

УДК 001.891:004:027.021

<https://doi.org/10.20913/1815-3186-2024-3-15-27>

Репозиторий исследовательских данных ГПНТБ СО РАН

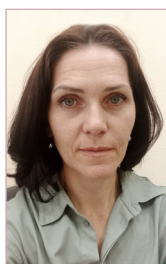
Т. А. Калюжная[✉], М. А. Плешакова

**Калюжная
Татьяна
Альбертовна,**

Государственная
публичная научно-
техническая
библиотека
Сибирского
отделения

Российской академии наук,
ул. Восход, 15, Новосибирск,
630102, Россия,
кандидат педагогических наук,
старший научный сотрудник
лаборатории информационно-
системного анализа

ORCID: 0000-0001-8637-2793

e-mail: Kalyuzhnaya@spsl.nsc.ru

**Плешакова
Мария
Александровна,**

Государственная
публичная научно-
техническая
библиотека
Сибирского
отделения

Российской академии наук,
ул. Восход, 15, Новосибирск,
630102, Россия,
кандидат педагогических наук,
заведующая лабораторией
информационно-системного
анализа, старший научный
сотрудник

ORCID: 0000-0002-2455-4636

e-mail: Pleshakova@spsl.nsc.ru

Аннотация. Исходные данные, накопленные в процессе научно-исследовательской деятельности, – ценный ресурс, который может стать базой для дальнейшего изучения, новой интерпретации, пересмотра результатов, объединения или сравнения с другими наборами сопоставимых данных в конкретной области знания, проведения метаанализа и др. Предоставление этих данных в открытый научный доступ позволяет повысить видимость результатов исследований. В статье акцентируется внимание на форме организации исходных исследовательских данных: онлайн-репозитории. Цель статьи – представить репозиторий научной библиотеки как информационную систему, обеспечивающую открытый доступ к исходным исследовательским данным для их продвижения в научном информационном пространстве. Проведенный анализ отечественных и зарубежных публикаций по теме «Репозиторий исследовательских данных» (РИД) показал, что отечественных работ, посвященных данному вопросу, пока недостаточно, хотя круг рассматриваемых специалистами проблем достаточно широк: от теоретического осмысления различных аспектов изучения РИД до описания практического опыта. Зарубежные публикации отмечаются более высоким уровнем обобщения и теоретизации практического опыта. Присутствие ряда российских репозитриев исследовательских данных на просторах интернета и их малое количество, зарегистрированное в реестре re3data.org, говорит о том, что для России это достаточно новое явление или оно распространено не широко. Создание репозитория исходных исследовательских данных в ГПНТБ СО РАН является одним из примеров движения к открытости наборов данных: он позволит специалистам, экспертам и рецензентам размещать, хранить и обеспечивать доступ к исходным массивам данных для верификации полученных выводов.

Ключевые слова: репозиторий, исследовательские данные, исходные данные, набор данных, ГПНТБ СО РАН, база данных, ИРБИС64

Для цитирования: Калюжная Т. А., Плешакова М. А. Репозиторий исследовательских данных ГПНТБ СО РАН // Библиосфера. 2024. № 3. С. 15–27. <https://doi.org/10.20913/1815-3186-2024-3-15-27>.

Статья поступила в редакцию 24.04.2024

Получена после доработки 18.07.2024

Принята для публикации 30.08.2024

© Т. А. Калюжная, М. А. Плешакова, 2024

Repository of SPSTL SB RAS Research Data

Tatyana A. Kalyuzhnaya✉, Maria A. Pleshakova

Kalyuzhnaya Tatyana Albertovna,
State Public Scientific Technological
Library of the Siberian Branch
of the Russian Academy of Sciences,
15 Voskhod St., Novosibirsk, 630102,
Russia,
Candidate of Pedagogical Sciences,
Senior Researcher

ORCID: 0000-0001-8637-2793
e-mail: Kalyuzhnaya@spsl.nsc.ru

Pleshakova Maria Alexandrovna,
State Public Scientific Technological
Library of the Siberian Branch
of the Russian Academy of Sciences,
15 Voskhod St., Novosibirsk, 630102,
Russia,
Candidate of Pedagogical Sciences,
Senior Researcher,
Head of the Laboratory of
Information and System Analysis

ORCID: 0000-0002-2455-4636
e-mail: Pleshakova@spsl.nsc.ru

Received 24.04.2024

Revised 18.07.2024

Accepted 30.08.2024

Abstract. The initial data accumulated in the course of research activities is a valuable resource that can become the basis for further study, reinterpretation, revision of results, combining or comparing with other sets of comparable data in a specific field of knowledge, conducting meta-analysis, etc. Providing this data to open scientific access makes it possible to increase the visibility of research results. The article focuses on the form of organization of the initial research data: an online repository. The purpose of the article is to present the repository of the scientific library as an information system providing open access to the original research data for their promotion in the scientific information space. The analysis of domestic and foreign publications on the topic “Research Data Repository” (RID) has shown that there are not enough domestic works on this issue yet, although the range of problems considered by specialists is rather wide: from theoretical understanding of various aspects of the study of RID to the description of practical experience. Foreign publications are marked by a higher level of generalization and theorization of practical experience. The presence of some Russian repositories of research data on the Internet and their small number registered in the Registry re3data.org suggest that this is a fairly new phenomenon for Russia or not widely spread. The creation of a repository of initial research data at the State Scientific Research Institute of the Siberian Branch of the Russian Academy of Sciences is one example of the movement towards openness of data sets: it will allow specialists, experts and reviewers to place, store and provide access to the initial data sets to verify the findings.

Keywords: repository, research data, source data, dataset, SPSTL SB RAS, database, Irbis64

Citation: Kalyuzhnaya T. A., Pleshakova M. A. Repository of SPSTL SB RAS Research Data. *Bibliosphere*. 2024. № 3. P. 15–27. <https://doi.org/10.20913/1815-3186-2024-3-15-27>.

Введение

Репозитории открытого доступа (ОД), содержащие материалы по результатам научных исследований, обеспечивают их длительное хранение и свободное использование (Гавриков, 2021; Засурский и др., 2020; Юдина, Федотова, 2020 и др.). Вопросы, касающиеся технологии их создания, наполнения различного рода документами, функционирования, влияния на результаты научных исследований, перспектив и особенностей развития и прочие, изучаются специалистами различных областей знания как в России, так и за рубежом (Гончаров, Колосов, 2022; Юмашева, 2023; Singh, 2016 и др.).

