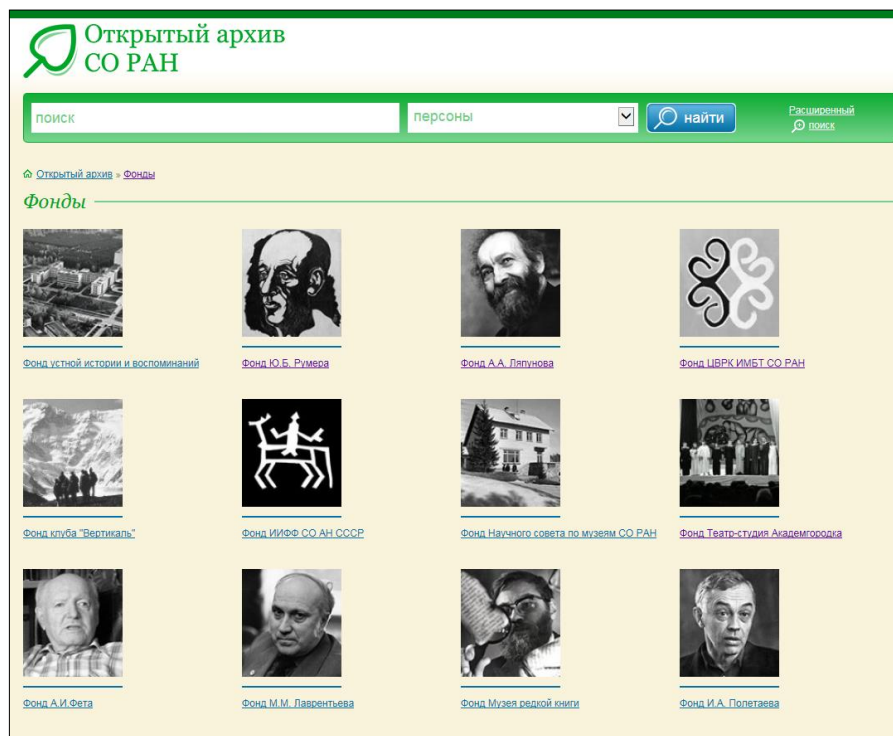
Fig. 4. The main page of the SB RAS Open Archive.

The creation of Internet-oriented information systems broadens the coverage range, making low-demand archives accessible to the general public, and provides access to collections which fall outside the current range of scientific interests of the SB RAS Scientific Archive and other state- or institution-run archives. Contributions to the Open Archive come from private collections on conditions negotiated with the collection owners.

The collection of the Open Archive includes personal collections of the mathematician Abram Fet, engineer Igor Poletaev, sociologist Tatiana. Zaslavskaya, her sister philologist Maya Cheremisina, mathematician Aleksey Lyapunov, theoretical physicist Yuri Rumer and others. Another group of collections is formed by the archives of scientific and educational organizations: the ethnographic collections of the Institute

of Archeology and Ethnography, Russian National Public Library of Science and Technology, Institute of Mongolian Studies, Buddhology and Tibetology, documents on the history of the Institute of Semiconductor Physics, Physics and Mathematics School, and Higher College of Informatics. The third group of collections is archives from social and creative organizations of Novosibirsk Akademgorodok: Vertical rock climbing club, Pod Integralom (*Under the Integral*) café club, Akademgorodok theater-studio, Open Society Institute, etc. The collections of spoken history and memories of people professionally and personally bound with Akademgorodok form yet another separate group. The Open Archive is continuously expanded, as new collections are added.

Открытый архив » Фонды » Фонд А.А. Ляпунова » Коллекции фонда А.А. Ляпунова » Переписка А.А. Ляпунова » Переписка военных лет » Письма А.А. Ляпунова 1943 г. » Письмо

Письмо

Дата: 1943-12-11

Описание документа: А.А. Ляпунов получил гвардейский значок. С юмором рассказывает о своих неудачах с оружием. Описывает свои будни. Говорит о хорошем самочувствии, о большом изменении во внешности.

