

ОТКРЫТЫЕ ДАННЫЕ КОЛЛЕКЦИЙ И НАХОДОК ГРИБОВ НА ТЕРРИТОРИЮ РОССИИ В GBIF

© 2024 г. Н.В. Филиппова^{1,*}

¹ Югорский государственный университет, 628011 Ханты-Мансийск, Россия

*e-mail: filippova.courlee.nina@gmail.com

Глобальная информационная система о биоразнообразии интегрирует находки видов растений, животных и грибов на международном уровне. В России активная оцифровка и мобилизация данных идет в области ботаники, зоологии, и менее активно в микологии. В публикации представлен анализ данных о находках грибов, опубликованных в GBIF на территорию России и обсуждаются проблемы и перспективы мобилизации данных о грибах на национальном уровне.

Ключевые слова: информатика биоразнообразия, коллекции, микота, открытые данные, оцифровка, фунга
DOI: 10.5281/zenodo.14182251

Открытые данные о находках видов животных, растений и грибов являются основной для решения фундаментальных и прикладных задач. Такие данные должны отвечать принципу FAIR (быть обнаружимы, доступны, совместимы и использоваться повторно). Принцип FAIR в области биоразнообразия обеспечивается существованием стандартов и платформ данных, самый распространенный стандарт Ядро Дарвина (Darwin Core), включающий около сотни терминов для описания образцов коллекций и находок видов.

Исходные данные о находках видов интегрируются в виде баз данных на уровне организации, и далее на национальных и международных порталах в зависимости от тематики или географического охвата. Для интеграции усилий на международном уровне и достижения договоренности между разными странами в области открытого свободного использования данных о биоразнообразии, в 2001 г. была создана организация Глобальной информационной системы о биоразнообразии (GBIF, gbif.org). GBIF в настоящее время является основным порталом о находках видов растений, животных и грибов и активно развивается в сторону интеграции данных молекулярного типа. В сферу деятельности GBIF как организации входит

образование, просвещение в области открытых данных, разработка систем управления самим порталом, управление сетью сотрудничества, поддержка использования данных, и пр.

На национальном уровне Россия не является официальным членом GBIF сети. Однако публикация и использование данных портала остаются возможными для Российских организаций на полномочной основе. В стране с 2005 г. ведется просветительская и организационная работа по области мобилизации данных на платформе GBIF: создан и поддерживается сайт на русском языке, ведется информационная рассылка в соцсетях, организуются образовательные мероприятия и конференции.

Анализ данных о находках грибов на территории России. Интерфейс GBIF предлагает ряд аналитических инструментов, где с помощью фильтров можно выбрать интересующие данные по заданным параметрам, а во вкладке «metrics» («default» и «custom») провести соответствующие анализы. Воспользовавшись этими функциями, можно изучить и визуализировать структуру и динамику публикационной активности по определенной стране и группе организмов.

Всего на территорию России опубликовано около 650 тысяч находок грибов, что составляет около 1.5% от всех находок грибов в мире, интегрированных в GBIF. Россия находится на 16-м месте в рейтинге стран по числу опубликованных на их территории находок грибов (рис. 1, А). По сравнению с другими группами организмов, грибы существенно отстают: из всех опубликованных российскими организациями находок на территории страны (всего около 6 млн), 52% представлено находками растений, 45% находками животных, и только 3% находками грибов (рис. 1, В). Публикационная активность страны более точно отражается не числом опубликованных находок на ее территорию, а числом находок, опубликованных из этой страны (т.е. российскими организациями). С этой точки зрения, из России опубликовано около 150 тысяч находок грибов, что составляет около 0.4% всех находок грибов в GBIF и Россия находится на 24-м месте в рейтинге стран (рис. 1, В).

Из страны есть опубликованные находки и на территории других стран (7% от всех находок, опубликованных из страны): Шпицберген, Чехия, Норвегия, Польша, Франция, Украина, Дания, Эстония и Финляндия. В основном это литературные данные (данные из набора об оцифрованных находках грибов в торфяниках по всему миру), и данные коллекций Полярно-Арктического ботанического сада и Всероссийской Коллекции Микроорганизмов.

Опубликованные на территорию страны находки грибов представлены поровну тремя типами: наблюдения (human observations), образцы коллекций (preserved specimen), молекулярные данные (material sample) и коллекции культур (living specimen). При этом, самой страной молекулярные данные почти не публикуются (7% от всех находок), а преобладают поровну данные наблюдений (47%) и данные образцов коллекций (43%) (рис. 1, Г).

