

УДК: 574.9:004.6

ИНФОРМАТИКА БИОРАЗНООБРАЗИЯ: МИРОВЫЕ ТЕНДЕНЦИИ, СОСТОЯНИЕ ДЕЛ В РОССИИ И РАЗВИТИЕ НАПРАВЛЕНИЯ В ХАНТЫ-МАНСЬИЙСКОМ АВТОНОМНОМ ОКРУГЕ

Филиппова Н.В.¹, Филиппов И.В.¹, Щигель Д.С.², Иванова Н.В.³, Шашков М.П.^{3,4}

1) Югорский государственный университет, г. Ханты-Мансийск, filippova.courlee.nina@gmail.com, fillip.83.pov@yandex.ru

2) Global Biodiversity Information Facility, Secretariat, г. Копенгаген, Дания, dschigel@gbif.org

3) Институт математических проблем биологии РАН - филиал ИПМ им. М.В. Келдыша РАН, г. Пущино, Natalya.dryomys@gmail.com

4) Институт физико-химических и биологических проблем почвоведения РАН, г. Пущино, mx.carabus@gmail.com

В публикации мы делаем краткий обзор истории и современных направлений Информатики биоразнообразия, а также стандартов, используемых для организации баз данных о биологическом разнообразии. Затем приводим примеры современных глобальных и локальных проектов по интеграции данных о биологическом разнообразии в мире и в России. В заключительной части публикации обсуждается вопрос организации региональной системы в Ханты-Мансийском автономном округе – Югре, где до сих пор биологические коллекции не оцифрованы и их данные не интегрируются в глобальные депозитарии.

Ключевые слова: информационные системы, базы данных, биологическое разнообразие, биологические коллекции, GBIF

Citation: Filippova N.V., Filippov I.V., Schigel D.S., Ivanova N.V., Shashkov M.P. 2017. Biodiversity informatics: global trends, national perspective and regional progress in Khanty-Mansi Autonomous Okrug // Environmental dynamics and global climate change. V. 8. No. 2. P. 46-56.

Цитирование: Филиппова Н.В., Филиппов И.В., Щигель Д.С., Иванова Н.В., Шашков М.П. 2017. Информатика биоразнообразия: мировые тенденции, состояние дел в России и развитие направления в Ханты-Мансийском Автономном Округе // Динамика окружающей среды и глобальные изменения климата. Т. 8. № 2. С. 46-56.

ВВЕДЕНИЕ

Биологическое разнообразие является важнейшим компонентом окружающей среды непосредственно и опосредованно влияющим на состояние глобальных экосистем и климата и отражающим их состояние [Cardinale et al., 2012; Pauls et al., 2013]. Данные о распространении биологических видов, общей картины распространения живого на Земле и ее изменении во времени необходимы для решения задач биогеографии, флористики, и других научных дисциплин, а также решения практических задач, таких как охрана природы и управление природными ресурсами. Для интеграции этих данных в прошлом использовались бумажные носители, например, карточные каталоги коллекций, на основе анализа которых создавались атласы о распространении видов и биогеографические сводки. Однако, с развитием и распространением информационных технологий сбор данных о биологическом разнообразии стал гораздо более детальным и оперативным, появились качественно новые возможности для интеграции, хранения, визуализации и анализа. Информационные технологии, применяемые для решения широкого класса подобных задач, объединяют под общим термином «Информатика биоразнообразия¹²» (ИБ) (Biodiversity informatics).

¹ Не путать с Биоинформатикой (Bioinformatics), смежной дисциплиной и набором инструментов для хранения и управления молекулярными данными (более полное определение термина см., например, <https://ru.wikipedia.org/wiki/Биоинформатика>).

² Существует несколько областей информатики, направленных на решение проблем других дисциплин. Таковы, в частности, биоинформатика (bioinformatics) и хемоинформатика, т.е. химическая информатика (cheminformatics, chemoinformatics). В указанных случаях для именованной дисциплины либо используется прилагательное, образованное от названия дисциплины, проблемы которой решаются, и термин "информатика", либо к этому термину добавляется сокращенное название дисциплины (в этом случае формируется одно слово). В случае интересующей нас дисциплины, *biodiversity informatics*, попытки образовать прилагательное от русского термина "биоразнообразие" дает в итоге интуитивно непривлекательный вариант "биоразнообразная

В задачи ИБ входит разработка стандартов хранения данных о биоразнообразии, методов их обработки и протоколов обмена, применение технологий распространения и обеспечения доступности информации через Интернет и пр. Одной из важных проблем является развитие глобальной таксономической системы, обеспечивающей согласованность между различными локальными системами. Начиная с июня 2001 г. и по настоящее время идет работа над такой системой на основе Каталога Жизни (Catalogue of Life – CoL, <http://www.catalogueoflife.org>). ИБ широко использует геоинформационные системы и системы управления базами данных для предварительной обработки (data-cleaning tools) и хранения данных.