Ly 255_085   Ly 255_086   Ly 255_087   Ly 255_088

Текст документа:

11/XII 43г.

Дорогая Таточка!

Теперь я обеспечен хорошей почтовой бумагой — получил праздничный подарок. Вообще, я получил огромное количество вещей, многие из которых были пределом моих мечтаний. У меня есть бинокль, секундомер, несколько готовален, правда, немного разрозненных, и множество специальных приборов, служащих для управления огнем, и для топографических работ. В заключение списка, у меня есть наган. Полчаса тому назад я его чистил. Это было развлечение для всего нашего штаба, который вечером преобразуется в некое подобие офицерского клуба. Как только Кацуба увидел, как я взялся за дело, он отобрал у меня револьвер, разрядил его и сказал: "Ну вот, теперь его можно дать Вам в руки". После этого

Fig. 6. An example of document description and visualization (A. A. Lyapunov's letters from the front, 1943) in the SB RAS Open Archive. Lyapunov says that he got a guards badge - a symbol of a soldier's courage. He also writes that he has paper for notes, binoculars, a stopwatch, cases with drawing instruments, many special devices for controlling shooting and for topographic work, and revolver! He was happy!

5   Analogies and Problems

Currently there are many resources created for the accumulation of historical and cultural heritage in a digital format. Millions of photographs from the LIFE photo archive, stretching from the 1750s to today, are now available for the first time through the joint work of LIFE and Google (2008). Digital collections of the Science History Institute (https://digital.sciencehistory.org/) includes 6,508 digitized items: artifacts, photographs, advertisements, letters, rare books. Library of Congress (https://www.loc.gov/) and digital collections of UNESCO (https://digital.archives.unesco.org/en/collection) are the most impressive ones.

Though, they have no catalogue helping to establish connections between documents. One of the main problems faced by the creators of these projects was financing. In 2015, UNESCO launched a fundraising project to digitize the archives of the Organization belonging to its predecessors, including the League of Nations International Institute for Intellectual Cooperation. Two years later, thanks to the generous support of the Japanese government, UNESCO launched a major two-year initiative. In partnership with the digitization company Picturae BV, in February 2018, a laboratory was established at the site of UNESCO Headquarters in Paris. Financing a project is a painful question for us as well.

The funding of any research projects by Russian foundations is such that they can willingly provide finance for the launch of the project but not for its support and development. At present, we are not raising sponsor funds since the project of A.P. Ershov's Archive has been virtually completed. The remaining digital projects of the Institute of Informatics Systems of the Siberian Branch of the Russian Academy of Sciences are carried out within the framework of the government assignment to the Institute on the theme "Research of the fundamentals of data structuring, information resources management, creation of information and computing systems and environments for science and education." The purpose of this study is the development of automated support methods for ontology design. The bottleneck in this direction so far is the creation of more accurate search tools, text recognition tools, and hiring qualified personnel.

6   Conclusion and Outlook

Since the mid-1980s, the European community has launched projects supporting specialists engaged in the preservation, conservation and dissemination of knowledge about the heritage with the help of digital reality: Framework Program for Research & Technological Development FR1, 1984–1987, prolonged until 2013, with HORIZON 2020 as its successor [14]. In addition to the programs supporting appropriate research, special-purpose centers were set up in some countries, such as the U.K. and France, to provide the long-term storage and access to software [7, 8]. Moreover, the European Commission is planning to launch a single European Open Science Cloud for storing, exchanging and reusing research data in a variety of areas and support its infrastructure.

