CREATION OF A REPOSITORY OF ORIGINAL RESEARCH DATA IN A SCIENTIFIC LIBRARY



doi 10.34920/2181-8207/2025/3-182



Tatyana Kalyuzhnaya,
State Public Scientific Technological Library of the Russian Academy of Sciences, Novosibirsk, Russia



Maria Pleshakova,
State Public Scientific Technological Library of the Russian Academy of Sciences, Novosibirsk, Russia

Annotation. The article discusses one of the forms of storing and making publicly available the source research data in the repository of the scientific library. Research data is used to obtain scientific knowledge is constantly accumulating and becomes material for further study, interpretation, as well as synthesis of the results of various studies and comparison or combination with other sets of comparable data, etc. The purpose of the article was to describe an information repository system that provides open access to the source data obtained during research, for their promotion in the scientific information space and to increase the visibility of research results.

An analysis of publications on the topic «Repository of research data» was carried out, which showed that the range of problems considered by specialists is quite wide: from theoretical understanding of various aspects of studying repositories to describing practical experience. In general, publications on the topic have a high level of generalization and theorization of practical experience, which indicates a good level of elaboration of the problem. Special attention is paid to the Russian experience in creating and maintaining research data repositories. A small number of Russian repositories registered in the registry re3data.org. This suggests that it is a fairly new phenomenon for the country. The creation of a repository of original research data in State Public Scientific and Technical Library of the Siberian Branch of the Russian Academy of Sciences (SPSTL SB RAS) is one example of the movement towards the openness of datasets. Such a repository will allow you to place, store and provide access to large source data arrays, which is important for obtaining grants, as well as verification of the findings by specialists, experts and reviewers.

Keywords: Repository, Research Data, Source Data, Dataset, SPSTL SB RAS, Database, IRBIS64

Аннотация. В статье рассматривается одна из форм хранения и предоставления в открытый доступ исходных исследовательских данных – репозиторий научной библиотеки. Исследовательские данные, которые используются для получения научного знания, постоянно накапливаются и становятся материалом для дальнейшего изучения, интерпретации, а также синтеза результатов разных исследований и сравнения или объединения с другими наборами сопоставимых данных и т. д. Целью статьи стало описание информационной системы – репозитория, обеспечивающего открытый доступ к исходным данным, полученным в ходе исследовательской работы, для их продвижения в научном информационном пространстве и повышения видимости результатов исследований.

Проведен анализ публикаций по теме «Репозиторий исследовательских данных», который показал, что круг рассматриваемых специалистами проблем достаточно широк: от теоретического осмысления различных аспектов изучения репозиториев до описания практического опыта. В целом публикации по теме имеют высокий уровень обобщения и теоретизации практического опыта, что свидетельствует о хорошем уровне разработанности проблемы. Особое внимание уделяется российскому опыту создания и ведения репозиториев исследовательских данных. Малое количество российских репозиториев, зарегистрированных в реестре re3data.org, говорит о том, что для страны это достаточно новое явление. Создание репозитория исходных исследовательских данных в научной библиотеке является одним из

примеров движения к открытости наборов данных. Такой репозиторий позволит размещать, хранить и обеспечивать доступ к большим исходным массивам данных, что важно для получения грантов, а также верификации полученных выводов специалистами, экспертами и рецензентами.

Ключевые слова: репозиторий, исходные исследовательские данные, наборы данных, научная библиотека, ГПНТБ СО РАН, база данных, ИРБИС64.

Today, the potential role of libraries in research data management is being actively discussed, which is considered as one of the areas of their work to support scientific research. The initial data from already completed studies become a necessary resource for re-examination and analysis, reinterpretation, possible revision of the results, and integration with other datasets in a specific field of knowledge. This task is performed by the research data repository (RDR), which is a type of repository created to access all research materials used before, during and after scientific research, as well as to preserve them [1]. Offering opportunities for storing, archiving, managing, accessing, and sharing research datasets, RDRs are becoming the primary tool for sharing primary scientific information. This is an important condition for supporting scientific conclusions and building an evidence base on various research issues. RDR provides for the management, sharing, access, and archiving of researcher's datasets; links source data sets with subsequent publications [2], which is the preferred way to share research data [3] and an important condition for supporting conclusions. In open scientific ecosystems, RDRs are used to accelerate global research, promote international collaboration, and innovate at levels previously considered impossible [4].