На временной шкале хронологии наблюдений, 70% данных на территорию России имеют происхождение в 21 в. (тот же процент сохраняется и для данных из страны). До 2000 г. структура данных почти полностью (90%) состоит из оцифрованных образцов коллекций (для данных из страны поровну коллекции и наблюдения).

Интересно проанализировать происхождение данных на территории России по наборам опубликованных данных (включая иностранного происхождения) (табл. 1): первые 10 наборов данных включают 77% находок видов и представлены в следующем порядке: 1) глобальный микробиом почв (Эстония), 2) наблюдения платформы iNaturalist исследовательского уровня, 3) фунгарий Эстонского университета естественных наук (ТААМ), 4) оцифрованные находки из литературы для севера Западной Сибири, 5) финская база данных *Aphyllophorales externi*, 6) набор данных Полярно-Арктического ботанического сада (ПАБСИ), 7) фунгарий Тартуского университета (ТУФ), 8) немецкий гербарий *Herbarium Senckenbergianum* (FR), 9) набор данных по ревизии данных краудсорсинга, 10) оцифрованные находки из литературы для юга Западной Сибири. Что касается рейтинга наборов данных, формирующих массив данных из страны: первые 10 наборов данных включают 82% находок и представлены в следующем порядке: 1) оцифрованные находки из литературы для севера Западной Сибири, 2) набор данных ПАБСИ, 3) набор данных по ревизии данных краудсорсинга, 4) оцифрованные находки из литературы для юга Западной Сибири, 5) фунгарий Югорского государственного университета, 6) учеты на площадках в районе Ханты-Мансийска в лесах, 7) метабаркодинг болот в районе Ханты-Мансийска, 8) учеты на площадках в районе Ханты-Мансийска в болотах, 9) Всероссийская коллекция микроорганизмов (ВКМ), 10) учеты на маршрутах в районе Ханты-Мансийска.