В настоящее время существует множество независимых систем, разрабатываемых отдельными институтами или коллективами, работающими на региональном уровне, а также систем специализирующихся на отдельных таксонах (обзор таких систем для территории России см. [Ivanova, Shashkov, 2017]). Таким образом, происходит накопление больших объемов первичной информации о биоразнообразии, но одновременно с этим происходит интегрирование разрозненных данных специализированных коллекций и информационных систем в единую унифицированную базу. Ярким успешным примером интеграции разрозненных данных о биологическом разнообразии является Глобальная информационная система по биоразнообразию (Global Biodiversity Information Facility – GBIF, <http://www.gbif.org>), а в области интеграции таксономической информации – Catalogue of Life. Однако, успешность совместной работы с данными о биоразнообразии на глобальном уровне зависит от соблюдения заинтересованными сторонами специальных правил и стандартов сбора, хранения и обмена данными о биоразнообразии.

Поскольку разработка информационных систем для управления данными о биологическом разнообразии только начинается в пределах Ханты-Мансийского Автономного Округа-Югры (ХМАО), актуальным является планирование и разработка ИС с учетом истории и современных мировых и национальных тенденций. Таким образом, *целью настоящей публикации* является обзор истории развития ИБ в мире и в России, описание состояния дел в настоящее время в ХМАО и планирование будущего развития ИС по биоразнообразию на региональном уровне.

ОСНОВНЫЕ НАПРАВЛЕНИЯ И МЕЖДУНАРОДНЫЕ ИНИЦИАТИВЫ В РАМКАХ ИНФОРМАТИКИ БИОРАЗНООБРАЗИЯ

Политическое основание. Важным основанием для организации международных программ по интеграции данных о биологическом разнообразии является Конвенция о Биологическом разнообразии (Convention on Biological Diversity, CBD) (<https://www.cbd.int>), ратифицированная Россией в 1995 г. В статье 17 Конвенции постулируется необходимость объединения усилий всех стран для оценки собственных биологических ресурсов. Очевидно, что изучение биологического разнообразия на мировом уровне или уровне крупных регионов возможно только благодаря согласованным действиям многих стран [Canhos et al., 2004]. В 2010 г. Стороны Конвенции приняли Стратегический план в области сохранения и устойчивого использования биоразнообразия на 2011-2020 гг. вместе с целевыми задачами (так называемые Айтинские задачи, Aichi Biodiversity Targets – <https://www.cbd.int/sp/targets>). Целевая задача 19 из этого списка направлена на усовершенствование, совместное использование и применение знаний, научных данных и технологий в области изучения биоразнообразия. Выполнение этой задачи невозможно без оцифровки биологических коллекций и открытия данных о распространении видов и динамике их ареалов.

История развития и ведущие проекты в области ИБ. Первая в мире программа по оцифровке данных биологических коллекций была начата в середине 70-х гг. в Австралии, где также был выпущен первый документ, посвященный стандартам данных в гербарных коллекциях (Herbarium Information Standards and Protocols for Interchange of Data). На основе этого опыта в 90-х гг. началось активное развитие национальных и международных программ в области ИБ в Европе и на Американском континенте. Международная сеть по вопросам биологического разнообразия

информатика".

Альтернативой может служить либо введение термина "биодиверсивная информатика", который, однако, может вызывать нежелательные ассоциации со словами "диверсия, биодиверсия". Наконец, можно использовать в качестве уточняющего слова не прилагательное, а существительное, что дает вариант "информатика биоразнообразия". Мы придерживаемся последнего варианта.

(Biodiversity Information Network) в это время служила для объединения усилий разных стран и развития локальных проектов, а также участвовала в обсуждении стратегии развития Конвенции.

В настоящее время существует несколько крупных международных проектов, направленных на объединение данных о биологическом разнообразии из разрозненных источников. Ведущей инфраструктурой является GBIF, образовавшийся на основе межправительственных соглашений в 2001 г. на совещании Megascience Forum на конференции Организации экономического сотрудничества и развития (Organisation for Economic Co-operation and Development, OECD). Целью проекта является создание глобального портала данных о биологическом разнообразии и решение следующих задач: 1) стимулирование оцифровки данных естественно-научных коллекций, 2) обеспечение доступности данных и их совместимости, 3) создание электронного каталога названий всех биологических организмов, 4) пропаганда основ и развитие теории ИБ [Canhos et al., 2004; Lane, Edwards, 2007]. На момент подготовки статьи через портал GBIF.org доступны данные о 737 816 656 находках 1 718 908 видов. Для организации информации в настоящее время предусмотрены следующие типы наборов данных: 1) метаданные о коллекциях или находках видов (metadata-only), 2) региональные и глобальные списки видов (checklist data), 3) данные о находках видов (occurrence data), 4) данные обследований на пробных площадях и учетов (sampling-event data) (<http://www.gbif.org/publishing-data/summary>).