The purpose of the article is to describe the repository as an information system that provides open access to source research data.

We adhere to the definition of «Research Data Repository», fixed in the registry of research data repositories re3data.org (https://www.re3data.org/suggest), is a subtype of a sustainable information infrastructure that provides long–term storage and access to research data.

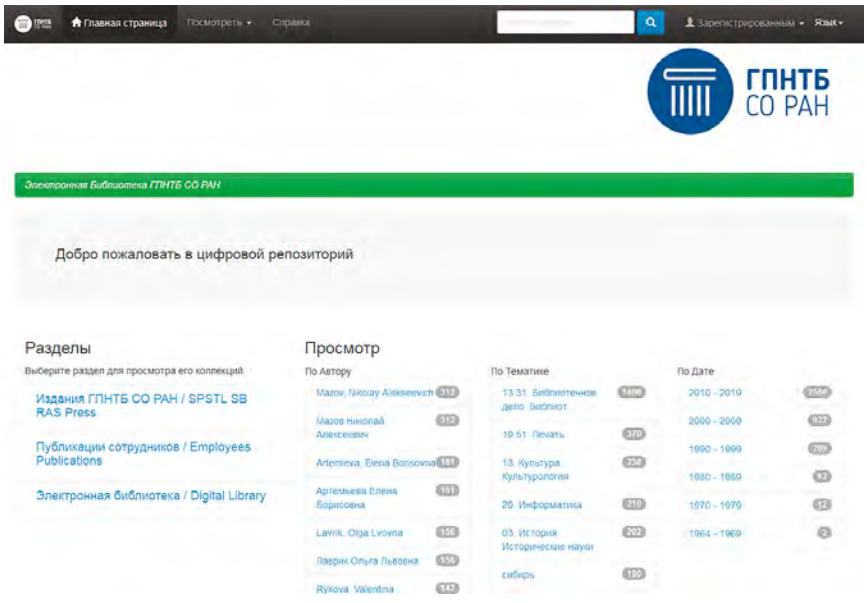
Literature Review

Initially, repositories of initial research data were developed to store quantitative clinical data and share it between researchers. The volume of such repositories can range from small, containing data from a single study conducted by one working group, or data from one academic institution, to extremely large (from several thousand datasets) and designed for numerous scientific studies [5]. The current range of problems discussed in publications on the topic is very wide: from a theoretical understanding of various aspects of studying RDR to a description of practical experience. The level of generalization and theorization of practical experience in publications on the topic of RDR is quite high, and this indicates that certain aspects of the repositories are well developed. At the same time, the opportunity to use the experience already gained by other developers is of great importance for practitioners involved in creating and maintaining research data repositories. An example of creating a repository for archiving language data based on Invenio RDM at the Leibniz Institute for the German Language (IDS) in Mannheim in the period from March 2022 to September 2023, H. Lange is divided. He explains in detail the design stages, provides criteria that influenced decision-making, and specific steps taken during the development of the repository. One of the main decisions that needs to be made, according to H. Lange, is the technology based on which the repository will be created. It can be either a new system, completely developed from scratch, or a setup based on existing software. Another option may be to use a commercial hosting provider. Each of these options has its advantages and challenges. The individual system, according to H. Lange, has the main advantage that it can accurately meet all the requirements of the organization [6]. This idea refers to another study analyzing the international landscape of geoscience research data repositories available on re3data.org ([1], because of which it became known that the software used for their development is mostly unknown. Widely used by government agencies and library communities advocating open access to scientific knowledge, Ckan, DSpace, DataVerse, e-Prints, and Fedora are rarely used in geosciences.