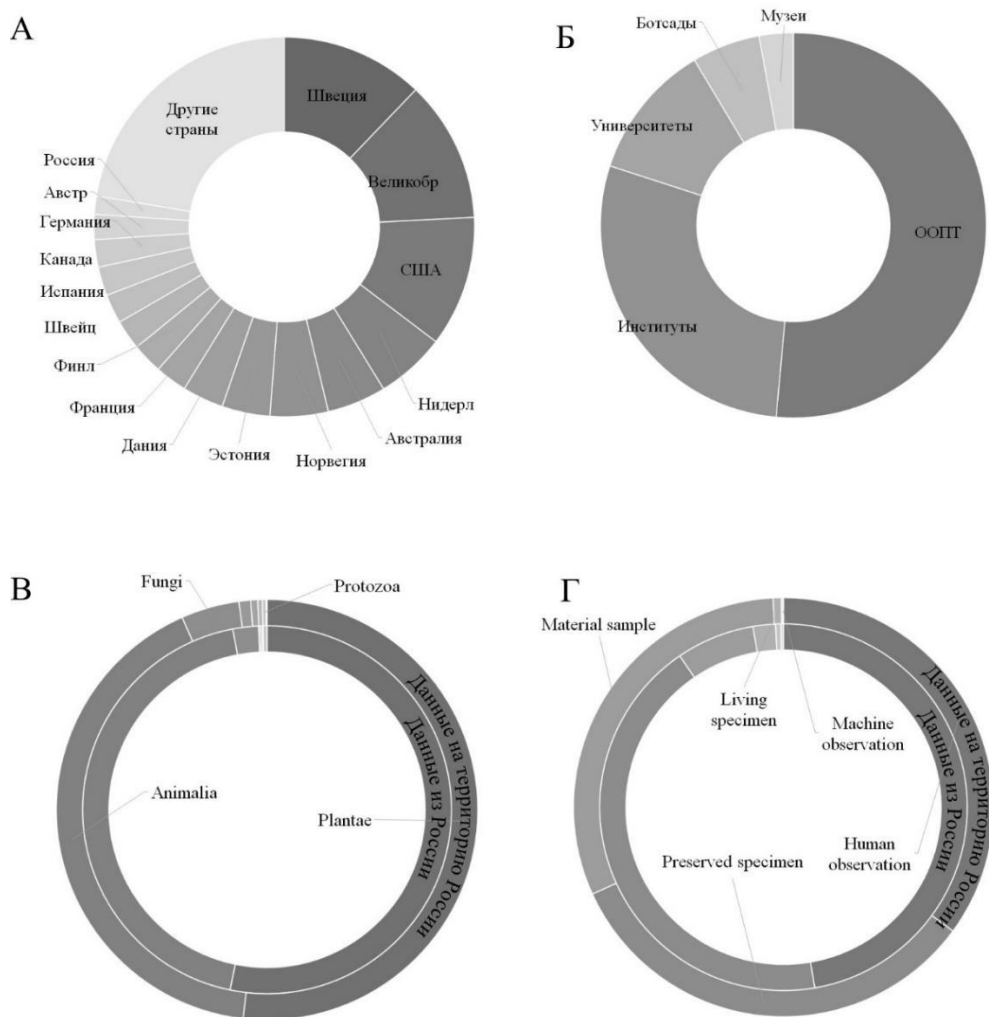


Рис. 1. Структура данных о находках грибов, опубликованных в GBIF: А – доля находок грибов по странам, Б – доля наборов данных о грибах, опубликованных в России организациями разного типа, В – доля находок из разных царств, опубликованных на территорию России и самой страной, Г – доля разных типов записей о грибах, опубликованных на территорию России и самой страной.

Всего из страны опубликовано 69 наборов данных. Около 40 наборов данных имеют менее 100 находок. Интересно провести классификацию всех наборов данных из России по характеру опубликованной в них информации. Для этого они были субъективно разбиты на 10 категорий: 1) оцифрованные коллекции, 2) оцифрованная литература, 3) данные учетов на площадках или маршрутные учеты, 4) данные любительских наблюдений, прошедшие ревизию, 5) молекулярные данные (метабаркодинг), 6) фенологические наблюдения, 7) инвентаризация (находки видов на определенной территории), 8) таксономическая ревизия, 9) находки охраняемых видов, 10) описания

растительности. По числу находок доминируют три типа: коллекции (35%), литературные данные (24%) и учеты на площадках (15%), а также распространены любительские наблюдения (9%) и данные метабаркодинга (7%). По числу наборов данных, больше всего опубликовано наборов по фенологическим наблюдениям (20 штук), 13 наборов по коллекциям, 11 по инвентаризации, 9 наборов данных по таксономическим ревизиям, остальные в меньшинстве.

Всего на территории страны зарегистрированы и публикуют данные о находках грибов 34 организации разного типа (рис. 1, Б): природоохранные территории (53%), научно-исследовательские инсти-

Таблица 1. Рейтинг из первого десятка наборов данных, опубликованных в GBIF на территорию страны (справа) и самой страной (слева)

Рейтинг наборов данных на территорию страны	N находок	%	Рейтинг наборов данных, опубликованных страной	N находок	%
Global soil organisms (EST)	180 890	28	Fungal literature records database of the Northern..	22 339	15
iNaturalist Research-grade Observations (RU)	158 340	24	L. IS dataset. Cyanoprocarvota, Lichens, Bryophyte	18 991	12
Estonian University of Life Sciences Institute of.. (RU)	31 482	5	CRIS data set	15 533	10
Fungal literature records database of the Northern.. (RU)	22 339	3	Crowdsourcing fungal biodiversity: revision of ina..	14 087	9
Aphyllorphorales externi (FI)	20 322	3	The Fungal Literature-based Occurrence Database fo..	11 493	8
L. IS dataset. Cyanoprocarvota, Lichens, Bryophyte (RU)	18 991	3	The Fungarium of Yugra State University	9 988	7
University of Tartu Natural History Museu... (EST)	18 562	3	Plot-based observations of terrestrial macrofungi..	9 794	6
Herbarium Senckenbergianum (FR) - Fungi (GERM)	15 898	2	DNA-based occurrence dataset on peatland fungal co..	9 749	6
CRIS data set (RU)	15 533	2	Plot-based observations of macrofungi in raised bo..	9 086	6
Crowdsourcing fungal biodiversity: revision of.. (RU)	14 087	2	VKM Microbial Cultures from Former USSR	2 807	2
Другие наборы данных (всего 322)	154 705	24	Другие наборы данных (всего 69)	28 216	19
Всего	651 149		Всего	152 083	

Примечание. Приводятся число находок и процент от общего числа находок. Названия источников даются в сокращенном виде.

туты (30%), университеты (11%), ботанические сады и музеи (по 3%). Первые две организации публикуют 90% всех находок: Югорский государственный университет (65%) и Полярно-Арктический ботанический сад-институт (24%) (табл. 2).