До сих пор не существует разработанной типологии репозиторий, однако на основании различных аспектов классификации различают тематические (предметные) или дисциплинарные (по тематическому охвату); институциональные, национальные и международные (по принадлежности репозитория) (Открытая наука..., 2020). По характеру содержащихся материалов из общего числа выделяют репозитории

исследовательских данных¹ (РИД). Этот тип репозитория заслуживает отдельного внимания, поскольку обеспечивает управление наборами данных исследователя, совместное использование, доступ к ним и архивирование; связывает наборы исходных данных с последующими публикациями (Uzwyshyn, 2016), что является предпочтительным способом обмена исследовательскими данными (Политика..., 2019) и важным условием подкрепления выводов. В открытых научных экосистемах РИД используются для ускорения глобальных исследований, содействия международному сотрудничеству и внедрения инноваций на уровнях, которые ранее считались невозможными (Uzwyshyn, 2022). Таким образом, исходные данные, накопленные в процессе научно-исследовательской деятельности, становятся ценным ресурсом и базой для дальнейшего

¹ Авторы опираются на следующее определение: исследовательские данные – это количественная или качественная информация, собранная учеными в ходе проведенного ими исследования (Косычева, Хорохорина, 2020). «Исследовательские данные могут быть получены в результате экспериментов, наблюдений, моделирования, посредством опросов или интервью, или другими способами, или созданы на основе уже существующих данных» (Там же, с. 7).

изучения, новой интерпретации, пересмотра результатов и/или объединения/дополнения/сравнения с другими наборами эквивалентных сопоставимых данных в конкретной области знания, проведения метаанализа. При этом предоставление исследовательских данных в открытый научный доступ позволяет повысить видимость результатов исследований.

В статье акцентируется внимание на особенностях работы с исходными исследовательскими данными, которые, как и опубликованные материалы, необходимо продвигать в научном информационном пространстве, а также на форме их организации и представления – онлайн-репозитории. Цель статьи – представить репозиторий научной библиотеки как информационную систему, обеспечивающую открытый доступ к исходным исследовательским данным.

В своей работе мы будем придерживаться определения, зафиксированного в реестре хранилищ исследовательских данных [re3data.org](https://www.re3data.org/): «Репозиторий исследовательских данных – это подтип устойчивой информационной инфраструктуры, которая обеспечивает долгосрочное хранение и доступ к исследовательским данным, являющимся основой научной публикации» (<https://www.re3data.org/suggest>).

Чтобы достичь поставленной цели, мы наметили следующие задачи:

- 1) дать краткий обзор литературных источников по проблеме РИД;
- 2) представить отдельные репозитории исследовательских данных в России и за рубежом;
- 3) описать проект создания репозитория исследовательских данных в ГПНТБ СО РАН.

Для решения первой задачи был проведен анализ отечественных и зарубежных публикаций по теме «Репозиторий исследовательских данных».

1. Обзор литературы

Анализ отечественных публикаций, посвященных репозиториям, показал, что внимание в них сосредоточено на хранилищах открытых данных, содержащих как опубликованные, так и неопубликованные (непубликуемые) материалы различных типов: статьи, диссертации и авторефераты, монографии, материалы конференций, учебники, методические пособия, курсы лекций, студенческие работы и т. д. Такие репозитории формируются из-за необходимости интеграции информационных ресурсов организаций (учреждений) в единое информационное пространство, например для создания информационно-образовательной среды вуза (Стукалова, 2023 и др.).

Вопросы создания и функционирования репозитория исходных исследовательских данных освещены значительно меньше. Первоначально они разрабатывались для хранения количественных

клинических данных и обмена ими между исследователями. Объем таких хранилищ может быть от небольшого, вмещающего данные одного исследования, проведенного одной рабочей группой, или данные одного академического учреждения, до сверхбольшого (от нескольких тысяч датасетов), рассчитанного на многочисленные научные исследования (Белякова и др., 2019).

По запросу «репозиторий данных» в РИНЦ было получено 559 публикаций, из которых по итогам просмотра аннотаций и полных текстов только 19 работ оказались релевантными нашей тематике.

Круг проблем, рассматриваемых в отобранных публикациях, достаточно широк: от теоретического осмысления различных аспектов изучения РИД до описания практического опыта. Особое внимание уделено тематическим репозиториям: вопросам создания (Борбат, Лищук, 2020; Куликов и др., 2023; Щекотилов и др., 2019 и др.), выбора независимых хранилищ для депонирования наборов первичных данных (Бычкова, Колосов, 2022; Чадин, 2018), описания исследовательских данных с помощью схем метаданных (Беглер, Ануфриев, 2021; Бычкова, Колосов, 2022; Гончаров, Колосов, 2022 и др.), управления исходными данными и участия в этой работе специалистов библиотек, включая подготовку соответствующих рекомендаций (Бычкова, Колосов, 2022; Редькина, 2019; Рычихин, Нуриев, 2020 и др.).

В зарубежных публикациях отмечен более высокий уровень обобщения и теоретизации практического опыта. Прежде всего, обоснованы специализации библиотекарей, чья деятельность связана с менеджментом исследовательских данных: библиотекарь службы данных, библиотекарь по курированию данных, библиотекарь по управлению исследовательскими данными («Data Librarian», «databrarian»); и сформулированы основные компетенции таких специалистов. Вопросы формирования таких служб уже не занимают столь значительного места в стратегических планах библиотек развитых стран, как это было еще 10 лет назад, когда повсеместно они являлись приоритетными, – теперь работа направлена на расширение услуг и переосмысление их характера. И эти «поиски» касаются в основном «восходящих» процессов в работе с исследовательскими данными, то есть возможностей оказания помощи ученым в обработке и анализе данных и их дальнейшей публикации (data articles, data paper), в то время как «нисходящие» процессы – сбор и курирование данных – успешно прошли стадию апробации и корректировки, и, в частности, в крупных научных библиотеках США в настоящее время наличие соответствующих специалистов и выполнение ими предписанных функций стали обыденным явлением (Gowen, Meier, 2020; Minarso, Suprpto, 2021; Newton et al., 2010).

Тем не менее работа с репозиториями не ограничивается названными процессами, что заметно по выделенным основным блокам РИД (Azeev, 2019):

1. Контент (кроме самих исследовательских данных, в таких репозиториях размещаются оригинальные отчеты, публикации, метаданные, статистические данные и другие материалы, связанные с первичными данными. Это делает возможным воспроизведение и валидацию результатов изысканий).

2. Процесс управления (обеспечивается компетентными специалистами, обладающими знаниями предметной области репозитория).

3. Инфраструктура (достаточное серверное пространство, база данных, программа управления данными и стратегическое планирование для обеспечения возможности развития).

4. Пользователи.

5. Доступ.