In general, special attention is paid to the study of thematic repositories in literature. The

We adhere to the definition of «Research Data Repository», fixed in the registry of research data repositories re3data.org (https://www.re3data.org/suggest), is a subtype of a sustainable information infrastructure that provides long–term storage and access to research data

The level of generalization and theorization of practical experience in publications on the topic of RDR is quite high, and this indicates that certain aspects of the repositories are well developed. At the same time, the opportunity to use the experience already gained by other developers is of great importance for practitioners involved in creating and maintaining research data repositories



problems of their creation and functioning are widely discussed [1, 7, 8]. Semeler, Sales, Pinto, González-Valiente, based on the types of data stored in these Earth sciences research data repositories (office documents, scientific and statistical data, images, and raw data), pointed to the interdisciplinary nature and diversity of research that is being conducted in this field [1]. Hypothetically, these conclusions can be extended to any field of knowledge, although the authors themselves consider conducting relevant research to be promising tasks. Nevertheless, this suggests the need to find flexible and scalable solutions that allow organizations to process large amounts of heterogeneous data using various tools and platforms. H. Zilong and F. Wei offers its own solution to this issue. In order to develop effective strategies for managing research data, preserving it, and modernizing the role of data in the research ecosystem, they presented the development of an integrated hybrid data warehouse architecture adapted to the needs of institutional repositories. The proposed architecture solves the problems associated with the increase in data volume, heterogeneity and multi-source nature of research data. The complex structure for the architecture of such repositories includes five fundamental levels: data collection, storage, processing, management and services. This structure can be refined and adapted during discussions with data management specialists, repository administrators, and researchers [9]. Another important issue related to the functioning of RDR is the description of research data using metadata schemes [10–12]. The study by T. R. Manu and N. Hasan provides a clear understanding of the metadata schemas available for research data repositories. After comprehensively considering the issue of choosing the appropriate metadata schema for the research data repository based on FAIRsharing analysis, which covers more than 1.600 metadata schemas in all major fields of science and technology, medicine, management, art, humanities and social sciences, they found that more than 1.500 metadata schemas with interdisciplinary topics are available on the FAIRsharing platform and have been used, but only 78% were active and ready for use. Therefore, the authors believe that before using the scheme, it is important to ensure its stability and applicability to all types of data [12].

The complex structure for the architecture of such repositories includes five fundamental levels: data collection, storage, processing, management and services. This structure can be refined and adapted during discussions with data management specialists, repository administrators, and researchers

The process of source data management and the participation of library specialists in it is described in detail, including the preparation of relevant recommendations [10, 13, 14]. The role of library specialists is most significant in the processes related to the formation of content. The specializations of librarians whose activities are related to research data management are also substantiated: a data service librarian, a data curation librarian, a research data management librarian («Data Librarian», «databrarian»). The main competencies of such specialists are formulated. The issues of forming such services no longer occupy such a significant place in the strategic plans of libraries in developed countries, as they did 10 years ago, when they were a priority everywhere. Now the work is aimed at expanding services and rethinking their nature. And these «searches» relate mainly to the «upstream» processes in working with research data, that is, the possibilities of assisting scientists in processing and analyzing data and their further publication (data articles, data paper). While the «downstream» processes – data collection and curation – have successfully passed the testing stage and adjustments, and in particular, in large scientific libraries in the United States, the availability of appropriate specialists and their performance of prescribed functions have now become commonplace [15–17]. Nevertheless, working with repositories is not limited to these processes, which is noticeable by the highlighted main blocks of RDR [18]:

1. Content (in addition to the research data itself, such repositories contain original reports, publications, metadata, statistics and other materials related to primary data. This makes

it possible to reproduce and validate the survey results).

2. The management process (provided by competent specialists with knowledge of the repository's subject area).

3. Infrastructure (sufficient server space, database, data management program, and strategic planning to enable development).

