



DOI: 10.19181/smtp.2024.6.4.7

EDN: TNIPPN

Научная статья

Research article

РЕПОЗИТОРИИ ИССЛЕДОВАТЕЛЬСКИХ ДАННЫХ В РОССИИ И ЗА РУБЕЖОМ



**Стукалова
Анна Александровна¹**

¹ Государственная публичная научно-техническая библиотека
СО РАН, Новосибирск, Россия

Для цитирования: Стукалова А. А. Репозитории исследовательских данных в России и за рубежом // Управление наукой: теория и практика. 2024. Т. 6, № 4. С. 122–142. DOI 10.19181/smtp.2024.6.4.7. EDN TNIPPN.

Аннотация. Одним из эффективных инструментов для обеспечения открытого доступа к результатам научных исследования являются репозитории различных типов. Наибольшее распространение получили тематические, национальные, исследовательские и институциональные репозитории, включающие, как правило, текстовые материалы: монографии, научные публикации, труды конференций и т. п. Помимо вышеперечисленных типов, в последние годы за рубежом активно развиваются тематические и универсальные репозитории исследовательских данных. Кроме текстовой информации, они включают числовые, аудиовизуальные данные, наборы данных, модели, компьютерные коды и т. п.

С целью оценки текущего состояния репозиторий данных в нашей стране было проведено исследование сайтов научных организаций трёх категорий, утверждённых Федеральным агентством научных организаций (ФАНО), согласно постановлению Правительства РФ от 8 апреля 2009 г. № 312 «Об оценке и о мониторинге результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения», и образовательных учреждений – участников программы «Приоритет 2030».

В результате проделанной работы было выявлено только три репозитория данных образовательных учреждений, контент которых в основном закрыт для удалённого пользователя. Среди научных организаций была найдена информация об исследовательских данных на двадцати сайтах. Однако большинство этих сведений представлено в виде перечня данных, без предоставления поиска этой информации по различным признакам и возможности управления их результатами. В то же время найдены успешные примеры функционирования отечественных репозиторий данных, включающие уникальные коллекции научных исследований и обеспечивающие открытый доступ к их результатам. Они предоставляют широкие поисковые инструменты для удалённого пользователя: возможность применения логических операторов, фасетной навигации, сортировки результатов поиска.

Выявлено, что отечественные репозитории данных практически не отражены в международных агрегаторах, а существующая информация о них зачастую некорректна. Это негативно влияет на видимость результатов исследований среди мирового научного сообщества.

В то же время размещение исследовательских данных в репозиториях, предоставление репозиториями данных в международных агрегаторах способствуют видимости, доступности и быстрому распространению результатов исследований. Это исключает необходимость повторного сбора данных, позволяет избежать дублирования работ.

Ключевые слова: репозитории, исследовательские данные, открытый доступ, агрегатор репозиториями, метаданные, поисковые возможности

Благодарности. Статья подготовлена по плану научно-исследовательской работы Государственной публичной научно-технической библиотеки Сибирского отделения Российской академии наук в рамках научного проекта № 122041100150-3 «Разработка модели функционирования научной библиотеки в информационной экосистеме открытой науки».

REPOSITORIES OF RESEARCH DATA IN RUSSIA AND ABROAD

Anna A. Stukalova¹

¹ State Public Scientific Technological Library, SB RAS, Novosibirsk, Russia

For citation: Stukalova A. A. Repositories of research data in Russia and abroad. *Science Management: Theory and Practice*. 2024;6(4):122–142. (In Russ.). DOI 10.19181/smtp.2024.6.4.7.

Abstract. One of the effective tools for providing open access to academic research results are repositories of various types. The most common are thematic, national, research and institutional repositories that usually contain text materials: monographs, scientific publications, conference proceedings, etc.

In addition to the types mentioned above, thematic and universal research data repositories have been actively developed abroad in recent years. Besides text information, they also include numerical, audiovisual data, datasets, models, computer codes, etc.

In order to assess the current state of data repositories in our country, we have conducted a study of websites of scientific organizations of the three categories that were approved by the Federal Agency for Scientific Organizations, in accordance with the Decree of the Government of the Russian Federation of April 8, 2009 No. 312 “On the Assessment and Monitoring of the Effectiveness of Scientific Organizations Performing Research, Development and Technological Work for Civil Purposes” as well as of educational institutions that are participants of the program “Priority 2030”.

This work has identified only three educational institution’s data repositories, the content of which is mostly closed for a remote user. Among scientific institutions, information on research data was found on 20 websites. However, most of this information is presented as a list of data, without providing search options by various attributes and managing their results. In the meantime, we have found successful examples of functioning of Russian data repositories that include unique collections of scientific research and provide open access to the research results. They provide extensive search features for a remote user: possibilities of applying logical operators and faceted navigation, as well as sorting of search results.

It has been found out that national data repositories are practically not covered in international aggregators, and the existing information about them is often incorrect. This has a negative impact on the visibility of research results within the global scientific community. At the same time, placing research data in repositories and making data repositories available in international aggregators promote visibility, accessibility and rapid distribution of research results. This eliminates the need for re-collection of data and prevents duplication of works.

Keywords: repositories, research data, open access, repository aggregator, metadata, search capabilities

Acknowledgements. The article was prepared according to the research plan of the State Public Scientific Technological Library of the Siberian Branch of the Russian Academy of Sciences, project “Development of a Model for the Functioning of a Scientific Library in the Information Ecosystem of Open Science”, No. 122041100150-3.

Важнейшими платформами сбора, хранения и обеспечения открытого доступа (ОД) к результатам научных исследований являются репозитории различных типов.

Согласно п. 3.1.21 ГОСТ Р 57723–2017 репозиторий – это «место, где хранятся и поддерживаются какие-либо данные вместе с историей их изменения и другой служебной информацией»¹. Они обеспечивают доступность, видимость и быстрое распространение результатов исследований, способствуют повышению цитируемости предоставленных материалов.

В зависимости от контента репозитория, способов их организации различают четыре основных типа: **тематические** (используются для публикации статей исследователей из разных организаций по одной или смежным дисциплинам), **исследовательские** (создаются и спонсируются, как правило, для прямой поддержки исследований в какой-либо области), **национальные** (создаются для развития разного рода исследований в государственном масштабе), **институциональные** (создаются в рамках конкретного научного центра или научного объединения) [1, с. 11–12]. Перечисленные типы репозитория, как правило, включают текстовые материалы: монографии, научные статьи, авторефераты и диссертации, труды конференций и т. п. Особенности их организации, проблемы функционирования подробно освещены в зарубежных [2; 3; 4] и отечественных [5; 6; 7; 8] публикациях.

Помимо вышеперечисленных типов, в последние годы получили развитие репозитории данных.

Репозитории исследовательских данных – это крупные инфраструктуры баз данных (БД), созданные для управления, совместного использования, доступа и архивирования наборов данных исследователей [9, с. 7].

Репозитории данных предоставляют более широкие наборы данных, включающие количественную или качественную информацию, собранную учёными в ходе проведённого ими исследования. В отличие от архивов, предоставляющих текстовые материалы, они могут включать числовые (таблицы, подсчёты,

¹ ГОСТ Р 57723–2017. Информационно-коммуникационные системы в образовании. Системы электронно-библиотечные: общие положения. М. : Стандартинформ, 2017. С. 4. URL: https://cchgeu.ru/upload/staff/bibl/СИБИД/gost_r_57723-2017.pdf (дата обращения: 02.05.2024).

измерения); аудиовизуальные данные (изображения, звукозаписи, видео); модели, компьютерные коды; специальные дисциплинарные данные (FITS² в астрономии, CIF³ в химии); специальные данные для прибора, конкретного оборудования. Исследовательские данные могут быть получены в результате экспериментов, наблюдений, моделирования, анкетирования, опросов или интервью или созданы на основе уже существующего опыта. На них строится или тестируется доказательство/теория, осуществляется исследовательский процесс [10].

Наиболее активное развитие репозитории данных получили в зарубежных странах. Опыт их создания и функционирования отражён в ряде публикаций [11; 12; 13; 14; 15]. Существует большое количество примеров тематических и универсальных репозиторий данных (табл. 1).

