

ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ДАННЫХ ГЛОБАЛЬНОГО ПОРТАЛА О БИОРАЗНООБРАЗИИ GBIF В ЭКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

© 2021 г. Н. В. Иванова^{а, *}, М. П. Шашков^{а, б}

^аИнститут математических проблем биологии — филиал Института прикладной математики им. М.В. Келдыша РАН, Россия 142290 г. Пушино, ул. проф. Виткевича, 1

^бИнститут физико-химических и биологических проблем почвоведения РАН — ФИЦ ПНЦБИ РАН, Россия 142290 г. Пушино, ул. Институтская, 2

*e-mail: Natalya.dryomys@gmail.com

Поступила в редакцию 14.04.2020 г.

После доработки 23.05.2020 г.

Принята к публикации 04.06.2020 г.

Глобальный портал GBIF является крупнейшим источником открытых научных данных о биоразнообразии России — на май 2020 г. стало доступно >4.3 млн записей о находках видов, среди них наиболее многочисленны сведения о находках сосудистых растений (>1.9 млн) и встречах птиц (>900 тыс.). Репрезентативность данных для остальных таксономических групп и большинства регионов страны пока остается низкой, хотя объем доступной информации продолжает увеличиваться за счет мобилизации данных научных коллекций и личных архивов исследователей, а также систем любительских наблюдений iNaturalist и eBird. Доступные через GBIF сведения представляют большой интерес для экологов, работающих с методами моделирования пространственного распространения видов. Однако существующие пробелы в данных для территории России ограничивают возможности их повторного использования, а для получения корректных результатов необходимо привлечение дополнительных источников информации.

Ключевые слова: информатика биоразнообразия, FAIR-данные, GBIF, анализ объединенных данных

DOI: 10.31857/S0367059721010066

Развитие информационных технологий привело к появлению нового подхода в проведении научных исследований, в рамках которого решение прикладных и фундаментальных задач основано не на результатах новых наблюдений или экспериментов, а на анализе массивов ранее собранных данных, полученных из разных источников. Это позволяет не только проводить исследования макрорегионального и глобального охвата, но и повышает их эффективность за счет повторного использования данных [1]. В области биологического разнообразия существует множество глобальных и региональных проектов, обобщающих сведения о распространении видов и предоставляющих к ним открытый доступ [2–6]. В России инициативы, направленные на обобщение и/или анализ объединенных данных, пока единичны [7], хотя в последние годы объем открытых данных о биоразнообразии значительно возрос [8], преимущественно в результате оцифровки отечественных научных биологических коллекций [9–11].

В настоящей работе описаны возможности использования данных глобального портала GBIF в

экологических исследованиях с учетом представленности информации для территории РФ.

МИРОВЫЕ ТЕНДЕНЦИИ В ОБЛАСТИ УПРАВЛЕНИЯ ЦИФРОВЫМИ ДАННЫМИ О БИОРАЗНООБРАЗИИ

По мере развития информационных технологий и накопления опыта работы с цифровыми данными о биоразнообразии в научной среде формируется представление о том, что исходные полевые данные сами по себе являются научным продуктом. Будучи собранными один раз, они должны быть многократно использованы для анализа в составе объединенных массивов. В рамках этого подхода широкое развитие получила FAIR-концепция [12], согласно которой массивы данных должны быть легко обнаружимыми через Интернет (Findable), доступными для исследователей (Accessible), совместимыми (Interoperable) друг с другом и используемыми многократно (Reusable). FAIR-концепция предполагает, что данные необходимо хранить в надежных репозиториях, в

Таблица 1. Крупнейшие международные ресурсы о биоразнообразии (данные по состоянию на 02.05.2020 г.)

Информационная система	Общее число записей	Из них для территории России	Экспорт данных в GBIF
GBIF https://www.gbif.org	1410467770	4328442	
iDigBio — портал цифровых научных коллекций https://www.idigbio.org	121428342	171310	Да
eBird — сеть любительских наблюдений за птицами https://ebird.org	561852542	171034	Да
OBIS — информационная система о распространении морских видов https://obis.org	58920363	57344	Да
Сеть любительских наблюдений iNaturalist https://www.inaturalist.org	36418528	623050	Да
VertNet — информационная система о позвоночных животных http://www.vertnet.org	21580767	17989	Нет
Canadensys — информационная система о канадских биологических коллекциях https://community.canadensys.net	5816085	4499	Да
Edaphobase — почвенно-зоологическая информационная система https://portal.edaphobase.org	373877	1383	Да

формате, понятном не только компьютеру, но и человеку, однозначно указывать их авторов и правила повторного использования (лицензию).

Разработка стандартов, методов и инструментов для оцифровки, хранения и управления данными о биоразнообразии является динамично развивающейся областью мировой науки [13, 14], получившей название “информатика биоразнообразия” (Biodiversity informatics) [15]. Одна из ключевых задач, решаемых в этом направлении, — мобилизация накопленных данных о биоразнообразии, т.е. обеспечение свободного доступа к ним через сеть Интернет.