4. Users.

5. Access.

The listed components are the basic set of system elements for storing and using research data. But such structures have much greater opportunities for scientific communication, and they have significant network functioning potential. The prospects for building a global scientific ecosystem are discussed, with RDR becoming an indispensable element, allowing researchers to publish, track, discover, and reuse data from other scientists and expand their research through available materials. Their future is seen precisely in integration into research networks (local, global), where, firstly, geographically distributed consortia of state or international levels will arise on the basis of combining repositories from different institutions into a single network.; Secondly, at the institutional level, RDR will be linked to the digital collections repository; thirdly, such integration will ensure the integration of datasets with open access academic journal platforms or directly link datasets through citations and reference links to published results [4].

Materials and Methods

This study used a descriptive approach. The search and analytical method were used to track RDR in the field of social sciences and humanities, including library and information activities, on the websites of the registry of research data repositories re3data.org and the «Catalog of Data Catalogs» of Info-culture (https://www.datacatalogs.ru). At various stages of the creation of the database «Repository of research data of the SPSTL SB RAS», the modeling method was used. It was used to optimize the process of forming a data warehouse and ensure compliance with the stated requirements. This research method allowed us to implement the following tasks: to develop the database structure, including the main fields and subfields for entering metadata; to determine the necessary data formats and establish requirements for the composition of metadata, considering the specifics of the description objects. The modeling method has become a tool that has contributed to the effective design, development and implementation of a repository of research data in the activities of a scientific library and has ensured its quality and ease of use. The use of the modeling method made it possible to achieve the level of functionality specified during the design.

This research method allowed us to implement the following tasks: to develop the database structure, including the main fields and subfields for entering metadata; to determine the necessary data formats and establish requirements for the composition of metadata, considering the specifics of the description objects

Preliminary Stage of Development of the Repository of Research Data of the SPSTL SB RAS

The need for a repository of research data arose in 2022 in connection with the start of a new scientific project, the tasks of which included conducting a longitudinal sociological study and saving the collected data for further consolidation and generalization. At the stage leading up to the development of the RDR, it was necessary to establish whether there is an existing open repository that meets our requirements: storing research data on humanitarian and social issues (including library science research data) of various types (raw, processed), different formats, and attaching related materials.

In the registry of research data repositories re3data.org At that time, about 3.000 repositories were registered, 24 of them Russian, 21 of which contained natural science data, three – humanitarian and social. These are the repository of the Moscow State Institute of International Relations (MGIMO), the University Information System RUSSIA (UIS RUSSIA) and Codex Sinaiticus. The MGIMO Repository serves as a center for academic resources, research publications, and educational materials related to the various fields of study offered at MGIMO and includes access to the institute's library resources. UIS RUSSIA was developed as an electronic library for research and educational purposes, but since 2003 it has focused on statistical databases to create infrastructure for educational courses. The Codex Sinaiticus project contains the Christian Bible in Greek, including the oldest complete copy of the New Testament, and functions to reunite the entire manuscript in digital form and make it available to a global audience. As you can see, none of the three repositories meet our needs.

UIS RUSSIA was developed as an electronic library for research and educational purposes, but since 2003 it has focused on statistical databases to create infrastructure for educational courses. The Codex Sinaiticus project contains the Christian Bible in Greek, including the oldest complete copy of the New Testament, and functions to reunite the entire manuscript in digital form and make it available to a global audience

However, we have noted that not all repositories are registered in re3data.org (the situation is typical for both Russian and foreign repositories), and this is assessed as an unfavorable factor, since it is not possible to ▶



track their existence and make an assessment of the representation of different types of RDR, that is, researchers cannot choose the most suitable repository for hosting and storing their own research data. Information about Russian RDRs, as well as international portals and data archives, including those from neighboring countries: Armenia, Kazakhstan, Uzbekistan, etc., potentially interesting to the Russian-speaking audience, is collected in the «Catalog of Data Catalogs» of Info-culture (https://www.datacatalogs.ru). The data topic is represented by 109 headings; data sets are sorted by category, repository software, region, resource owner, country, and content type. An analysis of the resources of the «Catalog...» showed that it also does not contain a suitable repository for us. Thus, there was not a single registered RDR in Russia that meets our requirements. In terms of thematic coverage, repositories on natural sciences are being created the most. There are few of them in the field of social sciences and humanities, but they are generally absent in the sciences of the library cycle, not only in the Russian segment, but also abroad [15].