Примерами крупных региональных сетей для решения задач ИБ являются Европейская (The European Network for Biodiversity Information) и Американская сеть данных по биоразнообразию (Inter-American Biodiversity Information Network), направленные на решение внутренних задач для каждого из регионов и обеспечения связи с GBIF. В Европе до этого успешно работает сеть Таксономических организаций (The Consortium of European Taxonomic Facilities, <http://www.cetaf.org>), способствующая объединению баз данных таксономических систем растений, наземных и морских животных. Кроме того, создавались сети для решения задач связанных с разработкой стандартов баз данных, например The European Natural History Specimen Information Network [Canhos et al., 2004].

В 2016 г. на сайте GBIF опубликован отчет по ситуации с оцифровкой коллекций в мире (<http://www.gbif.org/resource/83022>). Совместная с GRbio (<http://grbio.org>) работа над поддержанием этой информации в актуальном состоянии продолжается.

Одна из проблем организации интегрированных баз данных – это существование многих разрозненных и часто параллельно развивающихся таксономических систем. Поэтому отдельной задачей является объединение отличающихся таксономических систем, используемых в различных таксономических базах. В настоящее время функции такой глобальной таксономической системы наиболее полно выполняет CoL. Теоретические основы выработки консенсуса между различными системами обсуждаются в ряде работ (например: [Berebdsohn, Geoffroy, 2007]).

Международные стандарты и протоколы обмена между базами данных. Одной из проблем на пути интеграции различных баз данных является их совместимость (наличие одинаковых по смыслу полей, заполненных в стандартном формате) и разработка общепринятых протоколов для обмена данными, полученными из различных источников. Ведущей по разработке стандартов в настоящее время является группа Biodiversity Informatics Standards (альтернативное название Taxonomic Databases Working Group, TDWG – <http://www.tdwg.org>). Она проводит ежегодные рабочие встречи и конференции, и активно работает между сессиями на тематических встречах (<http://www.tdwg.org/activities>). В 90-х гг. группой разрабатывался глобальный формат Access to Biological Collection Data (ABCD) для стандартизации самых разных баз данных от палеонтологических коллекций до коллекций зоопарков и ботанических садов. Позднее был принят еще один более упрощенный стандарт Darwin Core (DwC) включающий глоссарий более 200 терминов предназначенных для обмена информацией о биологическом разнообразии (в частности биологических коллекций). В настоящее время DwC наиболее часто применяется при интеграции разрозненных баз данных и обмена данными, а также используется как основной стандарт в GBIF [Wieczorek et al., 2012].

Для обмена данными продолжают использоваться протоколы DiGIR, BioCase и TAPIR (<https://github.com/tdwg/tapir>). В настоящее время для публикации данных в сети GBIF активно развивается пакет свободно распространяемого программного обеспечения Integrated Publishing Toolkit (IPT) [Robertson et al., 2014]. IPT осуществляет проверку данных на соответствие стандарту Darwin Core, их архивирование (Darwin Core Archive) и публикацию на глобальном портале GBIF.org.

Вопросы авторских прав при публикации данных через интегрированные БД. Кроме вопросов организации общих стандартов и протоколов, ряд проблем связан с вопросами авторских прав и адекватного цитирования данных при их интеграции в глобальные базы. И хотя в последнее время научное сообщество все больше настаивает на открытости исходных данных, есть и обратные случаи, связанные с коммерческими интересами. Авторские права не применимы к первичным данным как таковым [Egloff et al., 2016], несмотря на то, что они действуют в отношении публикаций, программного обеспечения, иллюстраций и географических карт. В контексте GBIF, при консультации с научным сообществом, была разработана система лицензирования данных при публикации на основе лицензий Creative Commons (<http://www.gbif.org/newsroom/news/data-licensing-milestone>). Эта система позволяет публикующей организации заявить об отсутствии ограничений на использование данных, либо требовать ссылки на источник, либо запретить коммерческое использование данных. Следует заметить, что отсутствует механизм контроля за соблюдением лицензионных ограничений – как и с цитированием статей, адекватное цитирование данных находится под контролем цитирующего. На новом портале GBIF приводятся рекомендации по цитированию данных в разных контекстах (<https://demo.gbif.org/citation-guidelines>).

О современных правилах публикации данных о биологическом разнообразии. Близко с вопросом сохранения авторских прав стоит вопрос правильной публикации данных в сети и правильного цитирования таких публикаций. Важность публикации исходных данных в научных исследованиях в последнее время осознается все больше и ряд международных программ работают над разработкой политики и стандартов в этой области: например, The Committee on Data for Science and Technology – CODATA, Research Data Alliance – RDA и др. Особенности публикации данных о биологическом разнообразии занимается группа Интеграции данных о биологическом разнообразии внутри RDA (Biodiversity Data Integration Interest Group). Исходными данными для исследований в области биоразнообразия могут быть списки видов, таблицы с данными о находках видов или описания, выполненные на пробных площадях, записи голосов, видео, фотографии и другая информация, послужившая исходным материалом для написания статьи или имеющая научную ценность сама по себе [Penev et al., 2017].