Таблица 1

Примеры зарубежных репозиторий исследовательских данных

Table 1

Examples of foreign research data repositories

Тематика	Название репозитория	Ссылка на репозиторий
Биология	INSDC repositories	https://insdc.org
	The Genome Sequence Archive	https://ngdc.cncb.ac.cn/gsa
Медицина	The Cancer Imaging Archive	https://cancerimagingarchive.net
	The National Database for Clinical Trials related to Mental Illness	https://nda.nih.gov
Химия	ioChem-BD Computational Chemistry Datasets	https://iochem-bd.org
Науки о земле	The NASA Goddard Earth Sciences Data and Information Services Center	https://catalog.data.gov/dataset/goddard-earth-sciences-data-and-information-services-center-ges-disc
	National Tibetan Plateau/Third Pole Environment Data Center	https://data.tpdac.ac.cn/en/?MetaDataId=187
Астрономия	SIMBAD Astronomical Database	https://simbad.u-strasbg.fr/simbad
	UK Solar System Data Centre	https://www.ukssdc.ac.uk/
Общественные науки	Archaeology Data Service	https://archaeologydataservice.ac.uk
	Harvard Dataverse	https://dataverse.harvard.edu
	The Qualitative Data Repository	https://qdr.syr.edu
Универсальные	Dryad Digital Repository	https://datadryad.org
	Figshare	https://figshare.com
	Harvard Dataverse	https://dataverse.harvard.edu
	Open Science Framework	https://osf.io
	Zenodo	https://zenodo.org/
	B2Share	https://b2share.eudat.eu
	Science Data Bank	https://scidb.cn/en

² FITS (Flexible Image Transport System) – цифровой формат файлов, используемый в науке для хранения, передачи и редактирования изображений и их метаданных (электронных таблиц).

³ CIF (Crystallographic Information File) – стандартный текстовый формат файла, который используется для описания кристаллографической информации.

Например, The Qualitative Data Repository – это архив для хранения и обмена цифровыми данными по общественным наукам (рис. 1). Он включает аудиовизуальные, научные и статистические данные, изображения, текстовые документы. Репозиторий предоставляет пользователю простой поиск по ключевым словам с возможностью его дальнейшего уточнения с помощью фасетной навигации по годам, ФИО ответственных лиц, предметам, ключевым словам, странам. Возможно использование булевых операторов, поиск точной фразы. Также предоставляется расширенный вариант поиска по большому количеству признаков: названию, типу файлов, языку, ключевым словам, ответственным лицам, географическому охвату, словам из полного текста и т. д. На карточках объектов репозитория размещена информация о датах загрузки, размере и типе файла, количестве загрузок, краткое описание.



Рис. 1. Сайт The Qualitative Data Repository
Fig. 1. The Qualitative Data Repository website

Если в тематических хранилищах сосредоточены данные, связанные с определённой академической дисциплиной, то универсальные репозитории могут принимать данные от всех исследователей, независимо от их направления деятельности, местонахождения или источника финансирования.

Среди универсальных репозиториев наиболее популярным является Zenodo, разработанный в рамках европейской программы OpenAIRE и управляемый Европейской организацией по ядерным исследованиям (рис. 2).

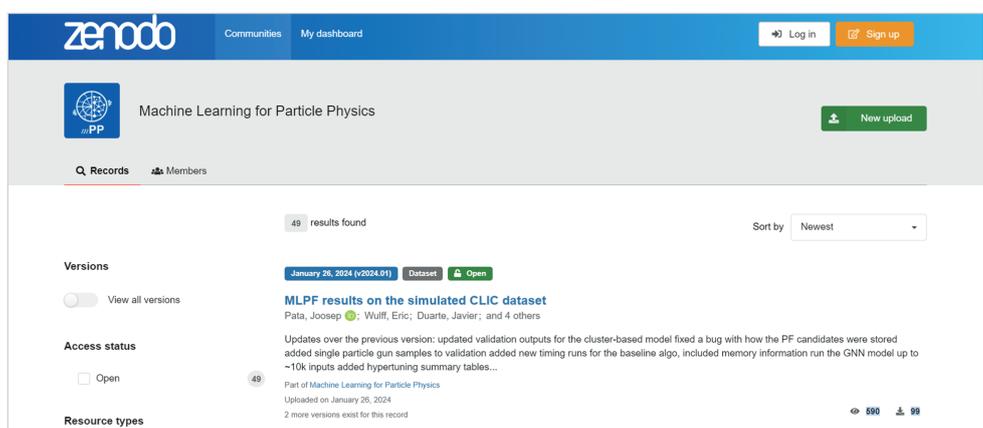


Рис. 2. Сайт Zenodo
Fig. 2. Zenodo website

Zenodo позволяет исследователям обмениваться и сохранять результаты своих исследований независимо от их тематики, размера и формата. Репозиторий включает публикации, изображения, наборы данных, программы, презентации, постеры, видео, лекции, физические объекты. На 3 апреля 2024 г. он включает более 3 500 000 объектов. Им автоматически присваиваются DOI. Поиск возможен по ключевым словам с применением фасетной навигации по типу публикации, формату, предмету, условиям доступа. Карточка объекта, помимо основных метаданных, включает сведения о количестве загрузок и просмотров, ссылки на подобные статьи.

Результаты поиска возможно отсортировать по релевантности, цитируемости, альтметрическим показателям, прямой и обратной хронологии, дате онлайн-публикации и т. д.

Благодаря бесплатным загрузке и доступу Zenodo делает научные результаты исследований видимыми и общедоступными.

Не менее известным хранилищем данных является крупнейшая в мире технология хранения и индексации научных данных – Figshare (рис. 3).

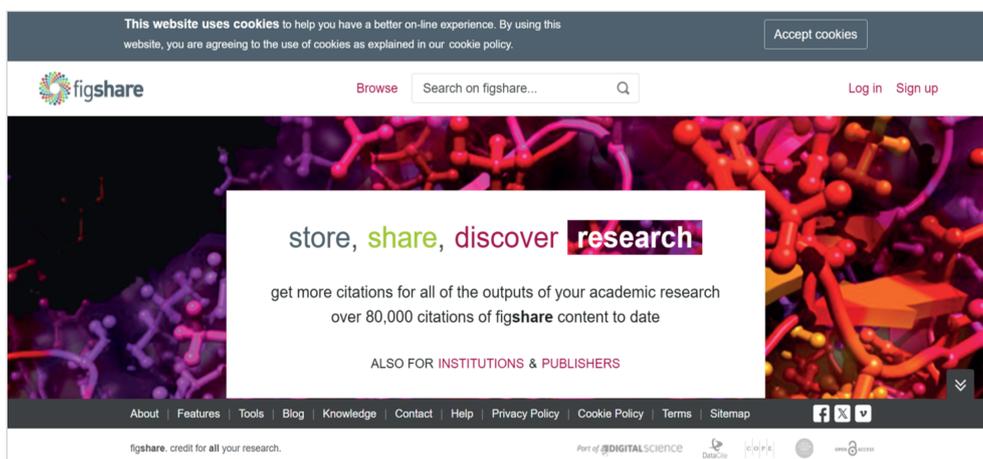


Рис. 3. Сайт Figshare
Fig. 3. Figshare website

Figshare работает более чем со 100 университетами, музеями, фондами, научными ассоциациями по всему миру. Пользователи могут загружать файлы в формате любого типа: документы, плакаты, рисунки, видеофайлы, 3D-модели, оцифрованные картины, музыкальные файлы, базы больших данных, наборы данных. Им автоматически присваиваются DOI. Загрузка контента бесплатна, предоставляется свободный доступ в соответствии с принципом открытых данных. Большинство файлов загружаются под лицензией Creative Commons, CC-BY и CC0 для наборов данных [16]. Figshare позволяет учёным публиковать исследования с отрицательным результатом⁴, отслеживает альтметрические показатели.

Для пользователей предоставлен простой и расширенный варианты поиска. Результаты поиска ранжируются следующим образом: сначала видны объекты, содержащие все либо несколько терминов, затем – документы, содержащие один из терминов, далее – объекты, содержащие похожие слова. Есть возможность использования булевых операторов, фасетной навигации по типу контента,

⁴ Fenner M. Figshare: Interview with Mark Hahnel // Front Matter Blog : [сайт]. 2012. February 16. URL: <https://blog.front-matter.io/posts/figshare-interview-with-mark-hahnel/> (дата обращения: 20.05.2024). DOI 10.53731/r294649-6f79289-8cw2r.