Taking into account the fact that there is no existing repository for storing the initial research data we received, we had to choose between developing our own system that meets our requirements as much as possible, which, however, will require significant resources, and a simpler option – adapting the finished software to our own needs

Taking into account the fact that there is no existing repository for storing the initial research data we received, we had to choose between developing our own system that meets our requirements as much as possible, which, however, will require significant resources, and a simpler option – adapting the finished software to our own needs.

Based on the analysis of literary sources, we focused on the consideration of two digital storage support systems – DSpace and Eprints as the most popular platforms for repositories that allow creating, distributing and storing digital materials. The choice was made for free software with various digital data formats DSpace as the most used for open repositories (it is used in more than 70% of organizations and universities in Russia). The possibility of the subsequent entry of the SPSTL SB RAS repository into the NORAa system and the associated need to use standard software also influenced the choice.

At the first stage of the repository organization, employees of the Computer and Multiplication Technology department installed the DSpace server part, downloaded records from separate databases of the SPSTL SB RAS running on the Russian library automation system (Figure 1), and tested the software, evaluating the capabilities of the DSpace platform. The system interface is simple and user-friendly, not overloaded with redundant information.

In 2023, the library refused to use the foreign DSpace system due to the lack of specialists to support it, it was decided to use the Russian library automation system IRBIS64.

Development of a Repository of Initial Research Data of the SPSTL SB RAS on the IRBIS64 Platform

IRBIS64 is a Russian multifunctional software package that provides automation of all standard library processes. The choice was due to the widespread use and successful administration of this system at the SPSTL SB RAS.

When designing, we were guided by the principles for research data repositories formulated by Abdul Azeez [18]: replication – the use of source data, metadata and statistical codes to replicate the research process; integration – combining (establishing links) between a set of raw data and a project implemented on its basis and/or prepared publications inside and outside the repository beyond it; open access – providing users with raw research data in accordance with the principles of openness while respecting legal and intellectual rights and confidentiality rules; management – ensuring permanent access to data and their long-term use. Therefore, the terms of reference included the following tasks:

1. Creation of the database «Repository of research data of the SPSTL SB RAS» (hereinafter referred to as the database «Repository...») with a set of necessary fields and subfields for entering metadata and forming a search image of the resource and search instructions (Figure 2).
2. Preparation of output formats.
3. Providing access to the database

«Repository...» on the website of the SPSTL SB RAS through the web-IRBIS.

4. Consideration of the issue of integrating the repository of the SPSTL SB RAS into the NORA system and the development of a transitional scheme.

In accordance with the terms of reference, the staff of the Automated Systems Department of the SPSTL SB RAS created the Repository database and provided it for official use to the database administrator. When defining the metadata schema, we focused on the IRBIS structure based on the UNIMARC exchange format, and taking into account the specifics of the description object, we added additional fields/subfields to the worksheets.

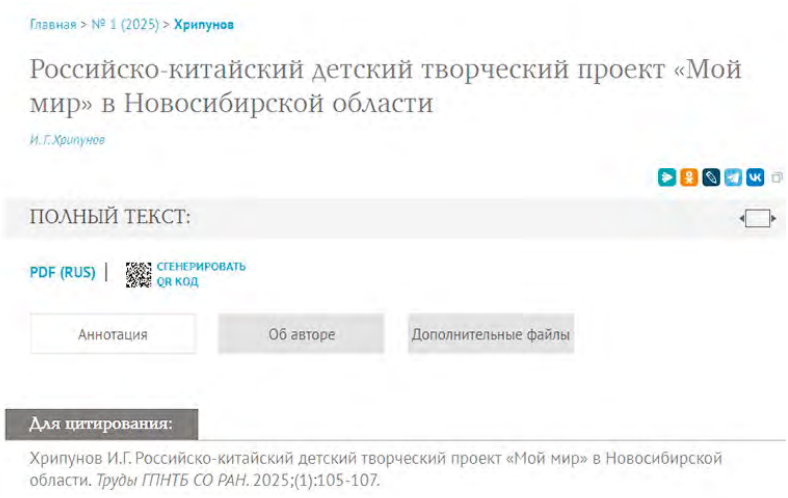
The specifics of the «Repository...» database for IRBIS64 were as follows:

The R&D/Grant Number field indicates the R&D registration number for which the research project is being conducted. The «Date/period of data collection» field is intended to record the period in which the initial scientific data was collected.