Для расширения списка используемых полей в GBIF предусмотрены расширения формата DwC, которые позволяют охватить больше параметров. Так, в наборах данных из России используется два расширения. «DNA derived data» используется только в четырех наборах данных, опубликованных Югорским государственным университетом и Центральным Сибирским Ботаническим Садам. Второе расширение «Measurement or Fact» — используется в двух наборах данных.

Важная информация о типовых образцах имеется только в одном

опубликованном наборе данных (ВКМ), всего 28 типовых образцов.

Находки, опубликованные в GBIF, могут содержать связанные с ними изображения. Всего 18% опубликованных находок, содержащиеся в 15-ти наборах данных, имеют изображения (25 тысяч образцов). Наиболее крупные из них: прошедшие ревизию данные платформы iNaturalist, коллекции трех фунгариев, и наборы данных о находках редких и охраняемых видов.

Отдельно стоит выделить наличие редких и охраняемых видов среди данных на территорию страны. Всего (в том числе иностранными организациями) опубликовано около 18 тысяч находок видов, отнесенных к одной из категорий МСОП: в том числе 16.5 тысяч находок LC, около 1 тысячи категории NT, и меньшая часть из

Таблица 2. Рейтинг российских организаций по числу опубликованных находок грибов в GBIF

Организация	Число находок	Процент от общего числа
Югорский государственный университет	91 611	65
Полярно-арктический ботанический сад-институт	34 524	24
Институт биологии Коми	3 933	3
Всероссийская коллекция микроорганизмов	2 807	2
Институт математических проблем биологии	1 332	1
Приокско-террасный биосферный заповедник	813	1
Центральный сибирский ботанический сад	811	1
Московский государственный университет	790	1
Природный парк «Кондинские озера»	560	0
Институт экологии растений и животных	473	0
Дарвиновский биосферный заповедник	450	0
Южный научный центр	388	0
Ботанический институт им. В.Л. Комарова	383	0
Уральский федеральный университет	336	0
Институт физико-химических и биологических проблем почвоведения	305	0
Государственный заповедник «Кивач»	282	0
Печоро-Илычский государственный заповедник	262	0
Ильменский государственный заповедник	197	0
Государственный заповедник «Малая Сосьва»	162	0
Государственный заповедник «Нургуш»	120	0
Сургутский государственный университет	115	0
Институт биологии внутренних вод	111	0
Центральный лесной государственный заповедник	106	0
Висимский государственный заповедник	105	0
Другие	572	0
Всего	141 584	

категорий VU, EN, DD, и CR. Из России эти данные опубликованы в 62 наборах данных, первые 10 охватывают около 90% всех находок, а именно: данные платформы iNaturalist, прошедшие ревизию; оцифрованные литературные находки; данные коллекций Фунгариев; данные учетов и мониторинга на площадках. Есть только один пример набора данных, целиком посвященного находкам видов, занесенных в региональную Красную Книгу Республики Коми (Degteva et al., 2021): всего в наборе данных около 7 тысяч находок, из них 1 тысяча – грибы.

Проблемы публикации открытых дан-

ных о находках грибов на территории России. В России отсутствует реестр коллекций грибов, поэтому сложно сделать прогноз скорости оцифровки и мобилизации данных наших коллекций. Учитывая, что грибы раньше рассматривались в царстве растений и традиционно хранились в ботанических коллекциях, можно сделать анализ по реестру Index Herbariorum Rossicum (<https://www.binran.ru/resources/current/herbaria/>). Метаданные о коллекциях на этом ресурсе хранятся в виде html страниц с перечнем около десятка полей. Поиск осуществляется только по городам и

регионам. Поэтому единственной возможностью извлечь информацию о коллекциях грибов является ручное перелистывание страниц и сбор данных из полей «структура гербария», «краткая информация о коллекции» и «краткие исторические сведения», и др. Прделав эту работу и интегрировав данные о коллекциях грибов и лишайников в виде таблицы, получаем следующее. Всего в Index Herbariorum Rossicum содержатся сведения о 255 гербарных коллекциях России, в том числе коллекций ботанических садов, институтов, заповедников, и пр. Из них образцы грибов упоминаются в 61 коллекции, лишайники в 66 коллекциях, а миксомицеты – в трех коллекциях. Из коллекций грибов, десять первых в рейтинге включают 90% от всех образцов в сумме, и хранят от 10 до 300 тыс. образцов. Среди коллекций лишайников, шесть первых в рейтинге коллекций охватывают около 95% всех образцов и содержат от 10 до 450 тыс. образцов. Коллекции миксомицетов хранят от 4 до 10 тысяч образцов. В сумме, если сложить все коллекции получается около 1.3 млн образцов грибов и миксомицетов, хранящихся в коллекциях на территории России.