In Russia, apparently, the critical mass required for making such decisions at the national level has yet to be achieved. The Russian State Archives have begun publicizing their meetings and reference apparatus fairly recently, later than other institutions keeping historical sources. The Archive of the Russian Academy of Sciences (RAS) is the umbrella association for launching a universal corporate resource (http://www.isaran.ru). The Science Archive of the Siberian Branch RAS, however, neither digitizes its collections nor represents itself in the Internet. This is an urgent issue of the SB RAS and Russian Ministry for Science and Education. The structural changes undergoing in the RAS Siberian Branch in connection with reforming the Russian Academy of Sciences have so far ignored the SB RAS archival activity. Therefore, the future of the SB RAS Science Archive is uncertain. This most valuable collection of documents on the development of Siberian science is in danger

of neglect because the SB RAS Presidium has no funds to maintain or, more importantly, to develop it. The SB RAS Science Archive established simultaneously with the RAS Siberian Branch in 1958 possesses a richest array of representative sources on the history of science in Siberia. It includes 86 collections and 52,219 files including 9,356 personal files. Until now, the Archive's collections have not been digitized for professional or public purposes, and the Archive has no electronic resources of its own (even though the SB RAS State Public Scientific-Technical Library has the Internet connection). With a view to preserving the unique historical documents, we need to digitize them and establish permanent repositories of datasets using cloud technologies. Within the framework of the project SB RAS Open Archive, which is in line with the all-Russia trend for the extensive use of information and communication technologies in the cultural and scientific spheres, the IIS has pioneered the organization of archival work in the RAS Siberian Branch. We expect that our experience will be in demand.

In 2014–2017, the research was partially funded by the RFBR grant 15-07-345A «Establishment and development of scientific schools of programming in leading scientific centers of the USSR», and joint project of RFBR and the Novosibrisk District № 19-49-540001«Institutes of Novosibirsk are named after them: life history of outstanding scientists of the XX century».

Translate by Tatiana Bulyonkova.

References

1. Lynch, C.: Jim Gray's fourth paradigm and the construction of the scientific record. The Fourth Paradigm: Data-Intensive Scientific Discovery. T. Hey, S. Tansley, K. Tolle (eds.). Redmond, Washington, Microsoft Research, 175–182 (2009).
2. Mirskaya, E.Z.: New information technologies in Russian science: history, results, problems and prospects. Science research: coll. Proc. A.I. Rakhitov (ed.). Moscow, INION RAN, 174–200 (2011) (In Russian).
3. Musher, S.L., Bredihin, S.V.: The history of the creation of the Internet in Novosibirsk Akademgorodok // Proceedings of SoRuCom-2017. The fourth international conference «Development of computer technology in Russia and in the former USSR: history and prospects». Zelenograd, 3–5 October 2017 / Ed. by A.N. Tomilin. M.: G.V. Plekhanov Russian university of economics, 236–242 (2017) (In Russian).
4. Krayneva, Irina, Troshkov, Sergey: Archival Information Systems: New Opportunity for Historians // Perspectives of System Informatics. 12th International Andrei P. Ershov Informatics Conference, PSI 2019, Novosibirsk, Russia, July 2–5, 2019, Revised Selected Papers, T. 11964 LNCS, pp. 41–49.
5. Tatarchenko, Ksenia: A House with the Window to the West: The Akademgorodok Computer Center (1958–1993). A dissertation presented to the Faculty of Princeton University in candidacy for the degree of Doctor of Philosophy. Adviser M.D. Gordin, 2006.
6. Boenig-Liptsin, M.: Making Citizens of the Information Age: A Comparative Study of the First Computer Literacy Programs for Children in the United States, France, and the Soviet Union, 1970–1990. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences (2015).
7. Doorn-Moiseenko, T.L.: Electronic Archives and their Role in the Development of the Information Infrastructure of Historical Science. In: Vorontsova E.A., Aiani V.Yu., Pe-

trov Yu.A. (eds.) Role of Archives in Information Support of Historical Science: a collection of articles, 101–117. Moscow, ETERNA (2017).

8. Schurer, K.: Anderson S.J. with the assistance of Duncan J.A.: A Guide to Hictorical Datafiles Held in Machine-Readable Form. Association for History and Computing. Cambridge, 339 p. http://www.aik-sng.ru/text/bullet/8/89-95.pdf (1992).