The data source (the «Data source» field) used in data collection and analysis can be a website, a document, a survey, an experiment, a field study, a video recording, an author's presentation, etc.; types of resources (the «Resource» field) included in the repository type of data: text, table, figure, graph, photograph, presentation, video, map, coordinates, database, etc.; format: doc/docx, xls/xlsx, pdf, jpg, ppt/pptx, mp4, mpeg, zip, etc.; resource volume: pages, pages, file, slide, byte.

The «Data collection procedure» field describes the steps, methods, and others involved in collecting the source data. In the «Retrospective» field, it is suggested to indicate the period that was studied. The «Link – external object» field contains links to the source datasets that are stored on the library server, indicating the resource location, title, and type of resource. In the «Related materials» field, links to presentations or video materials are provided, indicating the location, title and type of resource, as well as the start time of the speakers (for the video). The «Link to publications» field is filled in with a bibliographic description of the articles and a link to a remote resource for the relationship of the original research data with articles where these already processed, interpreted data are recorded.

The Repository database ... is updated with initial research data: lists of research objects, survey data, statistical samples,



information arrays, object-characteristic tables, etc.; materials related to the source data: presentations of reports at conferences and seminars (as a fact of testing research results), video presentations at scientific events, etc., links to publications.

All the source data received from the project participants is adapted to a single input format:

- based on the context of the data, the title of the resource is determined, an abstract is compiled and keywords are highlighted;
- the initial data presented in the form of tables, text files, presentations, and video files are processed: text documents and presentations are converted to *.pdf format; the start time of the speech is recorded for video files; a group of files of the same format (for example, *.xlsx) is archived (*.zip) (for example, statistical tables of magazine issues for different years according to a certain socio-demographic characteristic).

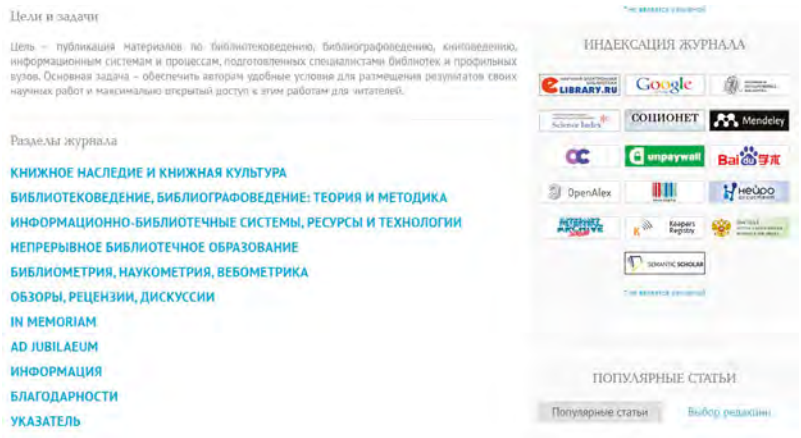
The industrial version of the repository appeared in July 2023. Using the web-IRBIS program, the Repository database was posted on the website of the SPSTL SB RAS in the Resources section (Figure 3); the scientific community was provided with open access to it.