Как видим, роль библиотечных специалистов наиболее значительна в процессах, связанных с формированием контента. Перечисленные компоненты – это базовый набор элементов системы для хранения и использования исследовательских данных. Но такие структуры имеют гораздо большие возможности для научной коммуникации. Уже сейчас обсуждаются перспективы построения глобальной научной экосистемы, незаменимым элементом которой становится РИД, позволяющий исследователям публиковать, отслеживать, обнаруживать, повторно применять данные других ученых и расширять свои исследования за счет доступных материалов. Обозначен потенциал сетевого функционирования РИД. Их будущее видится именно в интеграции в исследовательские сети (локальные, глобальные), где, во-первых, на основе объединения репозиторий разных учреждений в единую сеть возникнут географически распределенные консорциумы государственного или международного уровней; во-вторых, на институциональном уровне РИД будет иметь связь с репозиторием цифровых коллекций; в-третьих, такая интеграция обеспечит объединение наборов данных с платформами академических журналов открытого доступа или напрямую свяжет наборы данных через цитирование и справочные ссылки с опубликованными результатами (Uzwyszyn, 2022).

2. Обзор некоторых репозиторий исследовательских данных в России и за рубежом

В реестре хранилищ исследовательских данных re3data.org на 1 марта 2024 г. зарегистрировано 3195 репозиторий, из них 23 российских, к одному из которых нет доступа. По типу контента выделяются аудиовизуальные данные,

изображения, структурированная графика, стандартные офисные документы, архивные данные, обычный и структурированный тексты, сетевые данные, необработанные данные, научно-статистические данные, базы данных, программные приложения, данные конфигурации, исходный код и др. Большинство российских РИД, зарегистрированных в реестре (20 из 22), относятся к естественным наукам, два – к гуманитарным и социальным.

Однако нами отмечено, что регистрируются в реестре не все репозитории (ситуация характерна как для российских, так и зарубежных репозиторий), и это оценивается как неблагоприятный фактор, поскольку нет возможности отследить их существование и сделать оценку представленности разных типов РИД, то есть исследователи не имеют возможности подобрать наиболее подходящий репозиторий для размещения и хранения своих исследовательских данных.

Приведем ряд примеров репозиторий, не отмеченных в re3data.org, но обнаруженных нами.

Репозиторий психологических исследований и инструментов RusPsyDATA (<https://ruspsydata.figshare.com/>). Репозиторий, создателем и владельцем которого является Московский государственный психолого-педагогический университет (МГППУ), содержит **научные данные по психологии** (рис. 1), начал функционировать с июня 2021. В настоящее время не пополняется.

Электронный гербарий водорослей (<https://herbariumle.ru/?t=occ>) (рис. 2). Лаборатория альгологии Ботанического института им. В. Л. Комарова предлагает репозиторий образцов водорослей: он представляет собой специализированный инструмент с расширенной поисковой системой и использует специфичную схему метаданных.

Портал открытых данных Государственного геологического музея (<http://data.sgm.ru/dataset?q>) (рис. 3). В музее разработан портал, предоставляющий наборы данных в режиме онлайн. Любой научный сотрудник / преподаватель или студент вуза может загружать и использовать эти данные для исследовательских или образовательных целей.

Библиотека клеточных изображений (Калифорнийский университет, США) (<http://www.cellimagelibrary.org/pages/datasets>) позиционируется как исследовательский инструмент для анализа и получения новых результатов в клеточной биологии.

Среди репозиторий, содержащих спутниковые изображения, временные ряды данных о ветрах и температуре, историю погоды и пространственные климатические данные, можно отметить Глобальный атлас ветров (<https://globalwindatlas.info/ru>), созданный по совместной инициативе Технического университета Дании и Группы



Рис. 1. Пример наборов данных в репозитории МГППУ

Fig. 1. Example of datasets in the Moscow State University of Psychology and Education repository

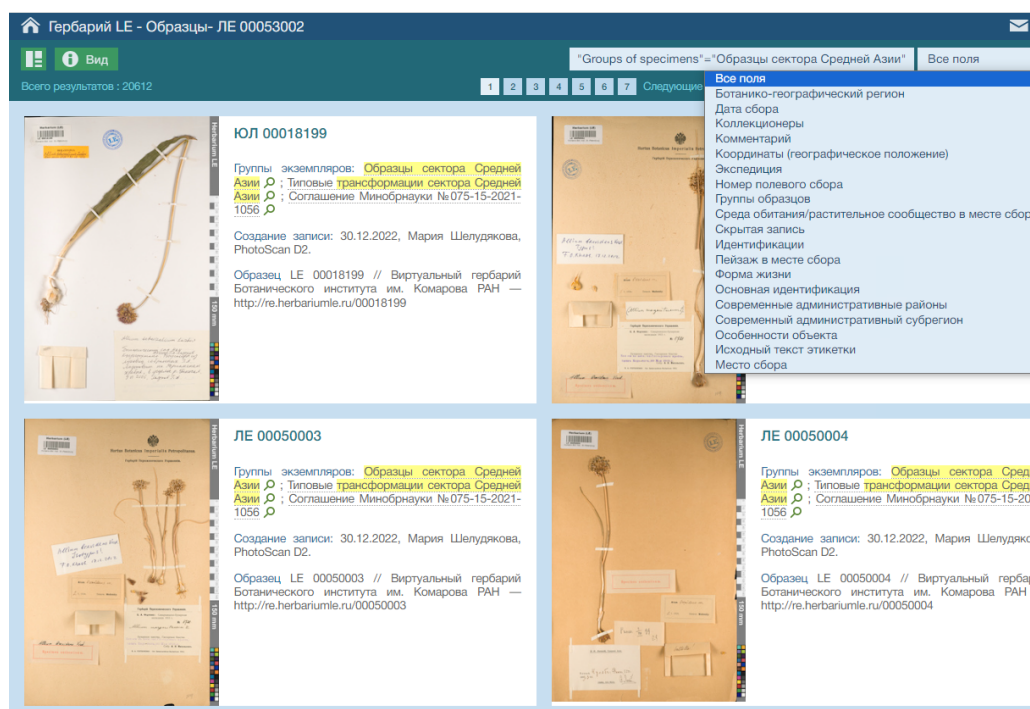


Рис. 2. Электронный гербарий водорослей

Fig. 2. Electronic herbarium of algae

Всемирного банка. Он обеспечивает доступ ко всем мировым данным о ветровых ресурсах.

База данных PISA (<https://www.oecd.org/pisa/data/>) содержит полный набор ответов учащихся, директоров школ и родителей, участвовавших в опросах о системе образования. Данные предлагаются профессиональным исследователям и статистикам для проведения собственного анализа.