Крупнейшим ресурсом, предоставляющим доступ к открытым интегрированным данным о находках биологических видов, сегодня является Глобальный портал о биоразнообразии GBIF.org (Global Biodiversity Information Facility) [3, 14, 16]: в мае 2020 г. было доступно >1.4 млрд. записей о находках >4.5 млн видов, организованных в >52 тыс. наборов данных. GBIF — это распределенная инфраструктура, состоящая из сети взаимосвязанных национальных “узлов” (national nodes) стран-участниц и таких крупных международных сообществ, как Европейское агентство по окружающей среде (European Environment Agency), рабочая группа Сохранение флоры и фауны Арктики (Conservation of Arctic Flora and Fauna, CAFF) и др. Благодаря использованию общих стандартов представления данных, портал GBIF

индексирует содержимое многих тематических информационных систем (табл. 1).

Через портал GBIF для повторного использования доступны метаданные (структурированные описания) наборов данных (resource metadata), таксономические списки для определенных территорий или групп организмов (checklist data), данные о находках видов (occurrence data), а также данные описаний, учетов или сборов, выполненных на пробных площадях, маршрутах или трансектах (sampling-event data). Интерес для экологов представляют прежде всего сведения о находках видов, имеющие точные географические координаты, содержащиеся в occurrence и sampling-event наборах данных. Сведения о методах проведения исследований и оценке обилия видов, обнаруженных на учетных площадях (маршрутах, трансектах), доступны только в sampling-event наборах данных, содержащих первичные сведения каждого учета как единого целого, что открывает возможности для последующего сопоставления результатов, полученных одинаковыми методами на разных территориях и/или в разное время. Сегодня sampling-event данные еще мало востребованы для комплексного повторного анализа. Как правило, исследователи используют только сведения о точках находок видов, учтенных в сборах или описаниях. На наш взгляд, это связано с тем, что sampling-event формат представления данных в GBIF появился всего несколько лет назад, и объем опубликованной информации (всего чуть

менее 1.5 тыс. наборов данных, или ~3% от их общего числа) пока не позволяет проводить корректные сопоставления результатов. Например, набор данных о растительности Дании [20] содержит 675000 геоботанических описаний, выполненных с 1868 г. по 2017 г. С момента его публикации в GBIF эти данные использованы в 106 научных статьях, посвященных в основном анализу экологических особенностей отдельных видов.

Открытая публикация первичных данных в GBIF способствует повышению эффективности научных исследований, создавая условия для оптимизации анализа и оценки качества данных, делает возможным проверку и оценку результатов исследований, тем самым противодействуя фальсификациям. Правила использования всех опубликованных данных однозначно описываются лицензиями Creative Commons. Каждому опубликованному в GBIF набору данных присваивается уникальный идентификатор цифрового объекта — DOI (Digital Object Identifier), что позволяет защитить результаты интеллектуальной деятельности авторов.

ИСПОЛЬЗОВАНИЕ ДАННЫХ ПОРТАЛА GBIF В НАУЧНЫХ ИССЛЕДОВАНИЯХ

Согласно статистике глобального портала, с 2003 г. в рецензируемых научных журналах вышло 4429 статей, содержащих результаты анализа объединенных данных, полученных через GBIF, и число таких публикаций возрастает с каждым годом (рис. 1). Начиная с 2017 г. каждый день в мире выходит в среднем 2 научные статьи, основанные на анализе данных GBIF. Ежегодно Секретариат GBIF публикует обзоры, в которых представлены наиболее значимые статьи.

Нами проанализированы публикации из обзоров 2017–2019 гг. (всего более 240 статей) [18–20]. Обнаружено, что в работах экологической тематики данные GBIF чаще всего используются для моделирования пространственного распространения отдельных видов или групп видов. Современные модели позволяют учитывать не только абиотические факторы, но и конкуренцию между видами, а также комплекс других биотических взаимодействий, что делает их мощным инструментом анализа экологических данных [21 и цитируемые работы]. Сведения портала GBIF позволяют проводить такие исследования не только на региональном, но и глобальном уровнях благодаря большому объему доступных данных о точках находок видов. Например, D. Righetti et al. [22] построены глобальные модели видовой разнообразия сообществ фитопланктона, учитывающие влияние температуры и некоторых других абиотических факторов. Основой исследования стали 1056363 записи о находках 1298 видов, по-

