

GBIF: публикация данных по биоразнообразию за 6 шагов



Для организаций из стран, не имеющих членства в GBIF¹

Ноябрь 2015

Общая процедура для онлайн публикации данных по биоразнообразию через портал GBIF.org включает в себя одобрение новых участников странами или международными организациями, которые уже являются членами GBIF. С 2015 года существует дополнительный путь для организаций, которые планируют опубликовать данные, из стран, не являющихся Членами GBIF. Этот путь проходит в шесть простых шагов.

В настоящее время GBIF публикует следующие типы данных: **данные о распространении организмов, списки видов и метаданные**, описывающие биологические данные, а с 2015 года и **sample-based данные**.

Общее описание публикации данных в GBIF: <http://www.gbif.org/publishingdata/summary>

Учебные пособия: <http://www.gbif.org/resources/for-publishers>

1. **Согласование с руководством**

Важно обсудить Ваши планы с администрацией, например, с заместителем директора по науке, так как данные будут опубликованы от имени организации (публикация данных отдельными лицами в данный момент не поддерживается). Вы можете объяснить преимущества публикации данных как возможность повысить обнаружимость Вашей организации в сети Интернет и ее участие в глобальном процессе, в дополнение к существующим возможностям, таким как научные публикации, обмен образцами и т.д. Имеет смысл подчеркнуть, что все данные опубликованные через GBIF однозначно указывают на публикующую организацию, для каждой из которых поддерживается собственная страница, на которой размещается вся необходимая информация, контакты и логотип, также создаются страницы для каждого набора данных.

О целях публикации данных через GBIF.org: <http://www.gbif.org/publishingdata/summary#whypublish>.

2. **Заполнение регистрационной формы**

Для сбора информации о новых публикующих организациях необходимо заполнить форму <http://www.gbif.org/publishing-data/how-to-publish#/intro>. Организация, соответствующая [критериям](#) GBIF, будет зарегистрирована и одобрена в качестве новой публикующей организации. Есть вопросы? Пишите на helpdesk@gbif.org.

3. **Стандартизация набора данных**

Исследователи собирают и хранят данные множеством разных способов. Для объединения и публикации данных, их нужно “привести к общему знаменателю” – упаковать в один из широко распространенных стандартов². GBIF построен на открытых стандартах. Мы рекомендуем пользоваться стандартом [Darwin Core](#) или другими протоколами, такими как стандарт [ABCD](#),

¹Список стран, имеющих членство в GBIF <http://www.gbif.org/participation/list>

²При этом исходные данные могут храниться в неизменном виде, экспорт / форматирование нужны лишь для публикации на национальном или глобальном портале. Стоит иметь в виду, что разработка системы в Darwin Core существенно повышает совместимость данных и экономит время.

используемый BioCAsE для публикации данных о распространении организмов, таксономических списков видов, а также [EML](#) для описания наборов данных (метаданные). Существует набор инструментов и документация, облегчающие этот процесс: обратите внимание на раздел Ресурсы: <http://www.gbif.org/resources/summary>. GBIF также поддерживает устаревшие протоколы [TAPIR](#) и [DiGIR](#). Данные должны быть отформатированы в соответствии с одним из этих стандартов до публикации через GBIF. Это важный момент для проверки качества данных: убедитесь, что формат соблюден во всех полях, проверьте отсутствие дубликатов и неполных записей, а также опишите все замеченные несоответствия.

Описание некоторых полей Darwin Core: EN <http://rs.tdwg.org/dwc/terms/> и RU <http://gbif.ru/node/5>

4. **Выбор способа публикации**

Для проверки и публикации данных мы рекомендуем Integrated Publishing Toolkit (IPT) <http://www.gbif.org/ipt>: Эта программа интегрирована в систему GBIF, включая автоматическое индексирование. Вы можете установить программу IPT на открытом сервере, или получить учетную запись в существующей инсталляции IPT, связавшись с ее администратором. Публикующая организация в любом случае будет отображаться на отдельной странице. Существуют и другие способы публикации данных, в том числе размещение в сети статического набора данных в виде Darwin Core archive (DwC-A).

Руководство по использованию IPT: <http://www.gbif.org/resources/2571>

Видеоруководство по установке IPT: <http://www.gbif.org/resources/78256>

Российские IPT-инсталляции:

Институт математических проблем биологии, г. Пущино (администратор Максим Шашков Max.carabus@gmail.com)

Институт растениеводства им. Н.И. Вавилова, г. Санкт-Петербург (администратор Игорь Лоскутов i.loskutov@vir.nw.ru)

Зоологический институт, г. Санкт-Петербург (администратор Роман Халиков Roman.Khalikov@zin.ru)

5. **Проверка набора данных**

Используя IPT, проверьте Ваш набор данных, чтобы обеспечить индексацию ваших данных. Вы также должны предоставить полное описание набора данных (метаданные). IPT позволяет назначить регулярное обновление данных, опубликованных через GBIF.

Руководство по IPT: <http://www.gbif.org/resources/2571>

6. **Публикация данных**

Нажмите на кнопку **Publish**. Проверьте страницу Вашей публикующей организации (**data publisher**), страницу набора данных (**dataset information**), активность (**activity**), т.е. сведения о скачиваниях, и статистику (**stats**), отображающую основные сведения о наборе данных. Статистика предоставляет возможность обнаружить и исправить проблемы интерпретации, например, в координатах.

См. раздел Data publishing для каждой страны: <http://www.gbif.org/country> а также статистику публикации данных по странам <http://www.gbif.org/analytics/country/about>.

Вопросы?

Можно задать по электронной почте (helpdesk@gbif.org), по телефону +45 35321470 или по факсу +45 35321480.

Дополнительные возможности

Рассмотрите вариант публикации метаданных в виде **data paper** <http://www.gbif.org/publishingdata/datapapers>

Участвуйте в переводе **IPT на русский язык** <https://crowdin.com/project/gbif-ipt>

Скоро: отслеживайте использование данных в литературе и в других контекстах при помощи **DOI**