Сведения о российских репозиториях исследовательских данных, а также международных порталах и коллекциях данных, в том числе стран ближнего зарубежья: Армении, Казахстана, Узбекистана и др., потенциально интересных русскоязычной аудитории, собраны в *Каталоге АНО «Инфокультура»* (<https://www.datacatalogs.ru>). Тематика данных представлена 109 рубриками; наборы данных сортируются

по категориям, программному обеспечению репозитория, региону, владельцу ресурса, стране, типу контента.

Представленный краткий обзор российских и зарубежных репозиторий исследовательских данных, так же как обзор публикаций, показывает, что для России работа по созданию репозиторий такого типа – достаточно новое явление. По тематическому охвату более всего создается репозиторий по естественным наукам. В области социально-гуманитарных наук их единицы, а по наукам библиотечного цикла они вообще отсутствуют, причем не только в российском сегменте, но и за рубежом (Gowen, Meier, 2020).

Принимая во внимание необходимость предоставления сообществу библиотечных специалистов открытого доступа к полученным

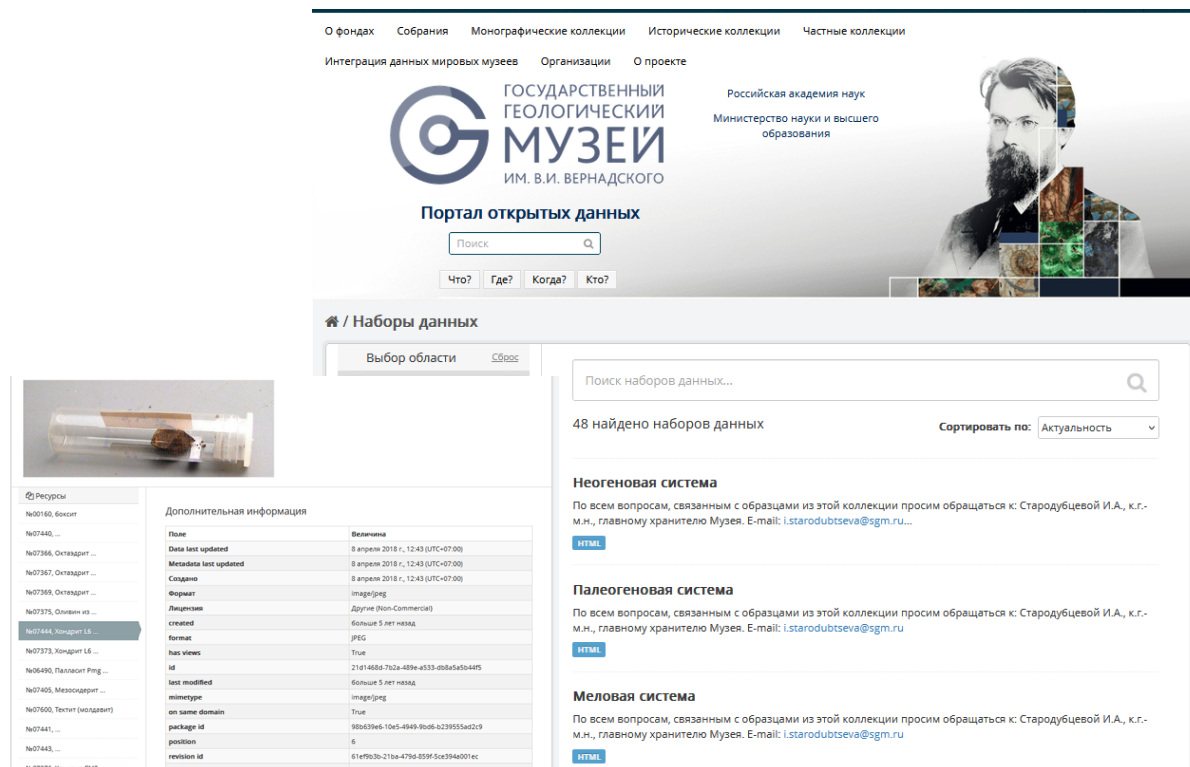


Рис. 3. Репозиторий Государственного геологического музея
Fig. 3. Repository of the State Geological Museum

в процессе исследовательской работы данным, а также возможности сохранять собственные данные, ГПНТБ СО РАН осуществила разработку репозитория исходных научных данных.

3. Репозиторий исходных исследовательских данных ГПНТБ СО РАН

В ГПНТБ СО РАН ведутся шесть проектов НИР. В начале работы над репозиторием мы ориентировались на исходные исследовательские данные по одному из них – «Современное состояние и тенденции развития коммуникаций российской науки с обществом». Цель: сделать видимыми и доступными в научном коммуникативном пространстве как первичные данные, так и исследовательские статьи.

При разработке репозитория мы придерживались четырех руководящих принципов, определенных для репозитория исследовательских данных Абдулом Азизом (Azeez, 2019):

1. Репликация – использование исходных данных, метаданных и статистических кодов для репликации процесса исследования.
2. Интеграция – объединение (установление связей) между набором необработанных данных и реализованным на его основе проектом и/или подготовленными публикациями внутри репозитория и за его пределами.
3. Открытый доступ – предоставление пользователям исходных исследовательских данных

в соответствии с принципами открытости при соблюдении юридических и интеллектуальных прав и правил конфиденциальности.

4. Управление – обеспечение постоянного доступа к данным и их долгосрочного использования.

Платформой для репозитория, предназначенного для сбора, хранения и последующего использования исследовательских данных, стала система автоматизации библиотек ИРБИС64. Выбор был обусловлен активным использованием и поддержанием этой системы в ГПНТБ СО РАН.

Репозиторий был создан в июле 2023 г. в рамках государственного проекта НИР № 122040600059-7. **На первом** этапе организации репозитория было разработано техническое задание на создание базы данных «Репозиторий исследовательских данных ГПНТБ СО РАН» (далее – БД «Репозиторий...») под управлением ИРБИС64 с набором необходимых полей для ввода метаданных (прил. 1, рис. 4). Сотрудники отдела автоматизированных систем сформировали БД «Репозиторий...» согласно техническому заданию и предоставили ее в служебное пользование администратору БД – ответственному за ввод информации.

При определении схемы метаданных мы ориентировались на структуру ИРБИС, основанную на обменном формате UNIMARK, а учитывая специфику объекта описания, добавили дополнительные поля.