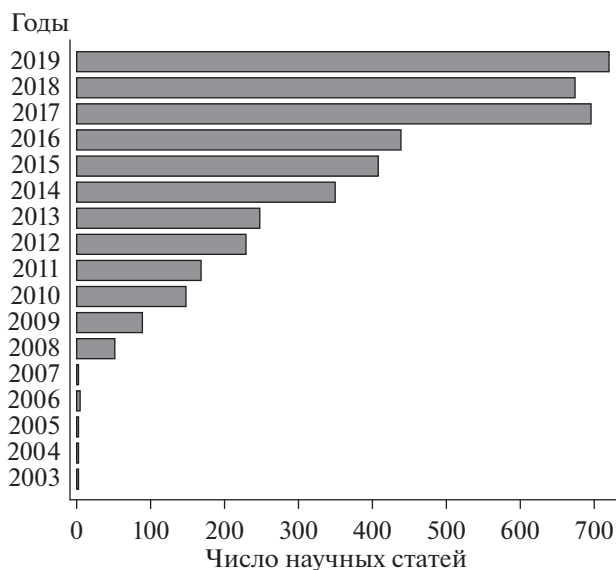


Рис. 1. Число научных статей, основанных на анализе данных, полученных через портал GBIF, опубликованных в рецензируемых журналах. По данным [20] и портала GBIF.

лученных из открытых источников, в т.ч. портала GBIF и информационной системы OBIS [2].

Построение моделей распространения видов на основе данных GBIF позволяет решать актуальные задачи в различных областях экологической науки. Важным направлением является оценка влияния глобальных изменений климата на распространение видов. Так, M.K. Dyderski et al. [23] с использованием более 1.8 млн записей GBIF построены модели распространения 12 европейских видов деревьев при разных сценариях изменения климата и выделены виды, ареалы которых могут значительно сократиться. B.R. Ribeiro et al. [24] разработали стратегию сохранения млекопитающих в бассейне р. Амазонка с учетом изменений климата. На основе данных GBIF ими построены модели экологических ниш 256 видов и выявлены территории с высоким и низким рисками исчезновения целевых видов.

GBIF является важным источником данных для модельного изучения факторов, способствующих расселению инвазивных видов. Например, C.S. Jarnevich et al. [25] исследованы факторы, влияющие на распространение инвазивного злака *Cenchrus ciliaris* L. в глобальном масштабе и в региональном — в национальном парке Сагуаро (Аризона, США). Выяснено, что глобальное распространение в большей степени лимитируется зимними температурами, в то время как на локальном уровне более важен фактор крутизны склона. L. Cardador и T.M. Blackburn [26] по данным GBIF построены модели распространения 776 чужеродных видов птиц и показано, что бли-

зость населенных пунктов является важным стимулом расширения их ареалов.

Особенно актуальна сегодня практическая задача в области экологии человека, которую позволяют решить данные GBIF, — это оценка рисков распространения опасных заболеваний на основе модельных прогнозов ареалов их переносчиков. Например, С.Л. Moyes et al. [27] показано, что риски заражения малярией в Малайзии выше в районах с интенсивным лесопользованием. D.M. Piggott et al. [28] на основе модельной оценки распространения переносчиков лихорадки Эбола в Западной Африке выделили страны, наиболее подверженные риску заражения.

Очевидно, что все эти исследования были бы невозможны без использования объединенных данных портала GBIF. В то же время надежность полученных результатов зависит от объема доступных сведений о находках видов, а репрезентативность данных GBIF значительно различается для разных регионов [16]. Поэтому на следующем этапе нами проанализирована представленность информации, доступной через глобальный портал для территории России.

ДААННЫЕ О БИОРАЗНООБРАЗИИ РОССИИ, ДОСТУПНЫЕ ЧЕРЕЗ GBIF

На 2 мая 2020 г. в GBIF для территории России доступны сведения о 4328442 находках видов, из которых 2523014 записей (58.3%) опубликованы отечественными исследователями, 3677733 записи (84.9%) имеют географические координаты. Значительный объем данных происходит из США (606412 записи), Великобритании (229330), Эстонии (177865) и Нидерландов (111889).

Число записей о находках растений (Plantae) несколько выше, чем животных (Animalia), — 49.4 и 41.7% соответственно (см. табл. 1). Сведения о сосудистых растениях (Tracheophyta) составляют 93.1% от всех данных о Plantae, 1593964 записи имеют географические координаты. Из 10 крупнейших наборов данных о сосудистых растениях России 8 опубликованы отечественными исследователями. Более половины находок сосудистых растений (55.0%) происходят из четырех наборов данных: 1) “Цифровой гербарий МГУ им. М.В. Ломоносова” — 376127 геопривязанных записей [29]; 2) “Наблюдения “исследовательского” уровня системы iNaturalist” [6] — 247108 геопривязанных записей [30]; 3) “Находки видов из “Флоры Сибири” (1987–1997 гг.)” — 169854 геопривязанные записи [31]; 4) “Сеточное картографирование сосудистых растений Владимирской области” — 123054 геопривязанные записи [32].

Данные о растениях в формате sampling-event для территории России весьма ограничены. Отечественными исследователями опубликованы

всего 4 набора данных этого типа, 3 из которых представляют собой геоботанические описания. В то же время в Европейском архиве данных о растительности EVA [33] хранится >46000 геоботанических описаний российских исследователей. Доступ к ним возможен только с разрешения авторов данных. Публикация этих описаний в открытом репозитории GBIF могла бы способствовать их более интенсивному повторному анализу и значительно повысить репрезентативность информации в масштабах страны.