типу записи, категории, источнику, лицензии. В результате поиска, помимо просмотра метаданных, можно получить ссылку на цитирование, загрузить полный текст или изображение, поделиться им в соцсетях или по электронной почте, добавить в коллекцию, заявить об авторстве, написать сообщение.

Результаты поиска возможно отсортировать по релевантности, цитируемости, альтметрическим показателям, прямой и обратной хронологии, дате онлайн-публикации и т. д.

Figshare позволяет авторам бесплатно загружать файлы любого формата до 5 Гб, использовать 20 Гб личного хранилища.

В российской практике репозитории данных активного развития не получили. Иногда исследовательские данные включены в институциональные репозитории (ИР) и электронные библиотеки научных и образовательных учреждений. Например, Электронная библиотечная система Алтайского государственного университета включает объекты интеллектуальной собственности: БД, изобретения, полезные модели, программы для ЭВМ, промышленные образцы, селекционные достижения. Каталог публикаций Института народнохозяйственного прогнозирования РАН помимо публикаций включает информацию об археологических объектах (рис. 4).

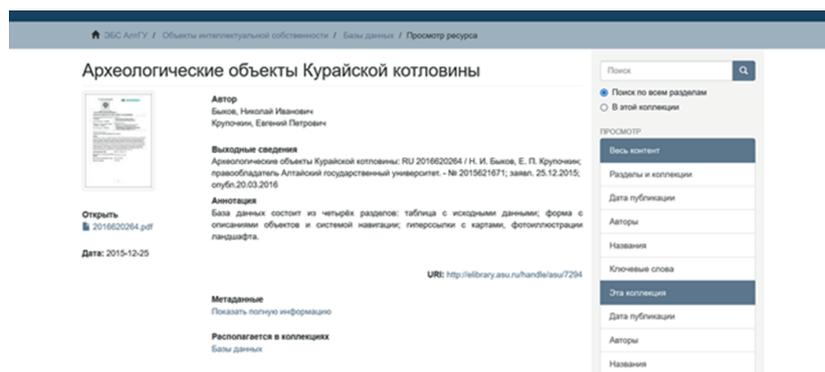


Рис. 4. Карточка объекта Каталога публикаций
Института народнохозяйственного прогнозирования РАН
Fig. 4. Card of an object in the Catalog of publications
of the Institute of National Economic Forecasting of the RAS

Но, как правило, эти документы представлены в ИР в небольших количествах.

С целью оценки текущего состояния отечественных репозиториев данных в 2023 г. было проведено исследование сайтов научных организаций трёх категорий, утверждённых Федеральным агентством научных организаций согласно постановлению Правительства РФ от 8 апреля 2009 г. № 312 «Об оценке и о мониторинге результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения»⁵, и образовательных учреждений – участников программы «Приоритет 2030». База исследования обоснована тем, что данные научные и образовательные учреждения являются ведущими организациями страны.

⁵ С полным списком организаций по категориям можно ознакомиться на сайте Indicator: Разобрали по категориям: списки научных организаций ФАНО после оценки их деятельности // Indicator : [сайт]. 2018. 3 апреля. URL: <https://indicator.ru/engineering-science/ocenka-rezultativnosti-akademicheskikh-institutov-fano.htm> (дата обращения: 20.05.2024).

Исследование сайтов образовательных и научных учреждений показало, что репозитории исследовательских данных практически не создаются в рамках этих учреждений. Всего в трёх вузах были найдены репозитории данных, но, как правило, доступ к ним закрыт для удалённого пользователя. В двадцати научных организациях найдена информация об исследовательских данных в ОД. Однако в большинстве из них сведения представлены в виде перечней данных, не предусматривающих поиск по определённым признакам. Выявлено лишь несколько архивов, которые можно отнести к репозиториям исследовательских данных. Зачастую они называются архивами, каталогами, БД, порталами и т. п. (табл. 2).

Таблица 2

Исследовательские данные научных и образовательных учреждений

Table 2

Research data from academic and educational institutions

Учреждение	Название репозитория	Адрес репозитория	Доступность
Национальный исследовательский университет «Высшая школа экономики»	Единый архив археологических и социологических данных	http://sophist.hse.ru	Находится на реконструкции. Доступ к некоторым коллекциям возможен только в локальной сети университета
Северо-Восточный федеральный университет им. М. К. Аммосова (СВФУ)	Репозиторий отчётов СВФУ	https://s-vfu.ru/sotrudniku/repository_svfu	Доступ разрешён только для сотрудников СВФУ
Южный федеральный университет	Цифровой репозиторий	https://hub.sfedu.ru/repository	Часть материалов в ОД
Физико-технический институт им. А. Ф. Иоффе РАН	Тематические БД, архивы и каталоги	http://old.ioffe.ru/index.php?go=physDB#	ОД
Специальная астрофизическая обсерватория РАН	Архив общего наблюдения	https://www.sao.ru/oasis/cgi-bin/fetch?lang=ru,%20	ОД
	БД по объектам локального тома	https://www.sao.ru/lv/lvgdb	
Институт прикладной астрономии РАН	Наблюдения	https://iaaras.ru/observations	ОД
Ботанический институт им. В. Л. Комарова РАН	Электронный гербарий водорослей	https://binran.ru/structure/podrazdelenia/laboratoriya-algologii/baza-dannykh/	ОД
Институт вулканологии и сейсмологии ДВО РАН (ИВиС ДВО РАН)	Геопортал ИВиС ДВО РАН	http://geoportal.kscnet.ru/data	ОД
Институт космофизических исследований и распространения радиоволн ДВО РАН	Данные текущих наблюдений	https://www.ikir.ru/ru/Data	ОД
Институт земного магнетизма, ионосферы и распространения радиоволн им. Н. В. Пушкова РАН	Данные наблюдений, модели, прогнозы	https://izmiran.ru/data	ОД
	БД магнитометрических наблюдений спутников «Прогноз»	https://izmiran.ru/projects/PROGNOZ	

Продолжение Табл. 2 см. на стр. 130

Продолжение Табл. 2

Учреждение	Название репозитория	Адрес репозитория	Доступность
Главная (Пулковская) астрономическая обсерватория РАН	Система астрометрических БД; БД пулковских фотографических пластинок; Интерактивная БД скоростей ГНСС-станций; Каталог OCARS; Предстоящие сближения и покрытия планет и радиоисточников; Интерактивная БД по солнечной активности; Extended time series of Solar Activity Indices; Объединённая БД магнитных полей солнечных пятен; Бюллетень «Солнечные данные»; Стеклоанная библиотека	https://www.gaoran.ru/?page_id=226	ОД
Институт биологических проблем Севера ДВО РАН (ИБПС ДВО РАН)	Гербарий ИБПС ДВО РАН (участник консорциума Цифрового гербария МГУ)	https://plant.depo.msu.ru/module/collectionpublic?openparams=%5bopen-id=157050739	ОД
Институт систематики и экологии животных СО РАН (ИСиЭЖ СО РАН)	Банк данных ИСиЭЖ СО РАН по численности животных Северной Евразии	http://eco.nsc.ru/data_bank.html	ОД
Главный ботанический сад им. Н. В. Цицина РАН (ГБС РАН)	Гербарий ГБС РАН (участник консорциума Цифрового гербария МГУ)	https://plant.depo.msu.ru/module/collectionpublic?openparams=%5bopen-id=130169625	ОД
Центральный сибирский ботанический сад СО РАН	VOTANY	http://herb.csbg.nsc.ru:8081	ОД
Государственный геологический музей им. В. И. Вернадского РАН (ГГМ РАН)	Портал открытых данных ГГМ РАН	https://sgm.ru/SCIENCE/open_data_sgm.php	ОД
Институт теории прогноза землетрясений и математической геофизики РАН	Карты и БД	https://itpz-ran.ru/ru/resultaty/maps-and-databases	ОД
Институт аналитического приборостроения РАН	Каталог приборов	http://iairas.ru/catalog.php	ОД
Институт проблем проектирования в микроэлектронике РАН (ИППМ РАН)	Библиотека схемотехнических решений	http://ippm.ru/index.php?page=journ&code=J1	ОД
Научно-исследовательский институт вакцин и сывороток им. И. И. Мечникова (НИИВС им. И. И. Мечникова)	Коллекция микроорганизмов III и IV групп патогенности НИИВС им. И. И. Мечникова	https://ckp-rf.ru/catalog/usu/676268/	На момент проведения исследования коллекция недоступна для удалённого пользователя
Уссурийская астрофизическая обсерватория ДВО РАН	Наблюдения	http://uafu.ru/observ_rus.php	ОД