Название элемента	№	Значение
1:Автор	1	Плешакова М.А. ГПНТБ СО РАН Pleshakova M.A. SPSTL SB RAS
	2	АКалужная Т.А. СПНТБ СО РАН EKalyuzhnaya T.A. GSPSTL SB RAS
2:Заглавие ресурса		АФорматы и тематика научно-популярных мероприятий в публичных библиотеках г. Новосибирска и пригорода BFormats and topics of popular
3:Аннотация/реферат		Представлены форматы и тематика научно-популярных мероприятий, проводимых в публичных библиотеках Новосибирска и пригорода. ГПНТ
31:Annotation/abstract		The formats and topics of popular science events held in public libraries of Novosibirsk and the suburbs are presented. GPNTB SB RAS, Koltsovskaya City
4:Ключевые слова	1	библиотечное мероприятие
	2	формат научно-популярных мероприятий
	3	тематика научно-популярных мероприятий
	4	публичная библиотека
	5	Новосибирск
41:Keywords	1	library event
	2	format of popular science events
	3	topics of popular science events
	4	public library
	5	Novosibirsk
5:Номер НИР/Гранта		122040600059-7
6:Дата/период сбора данных	1	A202303 B202304
8:Дата размещения в репозитории		20230601
9:Источник данных	1	информационные материалы на сайтах публичных библиотек г. Новосибирска
10:Процедура сбора данных		Сбор первичной информации осуществлялся методом сплошного просмотра материалов (планы и отчеты о деятельности библиотек, анонсы с
114:Ретроспектива		2016-2023 гг.
11:Ресурс	1	АТаблица Bxlsx Crus E14 л., 591 записи, 93 Кб
	2	Апрезентация Bpdf Crus E24 слайда, 792 Кб
	3	Авидео Bmp4 Crus
Основная / Систематизация / Технология / Добавочные		
Название элемента	№	Значение
964:Индекс ГРНТИ	1	13.31.23
16:DOI		
186:Название конференции и дата	1	АНаука, технологии и информация в библиотеках (LIBWAY - 2023) : Международная научно-практическая конференция, посвященная 105-летию
951:Ссылка-внешний объект	1	http://www.spsl.nsc.ru/fulltext/repoziyoriy/01/Форматы мероприятий.xlsx ТФорматы научно-популярных мероприятий (таблица)
952:Связанные материалы	1	http://www.spsl.nsc.ru/fulltext/repoziyoriy/01/Коммуникативный ресурс мероприятия.pdf Коммуникативный ресурс библиотеки
	2	https://www.youtube.com/watch?v=obzfNVD7NXto&t=38m33s ТБиблиотечные мероприятия (видео) B22.03.2023 E1
9511:Ссылка на публикации	1	ТПлешакова М.А., Калужная Т.А. Коммуникативный ресурс библиотечного мероприятия в рамках коммуникации науки и общества // Библиоте
19:Создатель ресурса		ГПНТБ СО РАН
997: Лицензия		АCC BY Bhttp://www.spsl.nsc.ru/fulltext/repoziyoriy/license/1cc by.png Chttps://creativecommons.org/licenses/by/4.0/legalcode.ru

Рис. 4. БД «Репозиторий исследовательских данных ГПНТБ СО РАН» (APM администратора)

Fig. 4. Database «Repository of research data of SPSTL SB RAS» (Administrator's office)

В качестве источника данных (поле «Источник данных»), которые используются при сборе и анализе данных, могут быть сайт, документ, опрос, эксперимент, полевое исследование, видеозапись, авторская презентация и др.; типы ресурсов (поле «Ресурс»), включаемые в репозиторий: текст, таблица, рисунок, график, фотография, презентация, видео, карта, координаты, база данных и др. – с указанием вида, формата, объема ресурса.

На **втором этапе** проводилось наполнение БД «Репозиторий...» исходными данными по теме НИР «Современное состояние и тенденции развития коммуникаций российской науки с обществом». Это данные анкетирования, перечни объектов исследования, статистические выборки, информационные массивы, объектно-характеристические таблицы и т. д.; связанные с исходными данными материалы: презентации² докладов на конференциях и семинарах; видео-выступления участников НИР на научных мероприятиях и др.; ссылки на публикации.

В поле «Ссылка – внешний объект» прописываются ссылки на исходные наборы данных, которые хранятся на сервере библиотеки, с указанием адреса размещения ресурса, заглавия и вида ресурса.

В поле «Связанные материалы» даются ссылки на презентации или видеоматериал,

с указанием адреса размещения, заглавия и вида ресурса, а также времени начала выступления участников НИР (для видео).

Поле «Ссылка на публикации» для взаимосвязи исходных исследовательских данных со статьями, где эти уже обработанные, интерпретированные данные зафиксированы, заполняется библиографическим описанием статей и адресом удаленного ресурса, где эти публикации размещены (например, РИНЦ, сайт научного журнала и т. п.).

Все полученные от участников проекта исходные данные адаптируются под единый формат ввода:

- исходя из контекста данных определяется заглавие ресурса, составляется аннотация и выделяются ключевые слова;
- исходные данные, представленные в виде таблиц, текстовых файлов, презентаций, видео-файлов, обрабатываются: текстовые документы и презентации приводятся к формату *.pdf; для видеофайлов фиксируется время начала выступления; группа файлов одинакового формата (например, *.xlsx) архивируется (*.zip) (например, статистические таблицы выдачи журналов за разные годы по определенной социальной-демографической характеристике).

Третий этап (заключительный) – создание репозитория – это предоставление информационной системы в открытый доступ научному сообществу через размещение БД

² Презентации являются фактом апробации результатов исследования.

«Репозиторий...» на сайте ГПНТБ СО РАН в разделе «Ресурсы» (рис. 5). С помощью программы веб-ИРБИС БД была выставлена в интернет, а также сформирована страница входа в репозиторий на сайте ГПНТБ СО РАН.

Итак, каждая запись содержит все необходимые метаданные для идентификации и нахождения требуемых исходных данных по тематическому запросу. В каждой записи указываются авторы, даются заглавие, аннотация и ключевые слова ресурса, в том числе на английском языке, приводится период проведения сбора данных и описание процедуры сбора, показываются источники, которые использовались при сборе и анализе, а также прописываются ссылки на полнотекстовые ресурсы, содержащие все исходные данные, и привязанные к ним материалы (видео и презентации). Поиск в репозитории возможен по следующим полям: автор, заглавие ресурса, аннотация, ключевые слова, индекс ГРНТИ, вид ресурса, источник данных, процедура сбора данных, конференция (место и дата проведения, название) (рис. 5).

С точки зрения защиты авторского права все наборы данных в репозитории ГПНТБ СО РАН имеют такую степень доступности, которую определил правообладатель, и находятся под лицензией Creative Commons CC BY (рис. 5).