Анализ пространственного распределения данных о находках сосудистых растений показал, что более высокая плотность записей отмечается в центре Европейской России: в Москве (24.6 записи/км²), Владимирской (5.3) и Московской (2.7) обл. (рис. 2а). По мере удаления от Москвы плотность записей уменьшается, но затем несколько увеличивается в южных субъектах РФ. Так, в Республике Адыгея она составляет 1.3 записи/км², в Республике Крым — 1.2, в г. Севастополе — 14.5. В азиатской части страны плотность записей низкая: в Республике Алтай отмечено 0.4 записи/км², в Бурятии — 0.3, в Приморском крае — 0.2, еще меньше в других субъектах Урала, Сибири и Дальнего Востока. В целом в 76 из 85 субъектов РФ плотность записей о находках сосудистых растений на 1 км² составляет <1. При этом важно отметить, что по абсолютному числу записей объем информации для азиатской части России весьма значителен и составляет 563638 (34.3% от всех имеющихся).

Чуть более половины (51.9%) данных о находках животных на территории России, доступных через GBIF, происходит из зарубежных источников (табл. 2). Среди них наиболее представлены данные о встречах птиц (Aves), 71.6% которых опубликованы российскими исследователями. Крупнейший набор данных — “Онлайн дневники наблюдений птиц” Зоологического музея МГУ им. М.В. Ломоносова [34], содержащий 331393 записи с географическими координатами находок птиц в России. Значительный объем данных происходит из международных сетей любительских наблюдений eBird [6, 33] и iNaturalist [28], а также “Атласа гнездящихся птиц Европы” [35]. Таким образом, доступные через GBIF данные о встречах птиц в России собраны в основном любителями, а не профессиональными орнитологами. В то же время российскими учеными накоплен большой объем данных учетов птиц. В большинстве своем эти сведения остаются неоцифрованными и хранятся в личных архивах. В GBIF российскими исследователями опубликовано только 3 набора первичных данных учетов птиц.

Анализ пространственного распределения доступных сведений о встречах птиц показал, что находки в основном приурочены к крупным городам и густонаселенным регионам, что объясня-