Было сформулировано 4 основных принципа (*FAIR*), которым должна соответствовать современная публикация данных. Данные должны быть:

- Доступны для поиска (*Findable*)
- Доступны для загрузки пользователем на локальный компьютер (*Accessible*)
- Совместимы (*Interoperable*)
- Доступны для повторного использования (*Re-usable*)

Важность публикации данных такого рода заключается в следующем: полученные данные могут быть использованы третьим лицом для перепроверки результатов; данные могут быть востребованы для решения другого рода задач, не поставленных в исходной работе; локальные массивы данных могут быть интегрированы в общие базы; открытые данные увеличивают возможности кооперации и сотрудничества между исследователями; создается дополнительная возможность распространения информации о результатах работы и ее цитирования другими исследователями [Penev et al., 2017].

Существует четыре основных способа публикации исходных данных:

- Электронное приложение к статье онлайн-журнала (является менее предпочтительным, поскольку в этом случае затруднен поиск и обращение к материалам, сомнительна перспектива их долговременного хранения)
- Размещение на специальных онлайн-ресурсах с получением идентификатора цифрового объекта – DOI (при этом в статье приводится ссылка на данные)
- Отдельная публикация с описанием исходных данных в формате статьи о данных (<https://demo.gbif.org/publishing/data-papers>), при этом сами данные должны быть сохранены как в предыдущем пункте.
- Публикация исходных данных внутри статьи в специализированных журналах, где предусмотрен такой формат (например, Biodiversity Data Journal).

Для размещения массивов исходных данных (второй и третий способ публикации) существуют как специальные ресурсы для хранения и публикации данных о биологическом разнообразии, так и ресурсы для хранения и публикации любых первичных данных научных исследований. Первые являются предпочтительными для биологических данных, так как кроме функции хранения выполняют еще роль интеграции соответствующих данных в общий ресурс с вытекающими отсюда

новыми свойствами. Например, четыре наиболее распространенных типа исходных данных должны быть опубликованы на следующих ресурсах:

- Списки видов (Checklist Data) – GBIF
- Наборы данных об отдельных находках (Occurrence Data) – GBIF
- Данные обследования на пробных площадях и учетов (Sampling event Data) – GBIF
- Молекулярные последовательности – GenBank, European Nucleotide Archive, Barcode of Life

Data Systems и др.

Формат статьи о данных поддерживается рядом современных журналов и разработан специально для публикации описания исходных данных, а не описанием результатов, полученных на их основе, как в обычных научных статьях. Статья о данных включает описание методов сбора данных, географическую привязку, и другие метаданные. Такая статья должна обязательно иметь ссылку на опубликованный в специальном ресурсе массив данных, например в GBIF, и присвоенный ему идентификатор цифрового объекта – DOI [Penev et al., 2017]. В России первым примером такой публикации является статья о базе данных находок Борщевика Сосновского *Heracleum sosnowskyi* [Chadin et al., 2017], опубликованная в разделе Data Papers журнала Phytokeys.

С принципами цитирования массивов исходных данных, опубликованных тем или иным способом, можно ознакомиться в статье L. Penev'a с соавторами [Penev et al., 2017] и на странице <https://demo.gbif.org/publishing/data-papers>.

Программное обеспечение для менеджмента биологических коллекций. На протяжении более двух десятков лет развития ИБ было создано немало продуктов для менеджмента баз данных биологических коллекций, которые постоянно эволюционировали вслед за развитием ИБ и информационных технологий. Современного сравнительного обзора не существует (некоторая информация доступна, например, по ссылке: <https://www.idigbio.org/content/biological-collections-databases>), однако в 2003 г. специалисты GBIF предпринимали такое исследование [Berendsohn et al., 2003]. В этой работе авторы провели анализ около 20 программных продуктов, опираясь на следующие основные параметры сравнения: условия распространения (стоимость ПО, открытый или закрытый исходный код), ограничение в объемах хранения данных и возможность масштабирования, поддерживаемые операционные системы, специализация для определенной таксономической группы и пр. Всего было проанализировано около 20 систем, представленных в то время на международном уровне. Кроме того, авторы обращают внимание на ряд важных параметров, не вошедших в анализ, но которые необходимо учитывать:

- Есть ли в системе раздел, связанный с указанием номенклатурных типов (их авторство, верификация и пр.)
- Есть ли в системе возможность работы с фотографиями (возможность привязки фотографий к образцам, локалитетам, таксономическим названиям; возможность описания метаданных фотографии)
- Есть ли в системе возможность гибкой настройки прав (доступ к разным типам операций, таблицам, ограничение доступа к определенной категории информации и пр.)
- Менеджмент коллекции: есть ли возможность отслеживания актов займа, обмена, и других кураторских операций
- Сохраняется ли в системе история предыдущих определений, цитируются ли дубликаты в других институтах с их определениями
- Множественные отношения: предусмотрены ли отношения типа несколько образцов – одна единица хранения, несколько единиц – один образец; отношения хозяин-паразит; образец ткани из определенного образца и пр.
- Есть ли перевод системы на язык пользователя.