Сегодня пополнение репозитория ГПНТБ СО РАН осуществляется в текущем режиме администратором по мере получения новых исходных

данных от участников проекта № 122040600059-7. В дальнейшем предполагается, что:

- наполнение репозитория будет вестись ответственным участником от каждой малой группы проекта;
- в репозиторий будут включаться наборы исходных данных по другим проектам НИР ГПНТБ СО РАН и сторонних организаций, физических лиц;
- будет разработан новый интерфейс (сейчас используется интерфейс веб-ИРБИС);
- добавится опция автоматического формирования ссылки для цитирования набора данных (прил. 2);

Таким образом, при создании репозитория исходных исследовательских данных в ГПНТБ СО РАН были соблюдены все требования, предъявляемые к РИД.

Заключение

Можно констатировать, что для России создание репозитория исходных исследовательских данных – это относительно новое явление. При этом потребность в таких хранилищах очевидна и обоснована рядом факторов. Прежде всего, характерным для цифровой эры геометрически прогрессирующим ростом объемов исходных данных и одновременно высоким риском их потери. Кроме того, вероятной необходимостью в оценке рецензентами корректности полученных

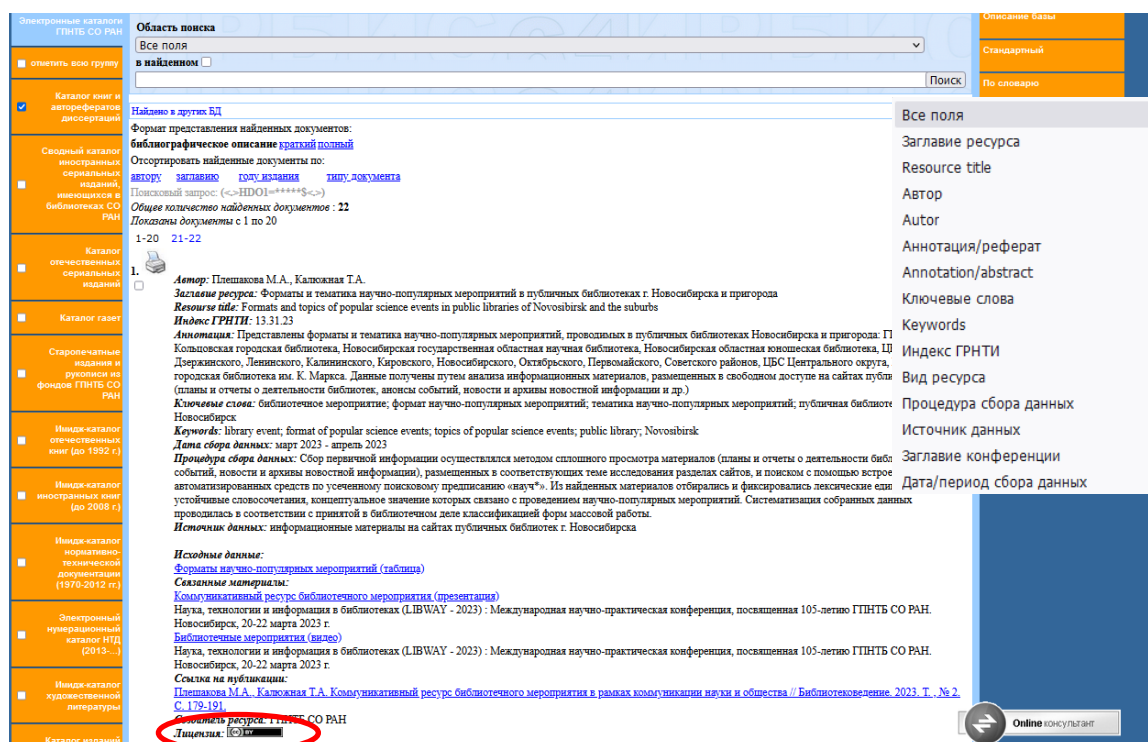


Рис. 5. Репозиторий исследовательских данных ГПНТБ СО РАН
(<https://www.spsl.nsc.ru/repozitorij-gpntb-so-ran>)

Fig. 5. Repository of research data of SPSTL SB RAS

в исследованиях результатов через обращение непосредственно к исходным данным. И наконец, возможностями обмена, повторного использования и объединения однотипных данных для проведения других исследований.

Обозначенные идеи стали предпосылками разработки репозитория ГПНТБ СО РАН, который изначально задумывался как хранилище исходных данных по одному из проектов НИР, реализуемых в библиотеке, однако его создатели видят более широкие перспективы. Поскольку ГПНТБ СО РАН является одним из ведущих научно-исследовательских институтов в области библиотекосведения, книговедения, библиографоведения и прикладной информатики, в рамках проводимой в ее стенах работы издаются четыре научных журнала³. Существующий репозиторий может стать хранилищем исходных данных в том числе к публикуемым в этих изданиях статьям. В мировой практике это уже стало нормой. Например, в соответствии с политикой использования данных Springer Nature⁴ рекомендуется, чтобы все наборы данных, подтверждающие анализ и выводы исследования, были общедоступными на момент публикации. В нашей стране такой опыт пока не распространен, тем не менее редакции научных журналов в качестве рекомендации могут предлагать авторам размещать наборы исходных данных в репозитории и давать на него ссылку при подаче статьи. При этом для рецензентов в личных кабинетах должна быть предусмотрена возможность перехода по указанной ссылке для оценки рукописи с точки зрения корректности использования данных. Репозиторий, в свою очередь, должен предлагать функцию размещения наборов данных без их немедленного публичного доступа. То есть исследователи смогут депонировать данные до тех пор, пока их рукопись не будет принята или опубликована.

Сегодня ссылка в научной статье на размещенные в репозитории исходные данные для их доступности научному сообществу (помимо

конечных результатов исследования) становится этическим требованием научных коммуникаций (Мжельский, 2022). Помимо этого, архивирование исследовательских данных в репозитории, при условии наличия идентификаторов, обеспечивает правообладателю такое же цитирование, как и для публикаций.

Статья подготовлена по плану НИР ГПНТБ СО РАН, проект «Современное состояние и тенденции развития коммуникаций российской науки с обществом», № 122040600059-7

Авторы прочитали и одобрили окончательный вариант рукописи.

Вклад авторов

Калужная Т. А.: обсуждение темы исследования; написание текста рукописи; обзор отечественных литературных источников; оформление метаданных и статьи в соответствии с требованиями журнала «Библиосфера».

Плешакова М. А.: обзор зарубежных литературных источников; научное редактирование текста рукописи.

Конфликт интересов

Калужная Т. А. является ответственным секретарем журнала «Библиосфера», но не имеет никакого отношения к решению редколлегии опубликовать эту статью. Статья прошла принятую в журнале процедуру рецензирования. Об иных конфликтах интересов авторы не заявляли.