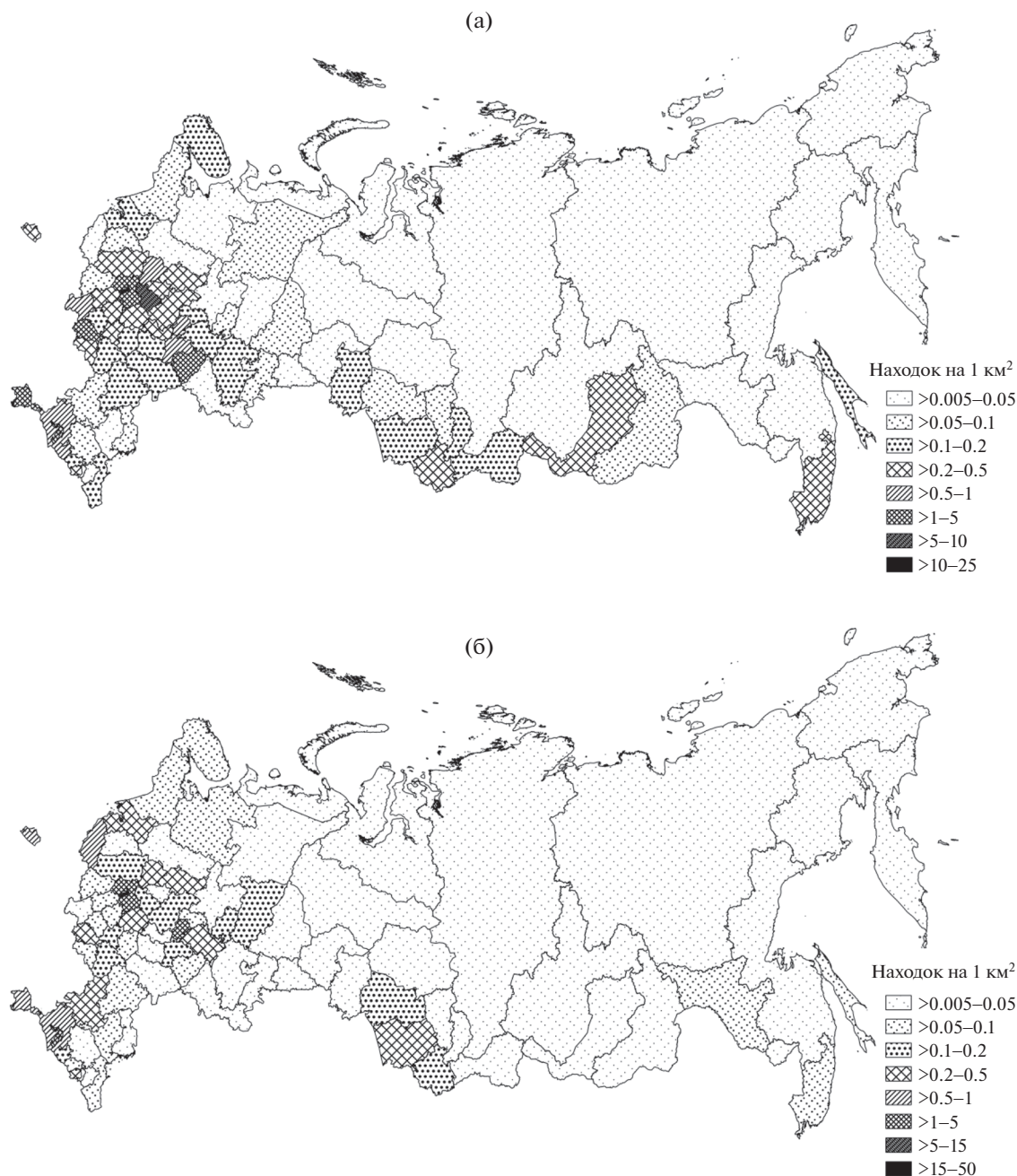


Рис. 2. Плотность записей о находках сосудистых растений (а) и птиц (б), доступных через глобальный портал о биоразнообразии GBIF для территории России.

ется большой долей любительских наблюдений (рис. 2б). Наибольшая плотность записей отмечена в Москве (45.7 встреч/км²) и Санкт-Петербурге (12.8), а также в Московской обл. (3.1). В остальных субъектах РФ, за исключением Республик Адыгея (1.7) и Чувашия (1.5), плотность записей <1 на 1 км². По абсолютному числу записей, помимо Московской обл. (135686 записи) и Москвы (117064), выделяются также Алтайский (54163) и Краснодарский (40320) края.

Второй по представленности данных группой позвоночных животных в GBIF для России являются земноводные (Amphibia) – 82.5% записей происходят из набора данных “Земноводные бывшего СССР” [37], опубликованного Институтом проблем экологии и эволюции им. А.Н.Северцова РАН (32561 геопривязанная запись). Доступные в GBIF сведения о находках других групп позвоночных животных для территории России весьма ограничены.

Таблица 2. Таксономический охват данных, доступных через портал GBIF.org, для территории России по состоянию на 02.05.2020 г.

Таксономическая группа	Число записей, доступных через GBIF.org		
	всего	из них для территории России	из них опубликовано российскими организациями
Animalia	1065586600	1805263	867334
В том числе:			
Chordata:	915116993	1302217	774446
Aves	853035930	985441	705349
Actinopterygii	30525833	58399	23403
Mammalia	16762112	207254	7934
Reptilia	6380942	6031	1151
Amphibia	5747325	39456	36469
Arthropoda:	123798598	370576	76883
Insecta	109844295	206603	15414
Malacostraca	4778105	20508	1227
Arachnida	3977520	26991	16885
Collembola	487139	44233	42870
Sarcoptiformes	225474	10585	9066
Plantae	287758965	2138776	1574686
В том числе:			
Tracheophyta	271832949	1992528	1461647
Bryophyta	9628695	102868	83818
Marchantiophyta	2309973	34144	28449
Fungi	19389961	178073	63490
Bacteria	14533384	34318	2531
Chromista	12577963	103565	265
Protozoa	1142534	29569	12499
Archaea	266318	328	0
Viruses	43869	407	0
incertae sedis	9168176	38144	2209

Примечание. Названия таксономических групп приводятся в соответствии с таксономическим справочником GBIF — GBIF Backbone [39], порядок — по убыванию общего числа записей.

Число записей о беспозвоночных невелико (см. табл. 2). Обращает на себя внимание крайне низкое число записей о находках насекомых (Insecta). Основная часть доступных сведений происходит из системы iNaturalist (22.8% записей) и зарубежных коллекций. Данные отечественных исследователей составляют лишь 7.5% записей и представлены в основном материалами Летописей природы российских заповедников, оцифрованными в рамках международного проекта “Летопись природы Евразии” [7].

Большой интерес для повторного анализа представляют опубликованные российскими исследователями первичные данные о сборах почвенных беспозвоночных — ногохвосток (Collembola), панцирных клещей (Oribatida), дождевых червей (Lumbricidae) и моллюсков (Gastropoda).

Доступные в GBIF данные о находках видов остальных крупных таксономических групп на территории России в основном происходят из зарубежных источников (см. табл. 2).