Продолжение Табл. 2 см. на стр. 131

Продолжение Табл. 2

Учреждение	Название репозитория	Адрес репозитория	Доступность
Институт космофизических исследований и аэронауки им. Ю. Г. Шафера СО РАН (ИКФИА СО РАН)	Данные ИКФИА СО РАН; Оперативный спутниковый мониторинг Якутии; Интерактивная карта грозовой активности в Якутии; Оперативные данные грозовой активности в Якутии; Ионосфера над Якутском; Уровень магнитной возмущенности; Данные риометра Тикси; Данные измерений интенсивности космических лучей в Якутске и Тикси	https://ikfia.ysn.ru	ОД
Институт астрономии РАН	Центр астрономических данных Института астрономии РАН	http://www.inasan.ru/divisions/dpss/cad	ОД

Среди образовательных организаций открытым репозиторием данных является цифровой репозиторий Южного федерального университета, который помимо текстовых материалов содержит наборы данных, презентации. Возможен простой поиск по ключевым словам с применением уточнения поиска по названию коллекции. Например, один из объектов включает имитационную модель сахарного завода (рис. 5). Метаданные объекта содержат ФИО автора, название, аннотацию, ключевые слова на английском и русском языках, DOI, статистические данные.

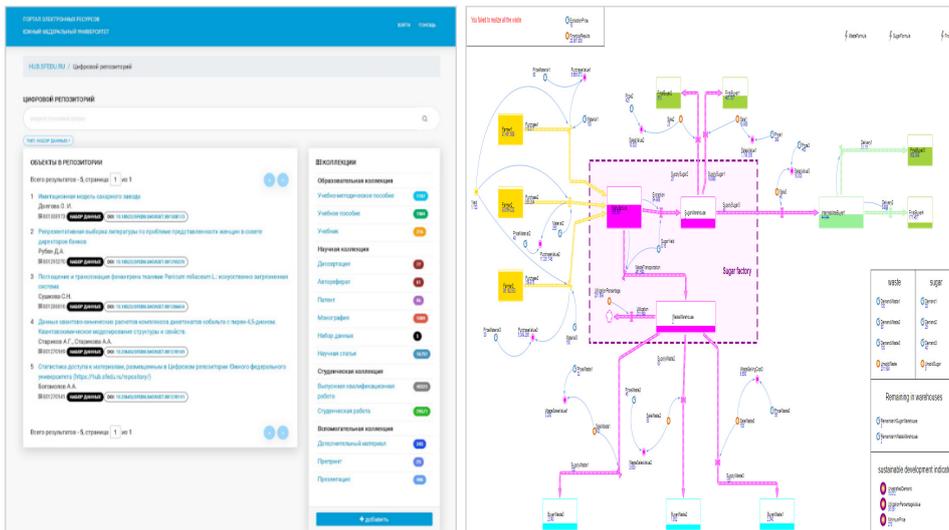


Рис. 5. Контент Цифрового репозитория Южного федерального университета
Fig. 5. Content of the Digital Repository of Southern Federal University

Среди научных учреждений одним из наиболее интересных примеров является Электронный гербарий водорослей Лаборатории альгологии Ботанического института им. В. Л. Комарова, включающий изображения и метаданные типовых образцов гербария. На карточках объектов размещены сведения о времени создания и создателях, регион обитания, ссылки для цитирования. Всего гербарий включает более 167 000 объектов (рис. 6).

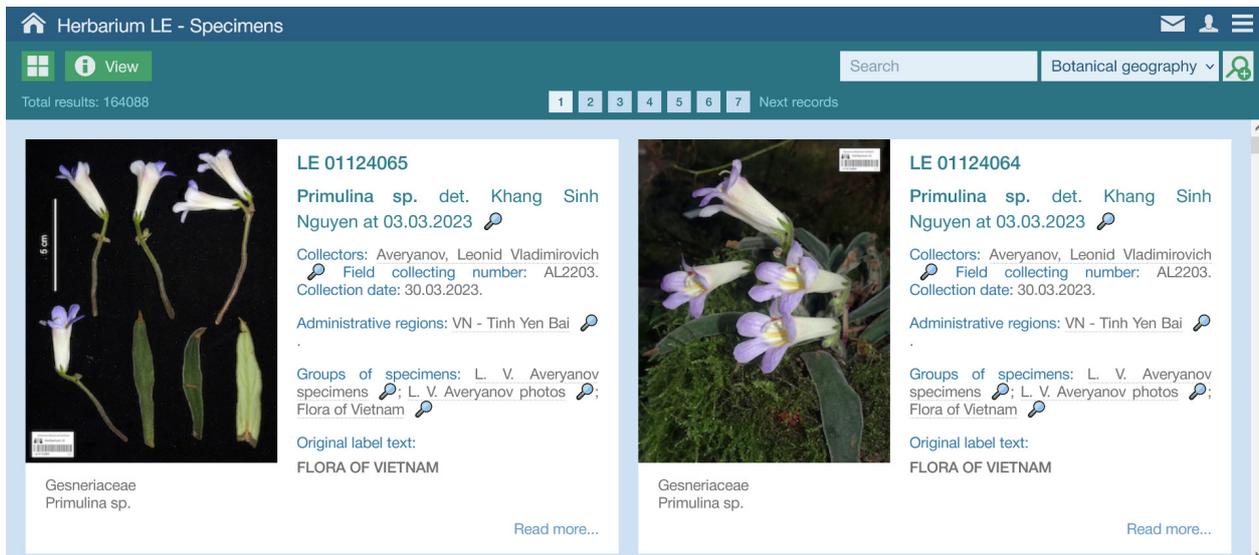


Рис. 6. Контент электронного гербария водорослей Лаборатории альгологии Ботанического института им. В. Л. Комарова

Fig. 6. Content of the electronic algae herbarium of the Algology Laboratory at Komarov Botanical Institute

Возможен простой поиск по большому количеству признаков: по областям знания, регионам, коллекциям, географическому положению, формам жизни и другим данным.

Другим успешным примером реализации репозитория исследовательских данных является Портал открытых данных Государственного геологического музея им. В. И. Вернадского. Все коллекции портала размещены по соответствующим инвентарям: «Минералы», «Горные породы и руды», «Ископаемая флора», «Ископаемые позвоночные», «Изделия из камня» (рис. 7). Каждая коллекция включает наборы данных, содержащих тексты, изображения, ссылки на источники. Портал включает также частные коллекции, которые содержат информацию о государственных деятелях, монографические и исторические коллекции.

Возможен простой вариант поиска по слову или словосочетанию. Также предусмотрен поиск по наименованию, месторождению, названию минерала, породы и руды при самостоятельном вводе термина или через справочник. Существует возможность поиска по географическому региону, геологическому периоду, лицам, ответственным за формирование коллекции.

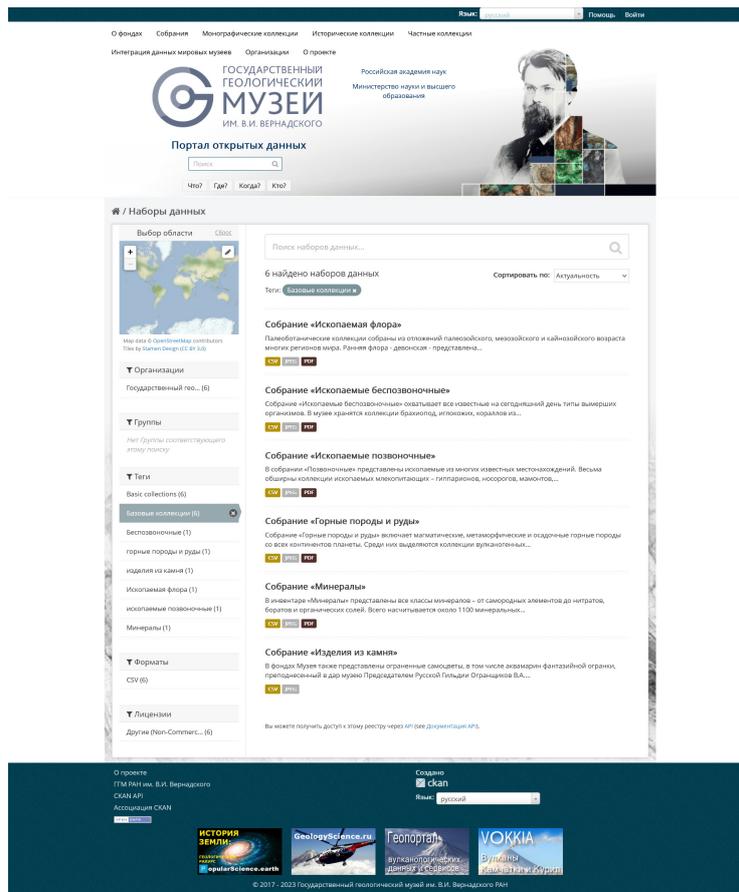


Рис. 7. Коллекции Портала открытых данных Государственного геологического музея им. В. И. Вернадского
Fig. 7. Collections of the Open Data Portal of Vernadsky State Geological Museum

Карточки объектов портала представляют собой изображения экспонатов и их характеристику, включающую название, размеры, год сбора и другие данные (рис. 8).