Приглашаем авторов для размещения исходных исследовательских данных в репозитории ГПНТБ СО РАН (<http://www.spsl.nsc.ru/repozitorij-gpntb-so-ran/>).

Благодарность: выражаем признательность сотруднику отдела автоматизированных систем ГПНТБ СО РАН В. И. Москвину за создание БД «Репозиторий...».

Список источников / References

Беглер А. М., Ануфриев Г. В. Формальная модель переменных как часть инфраструктуры для обеспечения принципов FAIR в поведенческих экспериментах // Когнитивная наука в Москве: новые исследования : материалы конф., Москва, 23–24 июня 2021 г. Москва, 2021. С. 484–489 [Begler AM and Anufriev GV (2021) A formal model of variables as part of the infrastructure for

providing FAIR principles in behavioral experiments. *Kognitivnaya nauka v Moskve: novye issledovaniya: materialy konf.*, Moskva, 23–24 iyunya 2021 g. Moscow, pp. 484–489. (In Russ.)].

Белякова О. А., Махнин П. Н., Сапегин С. В. Платформа для сбора, хранения и анализа мультимодальных данных, генерируемых субъектами онлайн коммуникаций // Мониторинг. Наука и технологии. 2019. № 4. С. 73–84 [Belyakova OA, Makhnin PN and 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. (In Russ.)]. DOI: <https://doi.org/10.25714/MNT.2019.42.011>.

³ <https://www.spsl.nsc.ru/nauchnaya-rabota>

⁴ <https://www.springernature.com/gp/authors/research-data-policy>

- Борбат А. М., Лищук С. В. Первый российский набор данных гистологических изображений патологических процессов молочной железы // Врач и информационные технологии. 2020. № 3. С. 25–30 [Borbat AM and Lishchuk SV (2020) The first Russian breast pathology histological images data set. *Vrach i informatsionnye tekhnologii* 3: 25–30. (In Russ.)].
- Бычкова Е. Ф., Колосов К. А. Репозитории исследовательских данных в области экологии и наук об окружающей среде // Научно-исследовательская работа в библиотечной теории и практике : ежегод. межведомств. сб. науч. тр. Москва, 2022. С. 19–29 [Bychkova EF and 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. (In Russ.)]. DOI: <https://doi.org/10.33186/978-5-85638-241-8-2022-19-29>.
- Гавриков О. В. Электронные библиотеки и репозитории в исследовательской деятельности // Инновационные подходы к организации проектно-исследовательской деятельности. Пермь, 2021. С. 141–155 [Gavrikov OV (2021) Electronic libraries and repositories in research activities. *Innovatsionnye podkhody k organizatsii proektno-issledovatel'skoi deyatel'nosti*. Perm, pp. 141–155. (In Russ.)].
- Гончаров М. В., Колосов К. А. Репозитории исследовательских данных как новое направление развития открытых архивов // Научно-исследовательская работа в библиотечной теории и практике : ежегод. межведомств. сб. науч. тр. Москва, 2022. С. 35–48 [Goncharov MV and 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. (In Russ.)].
- Засурский И. И., Соколова Д. В., Трищенко Н. Д. Репозитории открытого доступа: функции и тенденции развития // Научные и технические библиотеки. 2020. № 9. С. 121–142 [Zassoursky II, Sokolova DV and Trishchenko ND (2020) Open access repositories: functions and development trends. *Nauchnye i tekhnicheskie biblioteki* 9: 121–142. (In Russ.)]. DOI: <https://doi.org/10.33186/1027-3689-2020-9-121-142>.
- Косычева М. А., Хорохорина Г. А. Репозитории данных: теория и практика // Health, Food & Biotechnology. 2020. Т. 2, № 2. С. 7–11 [Kosycheva MA and Khorokhorina GA (2020) Data repositories: theory and practice. *Health, Food & Biotechnology* 2 (2): 7–11. (In Russ.)]. DOI: <https://doi.org/10.36107/hfb.2020.i2.s393>.
- Куликов Е. С., Федорова О. С., Толмачев И. В., Рязанцева У. В., Вражнов Д. А., Губанов А. В., Нестерович С. В., Шмырина А. А. Русскоязычный репозиторий открытых клинических данных SibMED Data Clinical Repository // Бюллетень сибирской медицины. 2023. Т. 22, № 2. С. 182–184 [Kulikov ES, Fedorova OS, Tolmachev IV, Ryazantseva UV, Vrazhnov DA, Gubanov AV, Nesterovich SV and Shmyrina AA (2023) Russian-language repository of the open clinical data “SibMED Data Clinical Repository”. *Byulleten' sibirskoi meditsiny* 22 (2): 182–184. (In Russ.)]. DOI: <https://doi.org/10.20538/1682-0363-2023-2-182-184>.
- Мжельский А. А. Воспроизводимость исследований и доступ к их исходным данным. Интерпретация международных рекомендаций и обзор лучших журнальных политик и практик // Научный редактор и издатель. 2022. Т. 7, № 2. С. 143–165 [Mzhelsky AA (2022) Data and reproducibility. Interpretation of international guidelines and journals' best policies and practices. *Nauchnyi redaktor i izdatel'* 7 (2): 143–165. (In Russ.)]. DOI: <https://doi.org/10.24069/SEP-22-53>.
- Открытая наука: цифровые сервисы и ресурсы // Университетская книга. 2020. № 8. С. 29–35 [(2020) Open Science: digital services and resources. *Universitetskaya kniga* 8: 29–35. (In Russ.)].
- Политика обмена исследовательскими данными (Data Sharing Policy) / сост. М. М. Зельдина. Москва : Ваше цифровое изд-во, 2019. 20 с. [Zeldina MM (comp.) (2019) The policy of the exchange of research data (Data Sharing Policy). Moscow: Vashe tsifrovoe izd-vo. (In Russ.)].
- Редькина Н. С. Подготовка библиотекарей в области управления исследовательскими данными // Ученые записки (Алтайская государственная академия культуры и искусств). 2019. № 3. С. 83–86 [Redkina NS (2019) Teaching of librarians' skills in research data management. *Uchenye zapiski (Altayskaya gosudarstvennaya akademiya kul'tury i iskusstv)* 3: 83–86. (In Russ.)]. DOI: <https://doi.org/10.32340/2414-9101-2019-3-83-86>.
- Рычихин А. К., Нуриев В. А. Организация жизненного цикла исследовательских данных // Системы и средства информатики. 2020. Т. 30, № 3. С. 81–96 [Rychikhin AK and Nuriev VA (2020) Organization of the life cycle of research data. *Sistemy i sredstva informatiki* 30 (3): 81–96. (In Russ.)].
- Стукалова А. А. Институциональные репозитории российских организаций: сравнительный анализ // Библиотековедение. 2023. Т. 72, № 4. С. 319–331 [Stukalova AA (2023) Institutional repositories of Russian organizations: a comparative analysis. *Bibliotekovedenie* 72 (4): 319–331. (In Russ.)]. DOI: <https://doi.org/10.25281/0869-608X-2023-72-4-319-331>.
- Чадин И. Ф. Zenodo и GBIF: инструменты для публикации наборов первичных данных // Вестник Института биологии Коми научного центра Уральского отделения РАН. 2018. № 3. С. 34–36 [Chadin IF (2018) Zenodo and GBIF: tools for publishing primary data sets. *Vestnik Instituta biologii Komi nauchnogo tsentra Ural'skogo otdeleniya RAN* 3: 34–36. (In Russ.)]. DOI: [https://doi.org/10.31140/j.vestnikib.2018.3\(205\).5](https://doi.org/10.31140/j.vestnikib.2018.3(205).5).
- Щекотилов В. Г., Лазарев О. Е., Лазарева О. С., Щекотилова С. Н. Репозиторий архивных карт