В целом, несмотря на значительный объем доступных через GBIF данных, репрезентативность сведений для большинства таксономических групп и регионов РФ остается низкой. На карте глобального биоразнообразия плотность находок видов в России составляет 0.24 записи/км², в то время как в странах с высоким уровнем оцифровки данных о биоразнообразии она значительно выше. Например, в Швеции на 1 км² карты GBIF приходится 144.4 записи, в Германии — 105.8, в

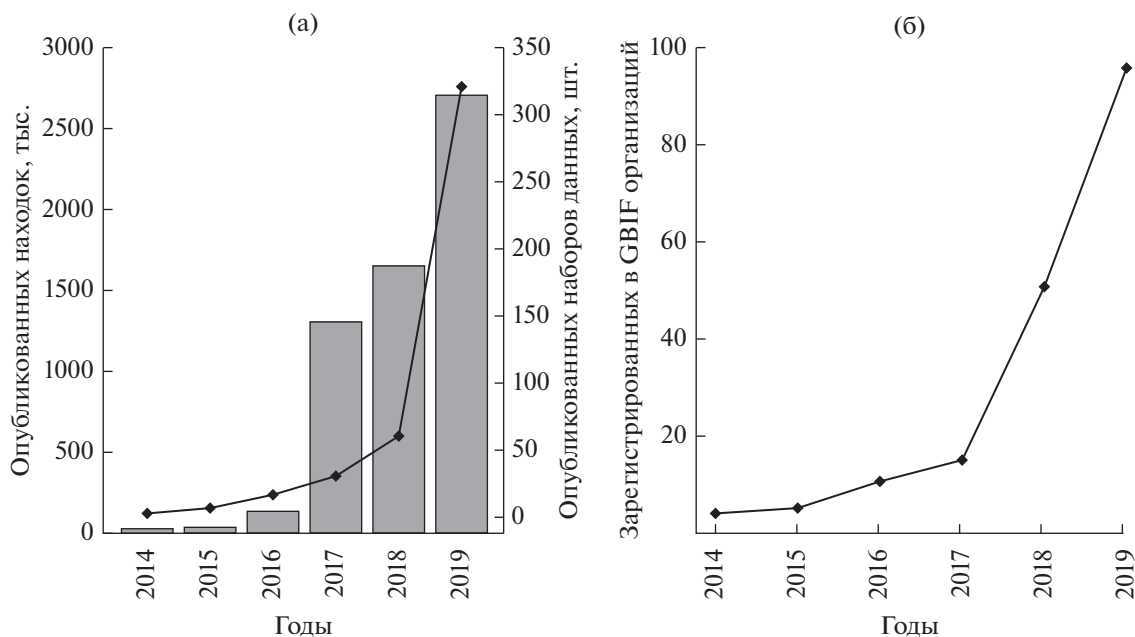


Рис. 3. Мобилизация данных о биоразнообразии через глобальный портал GBIF российскими исследователями: а — публикационная активность, б — динамика роста числа российских участников (publishers) сети GBIF.

Канаде, сопоставимой с Россией по территории и с меньшей плотностью населения, — 7.1. Недостаток данных затрудняет их повторный анализ и может привести к неверной интерпретации получаемых результатов.

В то же время с 2014 г. наблюдается непрерывный рост числа зарегистрированных в GBIF российских организаций и опубликованных ими наборов данных (рис. 3), поэтому стоит ожидать дальнейшего увеличения объема доступной информации. На май 2020 г. российскими исследователями опубликовано 344 набора данных (3016696 записей): 28 таксономических списков, 272 набора данных о находках видов и 39 sampling-event наборов данных, охватывающих 190 стран и территорий.

Безусловно, при отсутствии национальной информационной системы портал GBIF сегодня является крупнейшим, хотя и не исчерпывающим источником данных о биоразнообразии России. Заполнению существующих пробелов может способствовать оцифровка отечественных научных биологических коллекций. По нашей оценке [38], в российских гербарных коллекциях хранится не менее 15 млн листов (из них ~6 млн в фондах БИН РАН и ~1 млн в МГУ им. М.В. Ломоносова), а в зоологических — более 65 млн образцов. К настоящему времени оцифровано и доступно только чуть более 1% от всех имеющихся образцов. Мобилизация еще не оцифрованных данных не только позволит сохранить научное наследие, накопленное многими поколениями отечественных

ученых, но и повысит востребованность коллекций (в т.ч. региональных).

Большой вклад в пополнение данных GBIF вносят проекты по сбору любительских наблюдений (iNaturalist, eBird и др.). Возможности и ограничения использования таких данных в научных исследованиях неоднократно обсуждались в литературе [6]. По нашему опыту надежность этого источника информации определяется квалификацией экспертов, подтвердивших определение видов.

Таким образом, открытые данные портала GBIF, безусловно, необходимо использовать в анализе объединенных массивов. Тем не менее для корректной интерпретации результатов важно учитывать их неполноту для многих таксономических групп и регионов России.

ЗАКЛЮЧЕНИЕ

Объединенные данные о распространении видов становятся сегодня важным (а иногда основным) источником информации в исследованиях биоразнообразия. Они позволяют проводить на макрорегиональном уровне исследования ключевых функций экосистем с использованием современных методов статистического и пространственного анализа. Несмотря на то, что объем данных, доступных через глобальный портал GBIF для территории России, весьма значителен и составляет >4.3 млн записей, репрезентативность информации для большинства таксономи-

ческих групп и регионов страны остается низкой. Это затрудняет повторное использование данных в научных исследованиях и может привести к неверной интерпретации получаемых результатов. Для заполнения существующих пробелов в данных необходимо дальнейшее развитие проектов, направленных на оцифровку и мобилизацию научных биологических коллекций и личных архивов исследователей.

