

ИСПОЛЬЗОВАНИЕ ДАННЫХ GBIF.ORG РОССИЙСКИМИ ИССЛЕДОВАТЕЛЯМИ

Иванова Н. В.

*Институт математических проблем биологии РАН –
филиал ИПМ им. М. В. Келдыша РАН, Пушкино, Россия*

GBIF.ORG DATA USE BY RUSSIAN RESEARCHERS

Ivanova N. V.

*Institute of Mathematical Problems of Biology RAS – the Branch of Keldysh Institute
of Applied Mathematics of Russian Academy of Sciences, Pushchino, Russia,
ORCID: [0000-0003-4199-5924](https://orcid.org/0000-0003-4199-5924)*

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

Summary: by now 76 peer-reviewed articles from Russia that use GBIF data are visible via GBIF.org. Russian citation index provides 128 Russian GBIF-mediated articles and 78 conference abstracts. Not all authors follow the citation guidelines, and many GBIF datasets are incorrectly cited.

Keywords: GBIF, eLIBRARY.RU, citation guidelines

Объединенные данные о распространении видов становятся сегодня важным (а иногда и основным) источником информации в исследованиях биоразнообразия. Портал GBIF, Global Biodiversity Information Facility (GBIF, 2020), является крупнейшим в мире ресурсом таких данных. На январь 2019 г. через GBIF доступны более 1.3 млрд записей о находках видов, опубликованных более чем 1 500 организаций со всего мира, в т.ч. 3.8 млн записей для территории России. Все эти данные свободно доступны для повторного анализа и распространяются по одной из трех открытых лицензий сообщества Creative Commons (CC-0, CC-BY, CC-BY-NC). Каждому набору данных присваивается идентификатор цифрового объекта (DOI), с помощью которого GBIF отслеживает использование опубликованных данных в составе объединенных массивов и их цитирование в научной литературе. Для каждого набора данных отображаются все публикации, в которых процитирована хотя бы одна его запись.

Согласно статистике глобального портала, с 2000 г. в рецензируемых журналах опубликовано 7 162 статьи, связанных с GBIF, в т.ч. 4 185 публикаций содержат результаты анализа данных, полученных через глобальный портал (Peer-reviewed..., 2020). Из них российскими исследователями опубликованы 175 и 76 статей соответственно. Поиск в научной электронной библиотеке [eLIBRARY.RU](https://elibrary.ru) по ключевому слову «GBIF» показал, что с 2004-го по 2019 г. российскими исследователями опубликовано 128 статей и 78 материалов конференций по данной тематике. Среди этих публикаций только 5 статей включены в список, доступный на портале GBIF.

Эти результаты показывают, что данные глобального портала пока мало используются отечественными учеными, вероятно, из-за низкой репрезентативности для многих регионов России. В то же время объем национальных данных, доступных через GBIF, возрастает с каждым годом (Shashkov, Ivanova, 2019), открытые цифровые данные о биоразнообразии

постепенно вовлекаются в научный оборот, но отечественные публикации, подготовленные на основе их анализа, далеко не всегда оказываются «видимыми» через глобальный портал. Причина этого несоответствия заключается в том, что поисковая система GBIF при подсчете цитирований наборов данных использует Google Scholar, Scopus, Wiley Online Library, SpringerLink, NCBI Pubmed и bioRxiv, которые охватывают далеко не все российские научные издания. В результате большинство публикаций, индексируемых в eLIBRARY.RU, не учитывается. В настоящее время Секретариат GBIF работает над включением Российского индекса научного цитирования в свои поисковые алгоритмы, что позволит более полно отобразить публикации отечественных ученых на глобальном портале.

Важно, чтобы публикации результатов исследований, выполненных с использованием данных GBIF, не только были обнаружимы через глобальный портал, но и связаны с соответствующими наборами данных. Для этого необходимо, чтобы пользователи GBIF соблюдали правила цитирования использованных ими источников (Citation..., 2020).

При цитировании одного набора данных следует приводить в списке литературы информацию, указанную в разделе Citation на его веб-странице, – авторов первичных данных, название набора данных и номер версии, публикующую организацию, DOI, URL и дату обращения, как, например, в (Karyakin et al., 2018). Обычно исследователи не используют исходные наборы данных GBIF целиком, а формируют поисковые запросы по определенным критериям. Результату каждого такого запроса всегда присваивается DOI, который содержит параметры поиска, ссылки на источники происхождения и авторство всех найденных записей. В этом случае рекомендуется включать в ссылку DOI, указание на GBIF и дату выгрузки, как, например, в (GBIF, 2019). На наш взгляд, хорошей практикой также является перечисление авторов всех наборов данных, включенных в результат запроса. Для быстрого получения этих сведений удобнее выгружать данные не в CSV-таблице, а в формате Darwin Core Archive (Wieczorek et al., 2012). Список всех использованных наборов данных и их авторов доступен в архиве в файле citations.txt.

Из 76 статей российских авторов, обнаружимых через GBIF, наборы данных корректно процитированы (с указанием DOI) в 12 публикациях. Среди 128 статей, обнаруженных нами по ключевому слову «GBIF» через eLIBRARY.RU, только в 21 приведены ссылки на DOI использованных наборов данных.

Некорректная ссылка на набор данных, полученный через глобальный портал, не позволяет рецензентам и читателям оценить полноту и качество использованных данных, воспроизвести результаты анализа и согласиться с представленными выводами. Такие ссылки не учитываются в списках цитирований наборов данных, а, следовательно, не отражают реальной востребованности цифровых данных о биоразнообразии, посвященных разным систематическим группам.

Благодарности

Автор благодарит сотрудников Секретариата GBIF Дмитрия Щигеля и Даниэля Нусгаарда (Daniel Noesgaard) за консультации при подготовке тезисов.

Литература

- Citation guidelines. GBIF.org. URL: <https://www.gbif.org/citation-guidelines> (2020-01-27).
GBIF – Global Biodiversity Information Facility. Свободный и открытый доступ к данным о биоразнообразии.
URL: <https://www.gbif.org/> (2020-01-27).
GBIF Occurrence Download. GBIF.org. DOI: [10.15468/dl.z5n8gs](https://doi.org/10.15468/dl.z5n8gs) (2019-12-07).

- Karyakin I., Kamenskiy D., Grachev E., Shnayder E., Ebel A., Shtol D., Mashkov V. Birds of Northern Eurasia. Occurrence dataset. Sibecocenter LLC, 2018. DOI: [10.15468/hmqxgv](https://doi.org/10.15468/hmqxgv) (2020-01-27).
Peer-reviewed papers using data. GBIF.org. URL: https://www.gbif.org/resource/search?contentType=literature&literatureType=journal&relevance=GBIF_USED&peerReview=true (2020-01-27).
- Shashkov M., Ivanova N. Considerable Progress in Russian GBIF Community // Biodiversity Information Science and Standards. 2019. V. 3. P. e37015. DOI: [10.3897/biss.3.37015](https://doi.org/10.3897/biss.3.37015)
- Wieczorek J., Bloom D., Guralnick R., Blum S., Döring M., De Giovanni R., Robertson T., Vieglaiss D. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard // PLoS ONE. 2012. V. 7(1). P. e29715. DOI: [10.1371/journal.pone.0029715](https://doi.org/10.1371/journal.pone.0029715)