- и геокодируемых данных: целеполагание, аудитория, структура и функциональность // Историческая информатика. 2019. № 1. С. 152–159 [Shchekotilov VG, Lazarev OE, Lazareva OS and Shchekotilova SN (2019) Repository of archival maps and geocoded data: goal setting, audience, structure and functionality. *Istoricheskaya informatika* 1: 152–159. (In Russ.)]. DOI: <https://doi.org/10.7256/2585-7797.2019.1.29062>.
- Юдина И. Г., Федотова О. А. Репозитории научных публикаций открытого доступа: история и перспективы развития // Информационное общество. 2020. № 6. С. 67–79 [Yudina IG and Fedotova OA (2020) Repositories of open access scientific publications: history and prospects of development. *Informatsionnoe obshchestvo* 6: 67–79. (In Russ.)].
- Юмашева С. И. Медицинские репозитории открытого доступа: состояние и тенденции развития // Библиосфера. 2023. № 3. С. 83–95 [Yumasheva SI (2023) Open access medical repositories: status and development trends. *Bibliosfera* 3: 83–95. (In Russ.)]. <https://doi.org/10.20913/1815-3186-2023-2-83-95>.
- Abdul Azeez TA (2019) Open research data repositories: a librarian's perspective. *KELPRO Silver Jubilee Souvenir*. Kerala Libr. Professionals' Organization: 47–54.
- Gowen E. and 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. DOI: <https://doi.org/10.7710/2162-3309.2336>.
- Minarso C and Suprpto A (2021) Kompetensi data librarian dalam research data management: systematic literature review. *BACA: Jurnal Dokumentasi dan Informasi* 42 (1): 1–17.
- Newton MP, Miller CC and Bracke MS (2010) Librarian roles in institutional repository data set collecting: outcomes of a research library task force. *Collection Management* 36 (1): 53–67. DOI: <https://doi.org/10.1080/01462679.2011.530546>.
- Singh P (2016) Open access repositories in India: characteristics and future potential. *IFLA Journal* 42 (1): 16–24. DOI: <https://doi.org/10.1177/0340035215610131>.
- Uzwysyn RJ (2016) Research data repositories: the what, when, why and how. *Computers in Libraries* 36 (3): 18–21.
- Uzwysyn RJ (2022) Research data repositories and global scholarly ecosystem possibilities. *ResearchGate: social network*. URL: https://www.researchgate.net/publication/357286044_Research_Data_Repositories_and_Global_Scholarly_Ecosystem_Possibilities (accessed 02.04.2024).

Техническое задание на создание репозитория исследовательских данных ГПНТБ СО РАН

1. **Цель проекта:** создание репозитория ГПНТБ СО РАН (с постепенным формированием репозитория СО РАН) для хранения и использования исходных данных исследований, а также материалов конференций (видеовыступлений и презентаций докладов). Его подготовка обусловлена необходимостью размещения, хранения и обеспечения доступа к большим исходным массивам данных для получения грантов, а также верификации полученных выводов в научных публикациях экспертами и рецензентами.

2. **Задачи проекта:** формирование БД «Репозиторий исследовательских данных ГПНТБ СО РАН» (далее – «Репозиторий...»).

2.1. Создать БД «Репозиторий...» в системе автоматизации библиотек ИРБИС, включив следующие поля и подполя (см. табл.), для формирования поискового образа ресурса и поискового предписания.

2.2. Подготовить выходные форматы.

2.3. Предоставить доступ к БД «Репозиторий...» на сайте ГПНТБ СО РАН через веб-ИРБИС.

2.4. Рассмотреть вопрос об интеграции репозитория ГПНТБ СО РАН в систему НОРА и разработать переходную схему.

Таблица

Поля и подполя для БД «Репозиторий...»

Поля	Подполя	Словарь
Автор	Фамилия	
	Инициалы	
	Аффилиация	
Заглавие ресурса		
Аннотация / реферат		
Ключевые слова		
Номер НИР / Гранта		
Дата / период сбора данных		
Дата размещения в репозитории		
Источник данных		сайт, документ, опрос, эксперимент, полевое исследование, видеозапись, авторская презентация и др.
Процедура сбора данных		
Ретроспектива		
Ресурс	Вид	текст, таблица, рисунок, график, фотография, презентация, видео, карта, координаты, база данных
	Формат	doc/docx, xls/xlsx, pdf, jpg, ppt/pptx, mp4, mpeg, zip и др.
	Язык	
	Объём	стр., л., файл, слайд, байт
Индекс ГРНТИ		
DOI		
Название конференции, дата	Название конференции	
	Город	
	Дата конференции	
Ссылка – внешний объект		
Связанные материалы		
Создатель ресурса		

Приложение 2

Пример цитирования наборов данных

1. Если набор данных имеет DOI:

Hao, Z., AghaKouchak, A., Nakhjiri, N., Farahmand, A. (2014) Global Integrated Drought Monitoring and Prediction System (GIDMaPS). Data sets. figshare. <https://doi.org/10.6084/m9.figshare.853801>

2. Если данные находятся в открытом доступе репозитория, не выдающем DOI:

Калюжная Т. А., Плешакова М. А., Баженов С. Р., Паршиков Р. М. Статистика выдачи научно-популярных журналов из фонда ГПНТБ СО РАН (01.01.2023–31.08.2023). Наборы данных. Репозиторий исследовательских данных ГПНТБ СО РАН. 2023. Запись 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=*****