СПИСОК ЛИТЕРАТУРЫ

1. *Newman M.E.J.* Coauthorship networks and patterns of scientific collaboration // *Proceedings of the National Academy of Sciences*. 2004. V. 101. P. 5200–5205. <https://doi.org/10.1073/pnas.0307545100>
2. *Grassle J.F., Stocks K.I.* A Global Ocean Biogeographic Information System (OBIS) for the Census of Marine Life // *Oceanography*. 1999. V. 12(3). P. 12–14. <https://doi.org/10.5670/oceanog.1999.03>
3. *Wheeler Q.D.* What if GBIF? // *Bioscience*. 2004. V. 54. P. 717. [https://doi.org/10.1641/0006-3568\(2004\)054\[0718:WIG\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0718:WIG]2.0.CO;2)
4. *Constable H., Guralnick R., Wieczorek J.* et al. VertNet: A New Model for Biodiversity Data Sharing // *Plos ONE*. 2010. V. 8. № 2. P. E1000309. <https://doi.org/10.1371/journal.pbio.1000309>
5. *Burkhardt U., Russell D.J., Decker P.* et al. The Edaphobase project of GBIF-Germany — A new online soil-zoological data warehouse // *Applied Soil Ecology*. 2014. V. 83. P. 3–12. <https://doi.org/10.1016/j.apsoil.2014.03.021>
6. *Chandler M., See L., Copas K.* et al. Contribution of citizen science towards international biodiversity monitoring // *Biological Conservation*. 2017. V. 213. P. 280–294. <https://doi.org/10.1016/j.biocon.2016.09.004>
7. *Ovaskainen O., Meyke E., Lo K.* et al. Chronicles of nature calendar, a long-term and large-scale multitaxon database on phenology // *Scientific Data*. 2020. V. 7(47). <https://doi.org/10.1038/s41597-020-0376-z>
8. *Филиппова Н.В., Филиппов И.В., Щигель Д.С.* и др. Информатика биоразнообразия: мировые тенденции, состояние дел в России и развитие направления в Ханты-Мансийском автономном округе // *Динамика окружающей среды и глобальные изменения климата*. 2017. Т. 8. № 2. С. 46–56. <https://doi.org/10.17816/edgcc8246-56>
9. *Серегин А.П.* Цифровой гербарий МГУ — крупнейшая российская база данных по биоразнообразию // *Изв. РАН. Серия биол.* 2017. № 6. С. 30–36. <https://doi.org/10.7868/S0002332917060042>
10. *Калякин М.В.* Итоги и перспективы развития коллекционного дела в МГУ имени М.В. Ломоносова: зоологические коллекции // *Зоологические исследования*. 2018. № 20. С. 52–29.
11. *Kovtonyuk N.K., Han I.V., Gatilova E.A.* Digitization of vascular plant herbarium collections at the Central Siberian Botanical Garden, Novosibirsk, Russia // *Skvortsovia*. 2018. V. 4(3). P. 100–111.
12. *Wilkinson M.D., Dumontier M., Aalbersberg I.J.* et al. The FAIR Guiding Principles for scientific data management and stewardship // *Scientific Data*. 2016. V. 3. <https://doi.org/10.1038/sdata.2016.18>
13. *Pennisi E.* Boom in digital collections makes a muddle of management // *Science*. 2005. V. 308. P. 187–189. <https://doi.org/10.1126/science.308.5719.187>
14. *Edwards J.L.* Research and societal benefits of the Global Biodiversity Information Facility // *Bioscience*. 2004. V. 54. P. 485–486. [https://doi.org/10.1641/0006-3568\(2004\)054\[0486:RASBOT\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0486:RASBOT]2.0.CO;2)
15. *Bisby F.A.* The quiet revolution: Biodiversity informatics and the Internet // *Science*. 2000. V. 289. P. 2309–2312. <https://doi.org/10.1126/science.289.5488.2309>
16. *Yesson C., Brewer P.W., Sutton T.* et al. How global is the Global Biodiversity Information Facility? // *PLOS ONE*. 2007. V. 2(11). P. e1124. <https://doi.org/10.1371/journal.pone.0001124>
17. *Hennekens S.* Dutch Vegetation Database. Alterra, Wageningen UR. Sampling event dataset. 2018. <https://doi.org/10.15468/ksqxep> (22.03.2020).
18. GBIF Science Review 2017. Copenhagen: GBIF Secretariat, 2017. 55 p. <https://doi.org/10.15468/EAH0-3E08>
19. GBIF Science Review 2018. Copenhagen: GBIF Secretariat, 2018. 47 p. <https://doi.org/10.15468/VA9B-3048>
20. GBIF Science Review 2019. Copenhagen: GBIF Secretariat, 2019. 39 p. <https://doi.org/10.15468/QXXG-7K93>
21. *Санданов Д.В.* Современные подходы к моделированию разнообразия и пространственному распределению видов растений: перспективы их применения в России // *Вестн. Томского гос. ун-та. Биология*. 2019. № 46. <https://doi.org/10.17223/19988591/46/5>
22. *Righetti D., Vogt M., Gruber N.* et al. Global pattern of phytoplankton diversity driven by temperature and environmental variability // *Science Advances*. 2019. V. 5. № 5. eaau6253. <https://doi.org/10.1126/sciadv.aau6253>
23. *Dyderski M.K., Paź S., Frelich L.E.* et al. How much does climate change threaten European forest tree species distributions? // *Global Change Biology*. 2018. V. 24. I. 3. P. 1150–1163. <https://doi.org/10.1111/gcb.13925>
24. *Ribeiro B.R., Sales L.P., Loyola R.* Strategies for mammal conservation under climate change in the Amazon // *Biodiversity Conservation*. 2018. V. 27. P. 1943–1959. <https://doi.org/10.1007/s10531-018-1518-x>
25. *Jarnevich C.S., Young N.E., Talbert M.* et al. Forecasting an invasive species' distribution with global distribution data, local data, and physiological information // *Ecosphere*. 2018. V. 9. I. 5. e02279. <https://doi.org/10.1002/ecs2.2279>
26. *Cardador L., Blackburn T.M.* Human-habitat associations in the native distributions of alien bird species // *Journal of Applied Ecology*. 2019. V. 56. P. 1189–1199. <https://doi.org/10.1111/1365-2664.13351>