Опираясь на эту приблизительную оценку, можно сделать вывод, что на сегодняшний день только 5% (61 тысяча образцов, т.е. находок типа «preserved specimen») коллекций грибов и грибоподобных организмов России оцифрованы и опубликованы в GBIF. При этом число оцифрованных образцов растений составляет около 1.2 млн, а животных – 380 тыс. Далее обсудим возможные причины запоздалой оцифровки и публикации данных о грибах из нашей страны по сравнению с другими группами организмов (растения, животные).

Глобальный пересмотр систематики грибов с появлением молекулярно-генетических методов. Вероятно, это отнимает время и затраты коллекций на ревизию отдельных групп в свете современной систематики. И наоборот, отсутствие

современной ревизии откладывает оцифровку коллекций из-за их «устаревшего» в контексте современной систематики статуса. Так или иначе, оцифровка коллекций не должна сопровождаться их ревизией и приведением их к современной систематике, а является технической процедурой, сопутствующей этому процессу. Наличие оцифрованных коллекций в открытых базах должно способствовать их ревизии благодаря обнаружимости национальных коллекций для современных международных проектов по филогенетической систематике разных групп грибов.

Сканы сухих образцов грибов не несут информацию об образце, как это у растений в гербариях. Систематическое положение грибов нельзя определить по эксикатам, а, значит, зачем сканировать? Однако, к образцам могут быть привязаны не только сканы, но полевые дневники с описанием витальных признаков, микрофотографии, фото свежих образцов, и др. Наконец, оцифровка не обязательно сопровождается сканированием образца. В большинстве случаев бывает достаточно информации, прочитанной с этикетки и ее скана.

В стране нет единого портала коллекций грибов, что затрудняет инициативы оцифровки разрозненными усилиями. В ряде стран существуют узлы-агрегаторы коллекций грибов, хранящихся в разных учреждениях (например, портал грибов в США <https://www.mycportal.org/portal/>). В то же время инициативы оцифровки определяются бюджетом и планом в рамках отдельных организаций. Агрегация полученных данных может произойти позже при условии, что в базах данных был соблюден единый формат данных (например, DwC). В случае региональных коллекций Зап. Сибири начата агрегация коллекций из разных учреждений с использованием программного обеспечения «Specify» и создан первый портал сибирских коллекций грибов (<http://mycportal.ugrasu.ru/>), который должен восполнить этот пробел.

Отсутствие программного обеспечения для создания баз данных коллекций и публикации данных в GBIF. На российском рынке программного обеспечения отсутствуют продукты, которые можно было бы «просто установить» на компьютер или сервер для создания баз данных коллекций. Небольшие коллекции не могут себе позволить разработку соответствующего продукта самостоятельно. Примеры зарубежных продуктов в этой области существуют (например, Symbiota <https://symbiota.org/>, Specify <https://www.specifysoftware.org/>, EarthCape <https://earthcape.com/>, и др.). В текущей политической ситуации покупка и установка иностранного программного обеспечения для организаций может быть затруднительна. Выходом может быть использование свободно распространяемых продуктов с открытым исходным кодом (например, Specify). С другой стороны, остается возможность прямой публикации данных в GBIF с помощью загрузки таблиц формата «csv».

Нет соответствующей финансовой поддержки. Так, гранты поддержки проектов в области фундаментальной науки не оказывают поддержку инициативам оцифровки, ссылаясь на технический аспект («а нужна наука»). Тем не менее, есть примеры оцифровки коллекций растений с целью изучения флор, поддержанные грантами РФФИ. Важным является наличие поддержки, направленной непосредственно на оцифровку как технического проекта. Например, субсидия Минобрнауки «Биоресурсные коллекции» в последние годы позволила закупить соответствующее оборудование, подготовить персонал и создать порталы коллекций в целом ряде организаций.

Данные любительских наблюдений грибов не так легко приспособить в микологии, поскольку по фотографиям нельзя сделать определение до вида. Большая часть данных о находках видов растений и животных происходит из порталов любительских наблюдений, таких как

iNaturalist.org. Поскольку грибы в большей мере нельзя определить по внешним признакам, то этот поток открытых данных будет меньше для грибов. С другой стороны, есть проекты по ревизии и анализу любительских наблюдений макромицетов с включением новых находок в чеклисты региона и дополнением ареалов многих надежно определяемых видов (Filipova et al., 2022).