РАЗВИТИЕ ИНФОРМАТИКИ БИОРАЗНООБРАЗИЯ И ИНФОРМАЦИОННЫХ СИСТЕМ О БИОРАЗНООБРАЗИИ В РОССИИ

Объединение биологических коллекций в сети происходит в зависимости от территории, типа коллекции или конкретных задач проекта. Так или иначе, данные из локальных баз объединяются во все более и более глобальные. Однако существование региональных и специальных информационных систем, совершенно необходимо, поскольку ни одна современная глобальная система не в состоянии учесть все специфические нужды конкретных исследовательских и прикладных проектов или региональные особенности.

В России только начинается формирование сетевой структуры, объединяющей разрозненные базы данных [Иванова, Шашков, 2017]. Коллекции криптогамных организмов Полярно-альпийского ботанического сада-института Кольского НЦ РАН (КРАВГ) и пяти российских и зарубежных коллекций объединены в ИС CRIS [Мелехин и др., 2013; Давыдов и др., 2017; Мелехин, 2017]. В то же время, коллекции мохообразных аккумулируются и другой системой «Флора Мхов России» (<http://arctoa.ru/Flora>). Примером довольно масштабного проекта оцифровки биологических коллекций является ИС Московского Государственного Университета «Ноев Ковчег», где в одной системе объединились коллекции МГУ (<http://depository.msu.ru>), в т.ч. гербарий MW (<https://plant.depo.msu.ru>). В Иркутском научном центре СО РАН ведется разработка информационно-аналитической системы (ИАС) по Фиторазнообразию Байкальской Сибири [Верхозина, Федоров, 2017]. База данных гербария Института биологии Коми НЦ (SYKO) – крупнейшей гербарной коллекции на Европейском Северо-Востоке России, ведется в локальной автоматизированной информационной системе Adonis [Тетерюк и др., 2017].

Примером информационной системы, разработанной под задачи изучения ареала и экологии одного вида с использованием современных стандартов (Darwin Core) и экспортом данных в GBIF является ИС о находках охраняемого лишайника лобарии легочной *Lobaria pulmonaria* в России (<http://lobaria.ru>) [Иванова, Шашков, 2017]. Другим примером специализированной ИС для моделирования распространения инвазивного вида (также с экспортом данных в GBIF) является ИС Распространения Инвазивных Видов Растений – РИВР (<http://ib.komisc.ru/add/rivr>) [Чадин и др., 2017].

Рабочий список и описание Информационных систем, развивающихся в настоящее время в области биологического разнообразия в России, составляется неформальной рабочей группой GBIF в России. В списке в настоящее время содержатся записи о 30 информационных системах (<http://gbif.ru/occurrence>) и около 100 публикациях с описаниями систем (<http://gbif.ru/ruspublic>). К сожалению, большая часть Российских информационных систем в настоящее время имеют самостоятельные, плохо конвертируемые форматы, в них плохо реализована возможность доступа к первичным данным, а таксономические списки разрознены в пределах одной группы, правила использования и цитирования данных, как правило, не разработаны [Иванова, Шашков, 2014]. Кроме того, очень часто базы данных ведутся без продуманной структуры и возможности их интеграции, в таких продуктах как Microsoft Access и Excel. С другой стороны, есть примеры объединения данных на региональном и федеральном уровне, которые в отсутствие развитых информационных систем используют подходящие онлайн-ресурсы (как Google Spreadsheets: [Большаков и др., 2017]). Подробнее об истории и современном состоянии баз данных и информационных систем о биоразнообразии в России см. в публикации [Иванова, Шашков, 2014; Ivanova, Shashkov, 2017].

Широкие перспективы для развития национальной системы представляют программные решения Atlas of Living Australia (ALA) (<https://demo.gbif.org/programme/living-atlases>), бесплатно предоставляемые группой разработчиков из Австралии. При поддержке GBIF, ALA Tools используются такими странами как Испания, Португалия, Шотландия, Франция, Канада, Аргентина и Коста Рика для развития собственных систем на их основе. Использование и адаптация готовых решений ALA позволит России наверстать отставание в этой области.

Если говорить о конференциях как о показателе научной активности в определенной области, то с начала 90-х гг. ряд совещаний в России был посвящен проблемам ИБ. Первые два совещания по биологическим базам данных прошли в 1990 г. в Пущино-на-Оке и в Санкт-Петербурге (ЗИН РАН); тезисы докладов этих совещаний не издавались [Смирнов и др., 2006]. С 1993 г. при Ботаническом институте проходили регулярные совещания по Компьютерным базам данных в ботанических исследованиях (совещания повторялись в 1993, 1995 и 1997 гг.; тезисы последнего совещания вышли под ред. Гельтмана, Роскова [Гельтман, Росков, 1997].