27. Moyes C.L., Shearer F.M., Huang Z. et al. Predicting the geographical distributions of the macaque hosts and mosquito vectors of *Plasmodium knowlesi* malaria in forested and non-forested areas // *Parasites & Vectors*. 2016. V. 9. <https://doi.org/10.1186/s13071-016-1527-0>
28. Pigott D.M., Millea A.I., Earl L. et al. Updates to the zoonotic niche map of Ebola virus disease in Africa // *eLife*. 2016. V. 5. e16412. <https://doi.org/10.7554/eLife.16412>
29. Seregin A. Moscow University Herbarium (MW). Occurrence dataset. Version 1.124. Lomonosov Moscow State University, 2020. <https://doi.org/10.15468/cpnhcc> (22.03.2020).
30. iNaturalist Research-grade Observations. Occurrence dataset. iNaturalist.org, 2020. <https://doi.org/10.15468/ab3s5x> (22.03.2020).
31. Artemov I., Egorova A. Locations of plants on dot distribution maps in the Flora of Siberia (Flora Sibiraea, 1987–1997). Occurrence dataset. Central Siberian Botanical Garden SB RAS, 2020. <https://doi.org/10.15468/jb84wg> (22.03.2020).
32. Seregin A.P. A grid-based database on vascular plant distribution in the Meshchera National Park, Vladimir Oblast, Russia. Occurrence dataset. Lomonosov Moscow State University, 2014. <https://doi.org/10.15468/ahunho> (22.01.2019).
33. Chytrý M., Hennekens S.M., Jiménez-Alfaro B. et al. European Vegetation Archive (EVA): an integrated database of European vegetation plots // *Applied Vegetation Science*. 2016. V. 19. P. 173–180. <https://doi.org/10.1111/avsc.12191>
34. Ukolov I., Kalyakin M., Voltzit O. RU-BIRDS.RU, Birds observations database from Russia and neighboring regions. Occurrence dataset. Zoological Museum of M.V. Lomonosov Moscow State University. Version 1.4. Lomonosov Moscow State University, 2019. <https://doi.org/10.15468/5cjsx70> (22.03.2020).
35. Levatich T., Padilla F. EOD — eBird Observation Dataset. Cornell Lab of Ornithology. Occurrence dataset. 2019. <https://doi.org/10.15468/aomfmb> (22.03.2020).
36. Hagemeyer W., Blair M., Loos W. EBCC Atlas of European Breeding Birds. Occurrence dataset. Version 1.3. European Bird Census Council (EBCC), 2016. <https://doi.org/10.15468/adtfvf> (22.03.2020)
37. Petrosyan V., Kuzmin S. Amphibians of the Former USSR. Occurrence dataset. Version 1.11. A.N. Severtsov Institute of Ecology and Evolution, RUSSIAN ACADEMY OF SCIENCES, 2017. <https://doi.org/10.15468/wxz3yj> (22.03.2020).
38. Ivanova N., Shashkov M., Buyvolov Y. How much data do Russian biological collections contain? // Abstracts of the conference “Information Technologies in the Research of Biodiversity”. Irkutsk: ISDCT SB RAS, 2018. P. 75–76.
39. GBIF Backbone Taxonomy. Checklist dataset. GBIF Secretariat, 2019. <https://doi.org/10.15468/39omei> (22.03.2020).