Дополнительная информация	
Поле	Величина
Data last updated	8 апреля 2018 г., 13:07 (UTC+07:00)
Metadata last updated	8 апреля 2018 г., 13:07 (UTC+07:00)
Создано	8 апреля 2018 г., 13:07 (UTC+07:00)
Формат	image/jpeg
Лицензия	Другие (Non-Commercial)

Рис. 8. Карточка объекта Портала открытых данных Государственного геологического музея им. В. И. Вернадского
Fig. 8. Card of an object at the Open Data Portal of Vernadsky State Geological Museum

В настоящее время наборы данных на портале насчитывают более 82 000 экспонатов, более 98 000 предметов.

Одним из примеров корпоративного участия научных и образовательных учреждений в формировании репозитория исследовательских данных является Национальный банк-депозитарий живых систем. Это проект Московского государственного университета «Ноев ковчег», который посвящён созданию многофункционального сетевого хранилища биологического материала. Оно состоит из пяти направлений: «Растения», «Животные», «Микроорганизмы и грибы», «Биоматериалы человека», «Биологическая информация». В настоящее время в рамках направления «Растения» в формировании цифрового гербария принимают участие девять научных и образовательных организаций (рис. 9).

The screenshot displays the website of the National Bank of Living Systems (Ноев Ковчег). The main header includes the logo and name of the bank. Below the header, there are navigation links: «О системе», «Коллекции», «Контакты», «Ссылки», «Инфраструктура», and «Цитировать». A summary of the database statistics is provided: «Сейчас в базе данных (гербарные образцы):» with icons for samples (1,309,886), images (1,287,261), views (36,648), geotags (949,519), labels + OCR (642,233 + 664,836), collectors (18,531), views (8,219,026), maps (9,885), and answers in the Pomogator (882,102). Below this, a section titled «Вклад участников консорциума Цифрового гербария МГУ» lists various institutions and their contributions:

- MW (Гербарий Московского университета)**: 1,068,828 образцы, 1,046,562 изображения, 36,117 видов, 761,595 геопривязки, 512,710 + 553,319 этикеток + OCR.
- MNA (Гербарий Главного ботанического сада РАН, г. Москва), с 1.04.2019**: 155,366 образцы, 155,057 изображений, 4,673 вида, 109,140 геопривязки, 55,241 + 100,115 этикеток + OCR.
- IRKU (Гербарий Иркутского государственного университета), с 11.09.2020**: 42,297 образцы, 42,154 изображения, 1,026 видов, 35,519 геопривязки, 30,910 + 11,387 этикеток + OCR.
- KUZ (Гербарий Кузбасского ботанического сада СО РАН, г. Кемерово), с 14.05.2020**: 19,014 образцы, 19,104 изображения, 1,333 вида, 19,007 геопривязки, 19,009 + 5 этикеток + OCR.
- TUL (Гербарий Тульского государственного педагогического университета имени Л.Н. Толстого), с 25.12.2019**: 10,098 образцы, 10,122 изображения, 1,144 вида, 10,089 геопривязки, 10,097 + 0 этикеток + OCR.
- TULGU (Гербарий Тульского государственного университета), с 07.2021**: 5,054 образцы, 5,054 изображения, 767 видов, 5,047 геопривязки, 5,053 + 0 этикеток + OCR.
- KULPOL (Гербарий Музея-заповедника «Куликово поле», г. Тула), с 07.2021**: 3,657 образцы, 3,656 изображений, 550 видов, 3,657 геопривязки, 3,657 + 0 этикеток + OCR.
- MAG (Гербарий Института биологических проблем Севера ДВО РАН, г. Магадан), с 22.10.2020**: 2,604 образцы, 2,604 изображения, 99 видов, 2,535 геопривязки, 2,598 + 6 этикеток + OCR.

Рис. 9. Цифровой гербарий
Fig. 9. Digital herbarium

Депозитарий включает образцы, изображения, звуки, виды, геопривязки, карты атласа, книги и т. п.

Предусмотрен простой поиск по всем полям, роду/виду и расширенный поиск по большему количеству параметров (рис. 10).

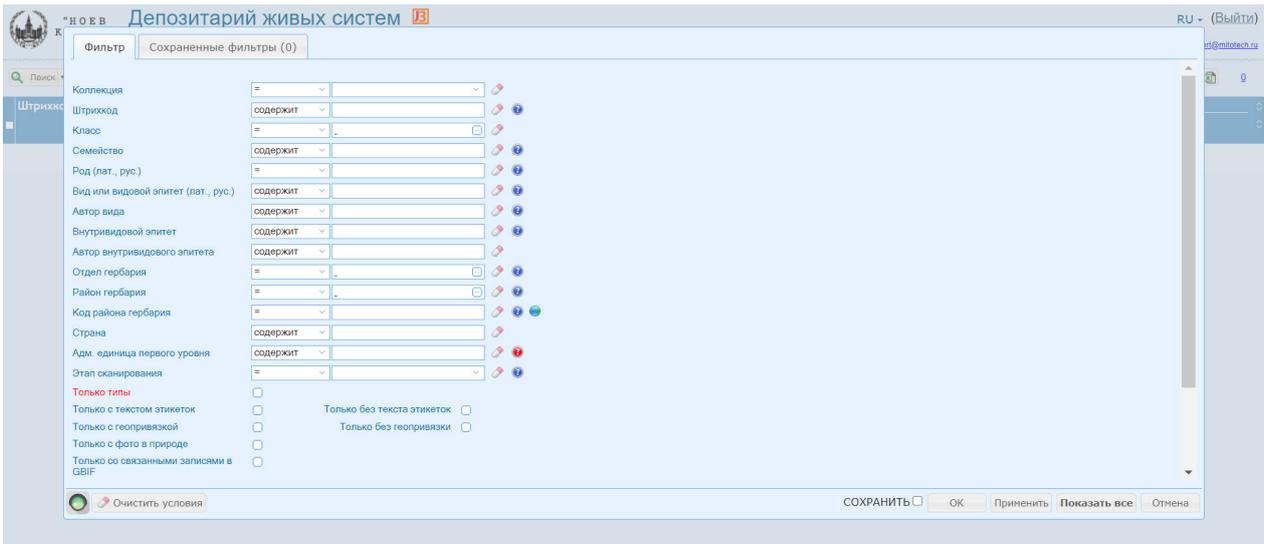


Рис. 10. Поисковые возможности депозитария живых систем «Ноев ковчег»
Fig. 10. Search capabilities of the Living Systems Depository “Noah’s Ark”

Карточки объектов содержат не только основные метаданные, но и сведения о месте сбора образца (рис. 11).