С 1994 г. Зоологическим институтом проведен ряд Международных симпозиумов посвященных ИБ [Смирнов и др., 2006]:

- 1994 г. «Базы данных и компьютерная графика в зоологических исследованиях» [Степаньянц и др., 1997]
- 1999 г. «Информационно-поисковые системы в зоологии и ботанике» [Рысс, Смирнов, 1999]
- 2001 г. «Информационные и телекоммуникационные ресурсы в зоологии и ботанике» [Рысс, Минтер, 2001]
- 2003 г. «Информационные системы по биоразнообразию видов и экосистем» [Алимов, 2003]
- 2005 г. «Электронные биологические коллекции: теории, стандарты и перспективы использования» («Taxonomic Databases Working Group 2005 Annual Meeting») [Berendsohn, Rissone,

2005]

- 2006 г. «Информационные системы и WEB-порталы по разнообразию видов и экосистем» (в пос. Борок) [Павлов и др., 2006]

Также работы по ИБ представляются на регулярной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (Digital Libraries: Advanced Methods and Technologies, Digital Collections – RCDL, <http://rcdl.ru>), которая проводится с 1999 г. Вопросы развития ботанических и зоологических баз данных на территории России и, в частности, Алтае-Саянского экорегиона, также обсуждались на конференции «Формирование баз данных – опыт, проблемы, решения» (г. Барнаул) [Шмаков и др., 2009].

С 2014 г. проводятся семинары, посвященные публикации данных в GBIF (см. <http://gbif.ru>).

В 2017 г. в городе Апатиты состоялась Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях» [Боровичев и др., 2017]. На конференции обсуждались вопросы использования международных стандартов данных по биоразнообразию и создания национальной системы о биоразнообразии России. Предполагается проводить эти конференции на регулярной основе.

Публикация в GBIF является одним из показателей активной работы страны в области ИБ на уровне современных стандартов. В последнее время начинается все более активная публикация Российских наборов данных (первый набор был опубликован Зоологическим институтом РАН в 2004 г., doi:10.15468/c9g3nw). К настоящему времени Российскими организациями опубликовано 20 наборов данных (включая крупные коллекции, коллекции заповедников и базы данных по изучению ареалов отдельных видов). Однако если сравнить это с общемировой картиной, к территории России относятся 0,3% всех опубликованных в GBIF данных и подавляющее большинство их (94%) опубликовано зарубежными организациями [Шашков, Иванова, 2017].

К СОЗДАНИЮ ИНФОРМАЦИОННОЙ СИСТЕМЫ О БИОЛОГИЧЕСКОМ РАЗНООБРАЗИИ В ЮГРЕ

Ранее в Югре была организована только одна ИС по биологическому разнообразию под названием UgraBio [Филиппов, 2012]. Эта система разрабатывалась в рамках работ по ведению Красной книги Ханты-Мансийского автономного округа и предназначалась для унифицированного сбора и первичного анализа информации о биоразнообразии округа. В качестве источников информации выступали отчеты сотрудников особо охраняемых природных территорий ХМАО, отчеты по итогам научно-исследовательских работ по ведению Красной книги ХМАО, научные публикации и собственные данные пользователей ИС. Идея создания БД по биоразнообразию ХМАО возникла в 2009 г. и к началу 2010 г., в качестве эксперимента, были оцифрованы и занесены в таблицу MS Excel часть карточек встреч краснокнижных видов. По мере оцифровки карточек встреч постепенно выработывалась структура таблиц и возникла необходимость перехода от громоздких несвязанных таблиц к реляционной базе данных. В начале 2010 г. все информационные таблицы были импортированы в СУБД MS Access. На этом этапе работы были поддержаны Департаментом экологии и экологической безопасности ХМАО. Началось планомерное наполнение БД, заключающееся в извлечении необходимой информации о находках видов из текстов отчетов экспедиций по изучению биоразнообразия округа, предоставляющихся в Департамент начиная с 2000 г. в рамках проведения НИР по ведению Красной книги. В дальнейшем, пополнение базы происходило за счет информации из отчетов НИР окружных ООПТ и опубликованных источников (статей из региональных сборников научных трудов и научных журналов). В 2011 г. была разработана web-версия системы UgraBio (ugrabiо.ru) с дифференцированным на роли многопользовательским доступом. К началу 2017 г. UgraBio содержит порядка 3000 точек встреч видов, собранных 180 пользователями (в основном это точки встреч видов, включенных в красную книгу ХМАО). Кроме того, система UgraBio содержит списки видов приуроченные к ключевым участкам (предварительный аннотированный список флоры и фауны ХМАО, составленный по данным UgraBio в 2011 г. насчитывал 3500 видов).

С 2016 г. начала развиваться информационная система Биологической коллекции ЮГУ на основе открытого ПО – Specify 6 (офлайн) и Specify 7 (онлайн) ([Filippova et al., 2017], <https://fungariumysu.org/ysu-bc>). Эта система обеспечивает хранение данных биологических коллекций в формате Darwin Core, их демонстрацию онлайн, функции менеджмента коллекции и потенциальный экспорт данных в другие системы, в том числе GBIF. В настоящее время в систему

загружены данные 4,5 тыс. образцов коллекции Фунгария ЮГУ. Система Specify имеет гибкую настройку для задач определенной коллекции, так для Фунгария ЮГУ была разработана схема необходимых таблиц и полей, созданы формы ввода, разработан формат этикеток и пр. Кроме информации об образцах, система позволяет хранить связанные с ними иллюстративные материалы (рисунки, фотографии), а также имеет ГИС модули для отображения на карте точек находок. Онлайн-версия базы данных установлена на сервере ЮГУ (<http://bio.ugrasu.ru>) и доступна для гостевого входа через логин и пароль *ugrabiodb*.