Results and Discussion

Each record of the research data repository of the SPSTL SB RAS contains all the necessary metadata to identify and find the required source data for a thematic query. It identifies the authors, provides the title, abstract, and keywords of the resource, including in English, provides the period of data collection and a description of the collection procedure, shows the sources used in the collection and analysis, and provides links to full-text resources

The «Data collection procedure» field describes the steps, methods, and others involved in collecting the source data. In the «Retrospective» field, it is suggested to indicate the period that was studied

Each record of the research data repository of the SPSTL SB RAS contains all the necessary metadata to identify and find the required source data for a thematic query



containing all the source data and related materials (videos and presentations). The repository can be searched by the following fields: author, resource title, abstract, keywords, index of the State Rubricator of Scientific and Technical Information, type of resource, data source, data collection procedure, conference (venue and date, name) (see Figure 3). The description of the datasets is accompanied by a citation link, which is generated automatically, for example:

Kalyuzhnaya T. A., Pleshakova M. A., Bazhenov S. R., Parshikov R. M. Statistics of the Issuance of Popular Science Journals from the Fund of the SPSTL SB RAS (01.01.2023–08/31/2023). Datasets. The Repository of Research Data of the SPSTL SB RAS. 2023. Entry 22. URL: https://webirbis.spsl.nsc.ru/irbis64r_01/cgi/cgiirbis_64.exe?I21DBN=REPOZITORIY&P21DBN=REPOZITORIY&Z21ID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=fullwebr&C21COM=S&2_S21P02=1&2_S21P03=HDO1=&2_S21STR=*****

From the point of view of copyright protection, all data sets in the repository of the SPSTL SB RAS have the degree of accessibility determined by the copyright holder and are under the Creative Commons CC BY license (Figure 3).

Replenishment of the repository of the SPSTL SB RAS is currently being carried out in the current mode by the librarian of research data as new source data is received. It is assumed that in the future, the repository will be filled by the scientists themselves, and the librarian will oversee this activity. To do this, it is planned to conduct training webinars on working with the repository, as well as provide technical support and advice. The librarian's main responsibilities for managing the research data repository will remain access and rights management.; ensuring long-term preservation, including support for data formats that will be relevant in the future; monitoring data usage and user activity; conducting regular assessments of data quality and relevance; maintaining documentation on data management processes and their changes; preparing reports on the status of the repository and its use.

For now, the development of a new user-friendly, simple and aesthetic interface remains a promising task, as it requires significant investments.

In Russia, the creation of repositories for storing initial research data is a relatively new phenomenon. However, the need for such storage facilities is obvious and justified by several factors. First, in the digital age, the volume of source data is growing exponentially and the risk of their loss is increasing

Conclusion

In Russia, the creation of repositories for storing initial research data is a relatively new phenomenon. However, the need for such storage facilities is obvious and justified by several factors. First, in the digital age, the volume of source data is growing exponentially and the risk of their loss is increasing. There is also a growing need for reviewers to assess the reliability of research results through access to source data. Finally, repositories provide opportunities for sharing, reusing, and combining similar data to conduct new research.

These considerations became the basis for the development of the repository of the SPSTL SB RAS. It was originally conceived as a repository of source data for scientific projects implemented in the library, but its creators see much broader prospects. The SPSTL SB RAS is one of the leading research institutes in the field of library science, book studies, bibliography and applied informatics, and publishes four scientific journals as part of its work. The existing repository can also serve as a source data repository for articles published in these publications. This has already become the norm in global practice: for example, Springer Nature's data usage policy recommends that all datasets confirming the analysis and conclusions of the study be publicly available at the time of publication. This practice has not yet become widespread in Russia, but editors of scientific journals may recommend that authors post sets of source data in repositories and provide links to them when submitting articles. At the same time, reviewers should be able to follow these links in their personal accounts to evaluate the manuscript in terms of the correctness of data use. The repository, in turn, should provide the ability to host datasets without immediate public access, which will allow researchers to deposit data before accepting or publishing their manuscript.

Today, links to the source data in

scientific articles posted in the repository are becoming an ethical requirement of scientific communications, ensuring the availability of data to the scientific community in addition to the final research results. In addition, archiving research data in a repository with assigned identifiers guarantees the copyright holder the same citation opportunity as for publications.

Gratitude: we would like to thank V. I. Moskvin, an employee of the Department of Automated Systems at the SPSTL SB RAS, for creating the Repository database.