Наборы данных в GBIF не являются результатом деятельности для отчетов по договорам и грантам. Наборы данных нельзя приравнивать к научным публикациям, поскольку они не проходят рецензирование и не индексируются в библиографических базах данных. По этой причине они не входят в стандартный перечень пунктов отчетности по научно-исследовательской работе. Однако в ряде институтов в качестве продукта учитывается разработка программного обеспечения и баз данных. Теоретически, набор данных, опубликованный в GBIF, можно зарегистрировать в системе Роспатент и получить свидетельство государственной регистрации на результат интеллектуальной деятельности (базу данных). В этом случае набор данных и сами данные должны иметь лицензию CC BY (атрибуция), которая позволяет свободно использовать и распространять базу данных с указанием автора, но при этом автор сохраняет права на базу данных. Это значит, что размещение базы в GBIF с такой лицензией не лишает нас права на регистрацию ее в Роспатенте. Наконец, самым верным решением на пути получения галочки в отчетности является публикация «статьи о данных» (data paper), посвященной структуре опубликованного набора и методике получения данных (такая статья приравнивается с обычной научной и индексируется).

Политика редколлегий журналов. Большую роль в публикации исходных данных на платформе GBIF играет политика журналов. Статьи формата аннотированных списков и инвентаризации территории

могут и должны содержать электронное приложение в виде набора данных, размещенных в GBIF. Сама статья может оставаться в прежнем формате аннотированного списка, или быть преобразована в статью о данных. Наличие аудита данных при этом обеспечивает их техническое качество перед публикацией в GBIF. К сожалению, некоторые микологические журналы в стране имеют прямо противоположное отношение: публикация данных в GBIF запрещается, поскольку это якобы нивелирует новизну. С другой стороны, в журнале «Микология и фитопатология» есть пример серии статей, где вместе со списком в классическом виде авторы размещают набор данных в GBIF (Rebriev et al., 2022). Важным представляется наличие в журналах специальных разделов для статей формата «статья о данных» (data paper), поощряющая публикацию наборов данных о находках видов в GBIF.

Субсидия из федерального бюджета на выполнение государственного задания

«Молекулярно-генетические методы в изучении и оценке состояния биоразнообразия Северных регионов (FENG-2024-0003)» № 1023041300017-6-1.6.4 от 13.03.2024.

СПИСОК ЛИТЕРАТУРЫ

- Degteva S., Bobretsov A., Bobrov Yu. et al.* Occurrences of threatened species included in the third edition of the Red data book of the Komi Republic (Russia) // *Biodiversity Data J.* 2021. V. 9. e73763. <https://doi.org/10.3897/BDJ.9.e73763>
- Bolshakov S., Kalinina L., Palomozhnykh E. et al.* Agaricoid and boletoid fungi of Russia: the modern country-scale checklist of scientific names based on literature data // *Biol. Communications.* 2021. V. 66 (4). P. 316–325. <https://doi.org/10.21638/spbu03.2021.404>
- Filippova N.V., Ageev D.V., Basov Yu.M. et al.* Crowdsourcing fungal biodiversity: revision of iNaturalist observations in Northwestern Siberia // *Nature Conservation Research.* 2022. V. 7 (1). P. 64–78. <https://doi.org/dx.doi.org/10.24189/ncr.2022.023>
- Rebriev Yu.A., Bogacheva A.V., Bulakh E.M. et al.* New species of macromycetes for regions of the Russian Far East. 3 // *Mikologiya i fitopatologiya.* 2022. V. 56 (4). P. 254–263. <https://doi.org/10.31857/S0026364822040080>

Open data on fungal collections and occurrences across Russia on the GBIF portal

N.V. Filippova^a

^a*Yugra State University, Khanty-Mansiysk, Russia*

The Global Biodiversity Information Facility (GBIF) integrates records of plant, animal, and fungal species at an international level. In Russia, active digitization and data mobilization efforts are ongoing in the fields of botany and zoology, while progress in mycology has been slower. This publication presents an analysis of open fungal records data published in GBIF for the territory of Russia and discusses the challenges and prospects for fungal data mobilization at the national level.

Keywords: biodiversity informatics, collections, digitization, FAIR, funga, mycota, open data