Проблемы развития информационных систем по биоразнообразию ХМАО связаны в основном с отсутствием в регионе специалистов в области ИБ.

Перспективы развития интегрированной региональной ИС по биоразнообразию на основе баз данных коллекций и других массивов данных. На наш взгляд, проект по созданию информационной системы для интеграции данных в рамках определенного региона, складывается из следующих задач:

- подготовка кадров и обучение персонала в области современных разработок ИБ
- разработка или установка и адаптация готовой информационной системы для интеграции данных и ее связь с глобальными проектами (GBIF)
- организация оцифровки баз данных локальных коллекций и импорт данных в ИС
- поддержка фундаментальных и практических исследований с использованием агрегированных данных (data intensive biodiversity research)
- разработка и продвижение образовательных программ для студентов и сотрудников биологических специальностей в ВУЗах.

В настоящее время в России практически не развита система подготовки кадров в области ИБ, однако GBIF делает шаги по внедрению образования в этой области в академические системы. В настоящее время подготовка в области может проходить через посещение специализированных семинаров и школ (например, TDWG annual conference). Регулярные семинары и мастер-классы проводятся группой GBIF в России (<http://gbif.ru/lastconf>).

Выбор между готовой системой или ее разработкой зависит от наличия специалистов в области ИБ и информационных технологий. В России есть примеры, где над разработкой ИС совместно работают два института биологического и информационного направления [Верхозина, Федоров, 2017]. Однако существует также ряд готовых свободных продуктов, разрабатываемых ведущими группами в области ИБ, которые широко используются по всему миру. В случае двух вышеупомянутых ИС в ХМАО были использованы оба подхода. Первая система для учета “краснокнижных” видов была создана самостоятельно, а в основу второй был взят свободный продукт Specify. На наш взгляд, использование готового продукта является более подходящим в случае, если нет достаточно большой группы специалистов готовых заниматься разработкой. Тем не менее, даже использование готовой ИС требует некоторой квалификации и наличия IT кадров для установки и поддержания системы в рабочем состоянии.

Перед установкой готового продукта требуется провести выбор наиболее подходящего из нескольких десятков доступных (<https://www.idigbio.org/content/biological-collections-databases>; см. также раздел *Обзор программного обеспечения*). В настоящее время мы находимся в состоянии выбора между двумя продуктами: Specify (<http://specifyx.specifysoftware.org>) и Symbiota (<http://symbiota.org>). Первый продукт уже использовался на протяжении двух лет для менеджмента и размещения онлайн Биологической коллекции ЮГУ, второй в целом имеет близкий функционал, но больше возможностей для интегрирования разных систем на одном портале. Экспорт данных из системы в GBIF может быть осуществлен через один из уже существующих узлов IPT или установки нового узла.

На территории ХМАО-Югры существует несколько десятков биологических коллекций небольшого масштаба. Поэтому работу по интеграции этих коллекций провести относительно не сложно, но опыт таких проектов здесь меньше, чем в центральных регионах России. Однако, мы уже имеем хорошее начало на этом пути в форме регулярных семинаров в Музее Природы и Человека «Биологические коллекции Югры: сбор, фиксация, хранение, научный оборот». Семинар проводился дважды (2015, 2017 гг.) и собирал кураторов коллекций и отдельных исследователей, работающих на территории Югры. В докладах были представлены материалы об истории, содержании и развитии около 20 коллекций округа. По итогам первого семинара опубликованы материалы докладов и каталоги коллекций [Белогая, Скучас, 2016]. Это мероприятие указывает на потребность исследователей в сотрудничестве и интеграции данных в общие базы, а также является основой для дальнейшего развития направления в округе.

Заключение. Открытые базы данных о биологическом разнообразии, кроме фундаментальных задач изучения биоразнообразия, могут быть востребованы, например, в области рационального использования природных ресурсов, мониторинга и охраны редких видов, прогнозирования изменения биоты под влиянием антропогенных и климатических факторов и развития биотехнологий. В Ханты-Мансийском округе-Югре это направление только начинает развиваться, и мы имеем первые примеры информационных систем более или менее отвечающих современным стандартам. Развитие интегрированной информационной системы для территории ХМАО будет способствовать улучшению понимания картины распространения видов на территории интенсивного недропользования. С другой стороны, использование современных методов работы с данными о биоразнообразии значительно улучшит качество и откроет новые перспективы для исследований.

БЛАГОДАРНОСТИ

Авторы выражают благодарность Алексею Резвову (внештатный переводчик и редактор, г. Москва) за помощь в обосновании перевода термина *Biodiversity informatics* на русский язык.