References

1. Semeler A, Sales L, Pinto A & González-Valiente C (2024) Geosciences research data repositories landscape: re3data.org as a datasource. Revista Ibero-Americana de Ciência da Informação 17: 524-548. https://doi.org/10.26512/rici.v17.n3.2024.53645

2. Uzwyshyn RJ (2016) Research data repositories: the what, when, why and how. Computers in Libraries 36(3): 18-21.

3. Zeldina MM (comp.) (2019) The policy of the exchange of research data (Data Sharing Policy). Moscow.

4. Uzwyshyn RJ (2022) Research data repositories and global scholarly ecosystem possibilities. ResearchGate: social network. Available at https://www.researchgate.net/publication/357286044_Research_Data_Repositories_and_Global_Scholarly_Ecosystem_Possibilities (accessed 02.04.2025).

5. Belyakova OA, Makhnin PN & Sapegin SV (2019) Platform for the collection, storage and analysis of multimodal data generated by online communication subjects. Monitoring. Nauka i tekhnologii 4: 73-84. https://doi.org/10.25714/MNT.2019.42.011

6. Lange H (2024) Setting up a Research Data Repository Based on Invenio RDM: An Experience Report. DNHB 2024: Digital Humanities in the Nodic and Baltic Countries 8th Conference, Reykjavík, Iceland, 27–31 May 2024, 6(1). https://doi.org/10.5617/dhnbpub.11515

7. Arend D, Junker A, Scholz U, Schüler D, Wylie J & Lange M (2016) PGP repository: a plant phenomics and genomics data publication infrastructure. Database, article ID baw033. https://doi.org/10.1093/database/baw033

8. Assante M, Candela L, Castelli D & Tani A (2016) Are scientific data repositories coping with research data publishing? Data Science Journal 15: 1-24. https://doi.org/10.5334/dsj-2016-006

9. Zilong H & Wei F (2024) Research data management in institutional repositories: an architectural approach using data lakehouses. Digital Library Perspectives 41: 145-178 https://doi.org/10.1108/DLP-02-2024-0022

10. Bychkova EF & Kolosov KA (2022) Repositories of research data in the field of ecology and environmental sciences. Nauchno-issledovatel'skaya rabota v bibliotechnoi teorii i praktike: ezhegod. mezhvedomstv. sb. nauch. tr. Moscow, pp. 19-29.

11. Goncharov MV & Kolosov KA (2022) Research data repositories as a new direction for the development of open archives. Nauchno-issledovatel'skaya rabota v bibliotechnoi teorii i praktike: ezhegod. mezhvedomstv. sb. nauch. tr. Moscow, pp. 35-48.

12. Manu TR & Hasan N (2023) Landscape of Metadata Schemas for Research Data Repositories: FAIRsharing Analysis. 26th International Symposium, ETD 2023, Gandhinagar, Gujarat, 26-28 October, 2023.

13. Amorim RC, Castro JA, Rocha da Silva J & Ribeiro C (2017) A comparison of research data management platforms: architecture, flexible metadata and interoperability. Universal Access in the Information Society 16(4): 851-862. https://doi.org/10.1007/s10209-016-0475-y

14. Rychikhin AK & Nuriev VA (2020) Organization of the life cycle of research data. Sistemy i sredstva informatiki 30(3): 81-96.

15. Gowen E & Meier JJ (2020) Research data management services and strategic planning in libraries today: a longitudinal study. Journal of Librarianship and Scholarly Communication 8(1): eP2336. https://doi.org/10.7710/2162-3309.2336

16. Minarso C & Suprapto A (2021) Kompetensi data librarian dalam research data management: systematic literature review. BACA: Jurnal Dokumentasi dan Informasi 42(1): 1-17.

17. Newton MP, Miller CC & Bracke MS (2010) Librarian roles in institutional repository data set collecting: outcomes of a research library task force. Collection Management 36(1): 53-67. https://doi.org/10.1080/01462679.2011.530546

18. Abdul Azeez TA (2019) Open research data repositories: a librarian's perspective. In: KELPRO Silver Jubilee Souvenir. Kerala Libr. Professionals' Organization, pp. 47-54.