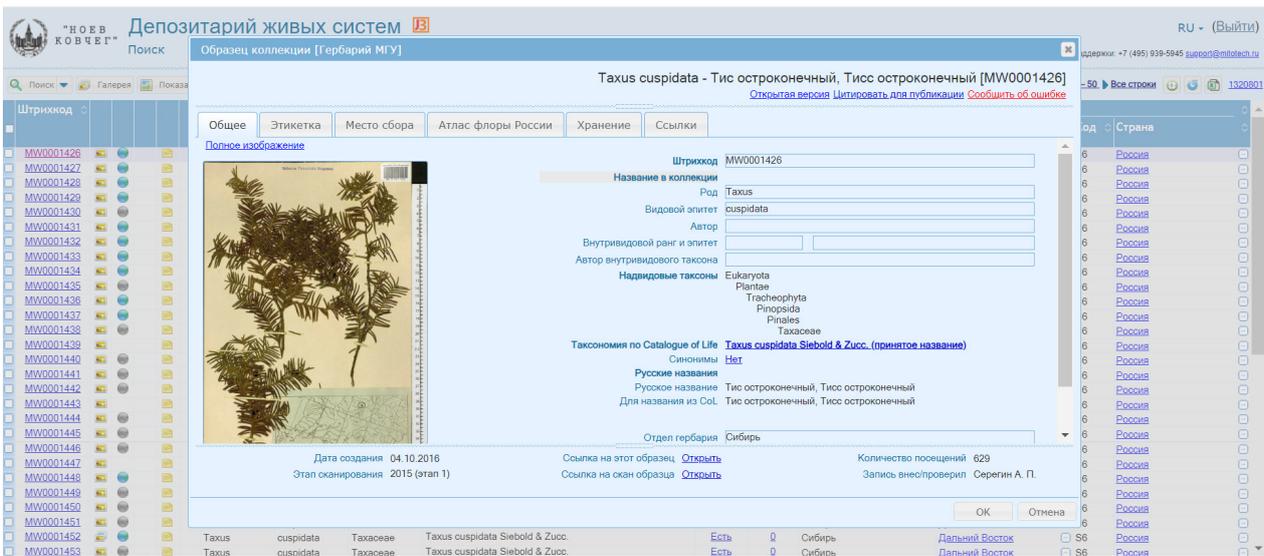


Рис. 11. Карточка объекта депозитария живых систем «Ноев ковчег»
Fig. 11. Card of an object in the Living Systems Depository “Noah’s Ark”

Чтобы облегчить ориентирование в огромном количестве разнообразных проектов ОД, были созданы агрегаторы репозиторий. Наиболее известным среди них является Реестр репозиторий исследовательских данных (re3data.org) – это открытый научный инструмент, который предлагает исследователям, организациям, библиотекам и издателям обзор существующих международных репозиторий исследовательских данных (рис. 12).



Рис. 12. Сайт re3data.org
Fig. 12. Re3data.org website

На 13 мая 2024 г. реестр содержит более 3000 хранилищ исследовательских данных со всего мира, охватывающих все академические дисциплины (рис. 13).

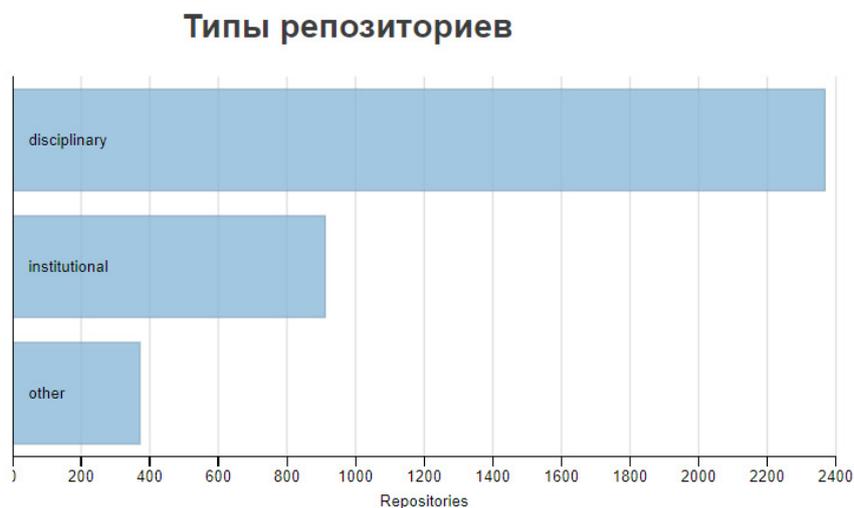


Рис. 13. Типы репозитория в re3data.org
Fig. 13. Types of repositories in re3data.org

Re3data.org включает текстовые данные, графики, изображения, БД, аудио-визуальные данные и т. п. (рис. 14).

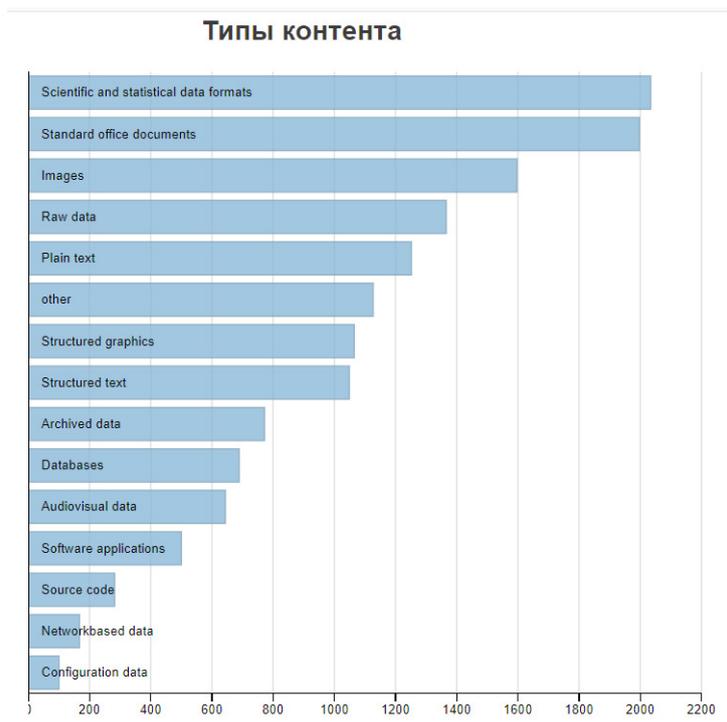


Рис. 14. Контент re3data.org
Fig. 14. Content at re3data.org

К сожалению, российские репозитории в re3data.org почти не представлены. В результате поиска по стране было найдено всего 23 репозитория, в половине из них Российская Федерация фигурирует в качестве соучастника, и только 12 репозитория, созданных в России, найдено в реестре.

Например, Мировой центр данных солнечно-земной физики отражён в реестре следующим образом: указано его название, предметная область, краткое описание, виды контента, ключевые слова, ссылки на репозиторий и т. п. (рис. 15).

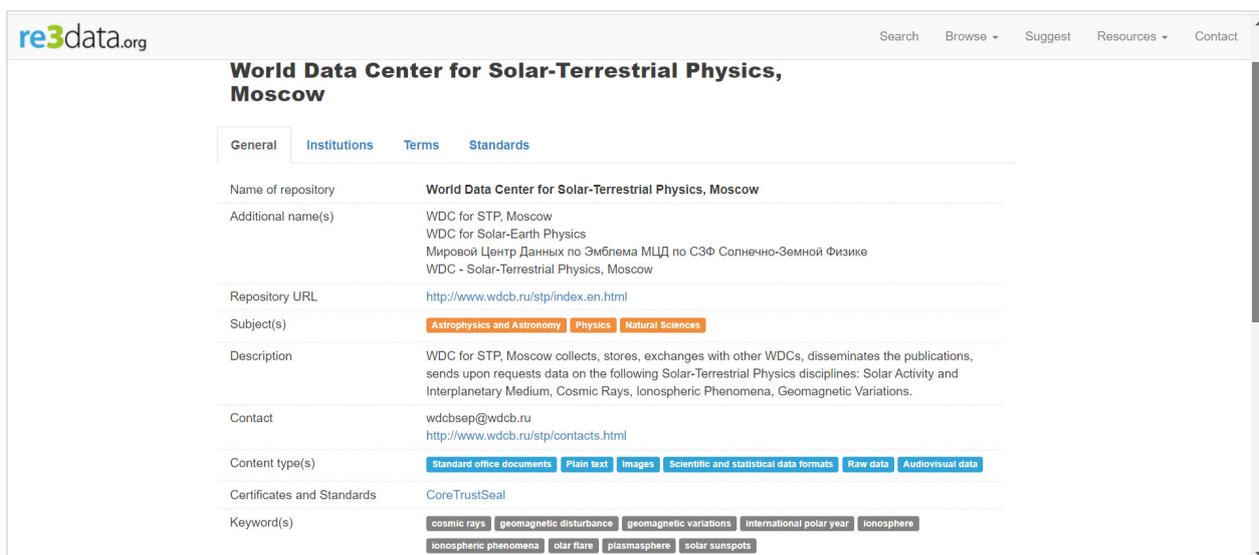


Рис. 15. Карточка репозитория в re3date.org
Fig. 15. Card of a repository at re3data.org

Однако из 12 репозиторий только в трёх дана корректная ссылка.

Из рассматриваемых научных и образовательных учреждений только два репозитория зарегистрированы в Re3data: Институт астрономии РАН и Институт органической химии им. Н. Д. Зелинского РАН. Однако в карточке последнего дана некорректная ссылка, при этом на сайте учреждения репозиторий данных найден не был.

Помимо Re3data, согласно рекомендациям по политике обмена исследовательскими данными, репозитории должны быть зарегистрированы в FAIRsharing.org [17, с. 8].

FAIRsharing – это информационная и образовательная служба, которая расширяет возможности потребителей по стандартам, БД, репозиториям и политикам в отношении данных, ускоряя поиск, выбор и использование этих ресурсов (рис. 16). В результате поиска российских репозиторий на данной платформе было найдено только два действующих репозитория: Всемирный центр радиационных данных (<http://wrdc.mgo.rssi.ru>) и Всероссийская коллегия микроорганизмов (<https://vkm.ru/rus/fungalDC.htm>).

Т. о. российские репозитории исследовательских данных практически не представлены на международных платформах.

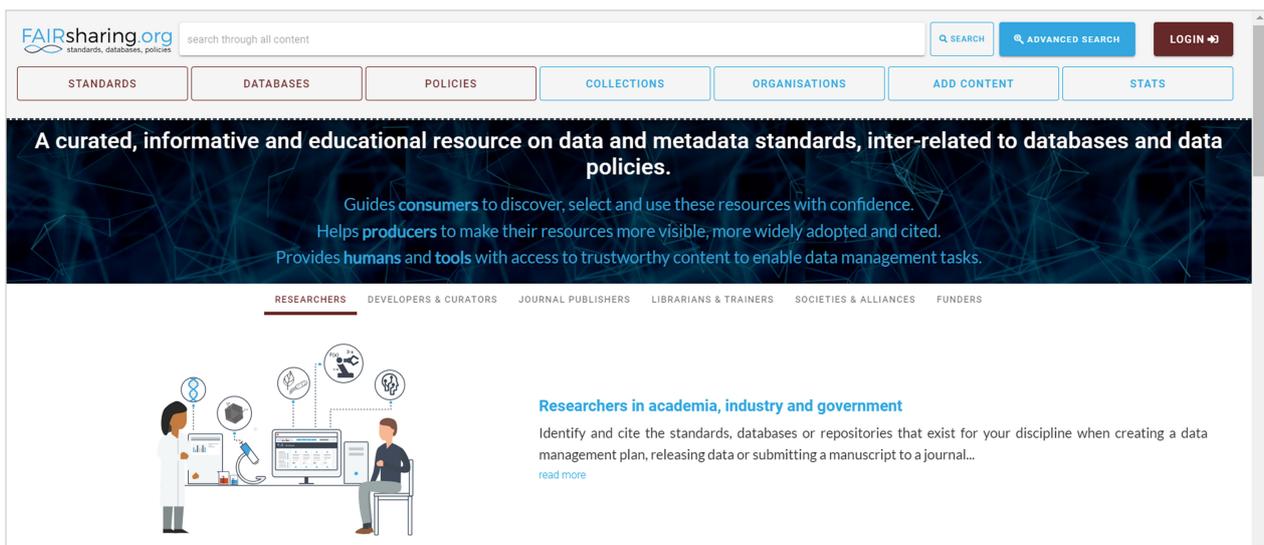


Рис. 16. Сайт FAIRsharing
Fig. 16. FAIRsharing website

ЗАКЛЮЧЕНИЕ

В то время как репозитории данных активно развиваются за рубежом, культура обмена исследовательскими данными в России пока находится на начальном этапе формирования.

Исследование сайтов отечественных научных и образовательных учреждений подтвердило тот факт, что в нашей стране репозитории данных практически не создаются. На сайтах вузов эти сведения, как правило, закрыты для удалённого пользователя, а большинство исследовательских данных научных учреждений предоставляются на сайтах в виде перечня, без возможности их поиска по определённым признакам и управления результатами поиска.

Отечественные репозитории практически не представлены в зарубежных агрегаторах Re3data, FAIRsharing, а имеющаяся в них информация в большинстве случаев некорректна. Это негативно влияет на интеграцию российских достижений в мировую науку.

В то же время размещение исследовательских данных в репозиториях, информирование о репозиториях в международных агрегаторах способствует видимости, доступности и быстрому распространению результатов исследования, исключает необходимость повторного сбора данных, позволяет избежать дублирования работ. При этом одним из необходимых условий их эффективного использования является ОД к информации, чтобы любой пользователь имел возможность исследовать, загружать и использовать данные репозитория для исследовательских или образовательных целей.

Репозитории исследовательских данных являются показателем качества деятельности научного или образовательного учреждения, показывают научную, социальную и экономическую значимость исследовательских работ и таким образом демонстрируют статус и общественное значение организации [18, с. 1325].

Таким образом, репозитории исследовательских данных играют большую роль в системе научной коммуникации, позволяя учёным распространять результаты своих исследований и использовать данные, полученные в ходе деятельности других исследователей.

СПИСОК ЛИТЕРАТУРЫ

1. *Михеенков А. В.* Самоархивирование и открытые репозитории: методические рекомендации. М. : Ваше цифровое издательство, 2018. 28 с. ISBN 978-5-6040408-1-2.
2. *Armbruster C., Romary L.* Comparing repository types: Challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication // *International Journal of Digital Library Systems (IJDLS)*. 2010. Vol. 1, № 4. P. 61–73. DOI 10.4018/jdls.2010100104.
3. *Santos D. B. G., Silva L. P. C., Coelho J. S.* Thematic repository as alternative model for science communication // 5th Iberian Conference on Information Systems and Technologies (Santiago de Compostela, Spain, June 16–19, 2010). Piscataway, NJ : IEEE, 2010. P. 418–423.
4. *Jain P., Bentley G., Oladiran M. T.* The role of institutional repository in digital scholarly communications : A paper presented at African Digital Scholarship & Curation Conference on 12–14 May, 2009 at CSIR Conference Center, Pretoria, South Africa. URL: https://researchgate.net/publication/237322259_The_Role_of_Institutional_Repository_in_Digital_Scholarly_Communications (дата обращения: 11.05.2024).
5. *Юдина И. Г., Федотова О. А.* Репозитории научных публикаций открытого доступа: история и перспективы развития // *Информационное общество*. 2020. № 6. С. 67–79. EDN ILTQDB.
6. *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. DOI 10.33186/1027-3689-2020-9-121-142. EDN HQTFGA.
7. *Рождественская М. Ю.* Репозиторий как реализация идей открытого доступа к научным публикациям: подходы к классификации // *Библиосфера*. 2015. № 2. С. 86–94. EDN TSIJHX.

8. Стукалова А. А. Платформы для интеграции ресурсов открытого доступа научных и образовательных учреждений // *Управление наукой: теория и практика*. 2024. Т. 6, № 2. С. 168–187. DOI 10.19181/smtp.2024.6.2.12. EDN OSKORH.
9. Косычева М. А., Хорохорина Г. А. Репозитории данных: теория и практика // *Health, Food & Biotechnology*. 2020. Т. 2, № 2. С. 7–11. DOI 10.36107/hfb.2020.i2.s393. EDN STKGDC.
10. Классификация данных – Руководство по управлению исследовательскими данными / разработчик: Н. С. Редькина // ГПНТБ СО РАН : [сайт]. URL: <http://www.spsl.nsc.ru/naukresursy-i-uslugi-gpntb-so-ran-dlya-nauki-i-biznesae-i-biznesu/rdm/klassifikaciya-dannux/> (дата обращения: 11.05.2024).
11. Zenodo in the spotlight of traditional and new metrics / I. Peters, P. Kraker, E. Lex [et al.] // *Frontiers in Research Metrics and Analytics*. 2017. Vol. 2. Article 13. DOI 10.3389/frma.2017.00013.
12. Thelwall M., Kousha K. Figshare: A universal repository for academic resource sharing? // *Online Information Review*. 2016. Vol. 40, № 3. P. 333–346. DOI 10.1108/OIR-06-2015-0190.
13. Data quality issues and content analysis for research data repositories: The case of Dryad / D. Rousidis, E. Garoufallou, P. Balatsoukas, M.-A. Sicilia // *Let's put data to use: Digital scholarship for the next generation : Proceedings of the 18th International Conference on Electronic Publishing*. Ed. by P. Polydoratou, M. Dobрева. Amsterdam ; Berlin ; Tokyo ; Washington, DC : IOS Press, 2014. P. 49–58. DOI 10.3233/978-1-61499-409-1-49.
14. Etz K., Kimmel H. L., Pienta A. National addiction and HIV Data Archive Program: Developing an approach for reuse of sensitive and confidential data // *The Journal of Privacy and Confidentiality*. 2023. Vol. 13, № 2. DOI 10.29012/jpc.853.
15. Payakachat N., Tilford J. M., Ungar W. J. National Database for Autism Research (NDAR): Big data opportunities for health services research and health technology assessment // *PharmacoEconomics*. 2016. Vol. 34, № 2. P. 127–138. DOI 10.1007/s40273-015-0331-6.
16. Hane P. J. Sharing research data – New figshare for institutions // *Charleston Hub* : [сайт]. 2013. September 16. URL: <https://charleston-hub.com/2013/09/sharing-research-data-new-figshare-for-institutions-2> (дата обращения: 20.05.2024).
17. Политика обмена исследовательскими данными (Data Sharing Policy) / сост.: М. М. Зельдина ; Некоммерческое партнёрство «Национальный электронно-информационный консорциум». М. : Ваше цифровое издательство, 2019. 20 с. ISBN 978-5-6040408-7-4. DOI 10.24108/978-5-6040408-7-4.
18. Патук М. И., Наумова В. В., Еременко В. С. Цифровой репозиторий “Geologyscience.ru”: открытый доступ к научным публикациям по геологии России // *Электронные библиотеки*. 2020. Т. 23, № 6. С. 1324–1338. DOI 10.26907/1562-5419-2020-23-6-1324-1338. EDN PFMJRJ.

REFERENCES

1. Mikheenkov A. V. Self-archiving and open repositories: Guidelines [Samoarkhivirovanie i otkrytye repozitorii: metodicheskie rekomendatsii]. Moscow : Vashe tsifrovoe izdatel'stvo; 2018. 28 p. (In Russ.). ISBN 978-5-6040408-1-2.
2. Armbruster C., Romary L. Comparing repository types: Challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication. *International Journal of Digital Library Systems (IJDLS)*. 2010;1(4):61–73. DOI 10.4018/jdls.2010100104.

3. Santos D. B. G., Silva L. P. C., Coelho J. S. Thematic repository as alternative model for science communication. In: 5th Iberian Conference on Information Systems and Technologies (Santiago de Compostela, Spain, June 16–19, 2010). Piscataway, NJ : IEEE; 2010. P. 418–423.
4. Jain P., Bentley G., Oladiran M. T. The role of institutional repository in digital scholarly communications : A paper presented at African Digital Scholarship & Curation Conference on 12–14 May, 2009 at CSIR Conference Center, Pretoria, South Africa. Available at: https://researchgate.net/publication/237322259_The_Role_of_Institutional_Repository_in_Digital_Scholarly_Communications (accessed: 11.05.2024).
5. Yudina I. G., Fedotova O. A. Open access scientific publications repositories: History and development prospects. *Information Society=Informatsionnoe obshchestvo*. 2020;(6):67–79. (In Russ.).
6. Zassoursky I. I., Sokolova D. V., Trishchenko N. D. Open access repositories: Functions and trends. *Scientific and Technical Libraries=Nauchnye i tehicheskie biblioteki*. 2020;(9):121–142. (In Russ.). DOI 10.33186/1027-3689-2020-9-121-142.
7. Rozhdestvenskaya M. Yu. A repository as ideas realization of open access for scientific publications: Classification approaches. *Bibliosphere=Bibliosfera*. 2015;(2):86–94. (In Russ.).
8. Stukalova A. A. Platforms for integration of open access resources of research and educational institutions. *Science Management: Theory and Practice*. 2024;6(2):168–187. (In Russ.) DOI 10.19181/sntp.2024.6.2.12.
9. Kosycheva M. A., Khorokhorina G. A. Data repositories: Theory and practice. *Health, Food & Biotechnology*. 2020;2(2):7–11. (In Russ.). DOI 10.36107/hfb.2020.i2.s393.
10. Red'kina N. S. Data classification – A guide to research data management [Klassifikatsiya dannykh – Rukovodstvo po upravleniyu issledovatel'skimi dannymi]. *SPSTL SB RAS*. Available at: <http://www.spsl.nsc.ru/naukresursy-i-uslugi-gpntb-so-ran-dlya-nauki-i-biznesae-i-biznesu/rdm/klassifikaciya-dannyx/> (accessed: 22.05.2024). (In Russ.).
11. Peters I., Kraker P., Lex E., Gumjenberger C., Gorraiz J. Zenodo in the spotlight of traditional and new metrics. *Frontiers in Research Metrics and Analytics*. 2017;2:13. DOI 10.3389/frma.2017.00013.
12. Thelwall M., Kousha K. Figshare: A universal repository for academic resource sharing? *Online Information Review*. 2016;40(3):333–346. DOI 10.1108/OIR-06-2015-0190.
13. Rousidis D., Garoufallou E., Balatsoukas P., Sicilia M.-A. Data quality issues and content analysis for research data repositories: The case of Dryad. In: Polydoratos P., Dobрева M., eds. *Let's put data to use: Digital scholarship for the next generation* : Proceedings of the 18th International Conference on Electronic Publishing. Amsterdam ; Berlin ; Tokyo ; Washington, DC : IOS Press; 2014. P. 49–58. DOI 10.3233/978-1-61499-409-1-49.
14. Etz K., Kimmel H. L., Pienta A. National addiction and HIV Data Archive Program: Developing an approach for reuse of sensitive and confidential data. *The Journal of Privacy and Confidentiality*. 2023;13(2). DOI 10.29012/jpc.853.
15. Payakachat N., Tilford J. M., Ungar W. J. National Database for Autism Research (NDAR): Big data opportunities for health services research and health technology assessment. *PharmacoEconomics*. 2016;34(2):127–138. DOI 10.1007/s40273-015-0331-6.
16. Hane P. J. Sharing research data – New figshare for institutions. *Charleston Hub*. 2013. September 16. Available at: <https://charleston-hub.com/2013/09/sharing-research-data-new-figshare-for-institutions-2> (accessed: 22.05.2024).
17. Research data sharing policy (Data Sharing Policy) [Politika obmena issledovatel'skimi dannymi (Data Sharing Policy)] / comp. by M. M. Zeldina ; Non-profit Partnership “National Electronic Information Consortium”. Moscow : Vashe tsifrovoe izdatel'stvo; 2019. 20 p. (In Russ.). ISBN 978-5-6040408-7-4. DOI 10.24108/978-5-6040408-7-4.
18. Patuk M. I., Naumova V. V., Eremenko V. S. Digital repository “Geologyscience.ru”: Open access to scientific publications on Russian geology. *Russian Digital Libraries Journal*. 2020;23(6):1324–1338. (In Russ.). DOI 10.26907/1562-5419-2020-23-6-1324-1338.

*Поступила в редакцию / Received 13.09.2024.
Одобрена после рецензирования / Revised 05.10.2024.
Принята к публикации / Accepted 30.11.2024.*

СВЕДЕНИЯ ОБ АВТОРЕ

Стукалова Анна Александровна *stukalova@gpntbsib.ru*

Кандидат педагогических наук, старший научный сотрудник, Государственная публичная научно-техническая библиотека СО РАН, Новосибирск, Россия
SPIN-код: 7830-6631

INFORMATION ABOUT THE AUTHOR

Anna A. Stukalova *stukalova@gpntbsib.ru*

Candidate of Pedagogy, Senior Researcher, State Public Scientific Technological Library, SB RAS, Novosibirsk, Russia
ORCID: 0000-0003-2202-943X