ЛИТЕРАТУРА

- Berendsohn W., Güntsch A., Röpert D. Survey of existing publicly distributed collection management and data capture software solutions used by the world's natural history collections. GBIF. 2003. 44 p.
- Berendsohn W.F., Geoffroy M. Networking taxonomic concepts - uniting without 'Unitary-ism' // Biodiversity databases : techniques, politics, and applications. Boca Raton, London, New York: CRC Press, 2007. P. 13-22.
- Berendsohn W.G., Rissone A. (Eds) Taxonomic databases working group, 2005 annual meeting 11-18 september 2005, St.Petersburg, Russia: abstracts. SPb, 2005. 42 p.
- Canhos, V.P., Souza, S., Giovanni, R., Canhos, D.A.L. Global biodiversity informatics: setting the scene for a "new world" of ecological modeling // Biodiversity Informatics. 2004. V. 1. P. 1-13.
- Cardinale B.J., Duffy J.E., Gonzalez A., Hooper D.U., Perrings C., Venail P., Narwani A., Mace G.M., Tilman D., Wardle D.A., Kinzig A.P. 2012. Biodiversity loss and its impact on humanity // Nature. V. 486. No. 7401. P. 59-67.
- Chadin I., Dalke I., Zakhoshy I., Malyshev R., Madi E., Kuzivanova O., Kirillov D., Elsakov V. Distribution of the invasive plant species *Heracleum sosnowskyi* Manden. in the Komi Republic (Russia) // PhytoKeys. 2017. V. 77. P. 71-80. doi:10.3897/phytokeys.77.11186
- Egloff W., Agosti D., Kishor P., Patterson D., Miller J.A. Copyright and the Use of Images as Biodiversity Data // Research Ideas and Outcomes. 2016. V. 3. P. e12502. doi: <https://doi.org/10.1101/087015>
- Filippova N.V., Bulyonkova T.M., Karpov D.V., Lapshina E.D. 2017. Fungarium of Yugra state university and its database // Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях». Апатиты, Мурманская область 28-31 марта 2017 г.: Тезисы докладов. С. 137-142.
- Ivanova N.V., Shashkov M.P. 2017. Biodiversity Databases in Russia: Towards a National Portal // Arctic Science. doi: 10.1139/AS-2016-0050
- Lane M.A., Edwards J.L. 2007. The Global Biodiversity Information Facility (GBIF) // Biodiversity databases : techniques, politics, and applications. Boca Raton, London, New York: CRC Press. P. 1-4.
- Los W., Hof C.H.J. 2007. The European Network for Biodiversity Information // Biodiversity databases: techniques, politics, and applications. Boca Raton, London, New York: CRC Press. P. 5-12.
- Muller M.R. 2004. An Analysis of the Implications of Intellectual Property Rights (IPR) on the Global Biodiversity Information Facility (GBIF). GBIF. 43 p.
- Pauls S.U., Nowak C., Bálint M., Pfenninger M. 2013. The impact of global climate change on genetic diversity within populations and species // Molecular Ecology. V. 22. No 4. P. 925-946.
- Penev L., Mietchen D., Chavan V., Hagedorn G., Smith V., Shotton D., Tuama É.Ó., Senderov V., Georgiev T., Stoev P., Groom Q., Remsen D., Edmunds S. 2017. Strategies and guidelines for scholarly publishing of biodiversity data // Research Ideas and Outcomes. V. 3. P. e12431. doi:10.3897/rio.3.e12431
- Robertson T., Döring M., Guralnick R., Bloom D., Wieczorek J., Braak K., Otegui J., Russell L., Desmet, P. 2014. The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet // PLoS ONE. V. 9. № 8. P. e102623.
- Wieczorek J., Bloom D., Guralnick R., Blum S., Döring M., Giovanni R., Robertson T., Vieglais D. 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard // PLoS ONE. V. 7 (1). P. e29715. doi:10.1371/journal.pone.0029715
- Алимов А.Ф. (под ред.) 2003. Информационные системы по биоразнообразию видов и экосистем: международный симпозиум, 1-4 Декабря 2003, Зоологический институт РАН, Санкт-Петербург, Россия: научная программа и тезисы докладов. СПб. 116 с.
- Белогой О.И., Скучас Ю.В. (под ред.) 2016. Биологические коллекции Югры: сбор, фиксация, хранение, введение в научный оборот. Материалы научно-методического семинара в Музее Природы и Человека. 150 с.
- Большаков С.Ю., Филиппова Н.В., Потапов К.О., Агеев Д.В., Волобуев С.В. 2017. Google Spreadsheets как базовый инструмент для управления данным о биоразнообразии // Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях». Апатиты, Мурманская область 28-31 марта 2017 г.: Тезисы докладов. Апатиты. С. 21-23.

Боровичев Е.А., Давыдов Д.А., Королева Н.Е. (под ред.) 2017. Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях». Апатиты, Мурманская область 28-31 марта 2017 г.: Тезисы докладов. Апатиты. 148 с.

Верхожина А.В., Фёдоров Р.К. 2017. Развитие информационно-аналитической системы по фиторазнообразию Байкальской Сибири // Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях». Апатиты, Мурманская область 28-31 марта 2017 г.: Тезисы докладов. Апатиты. С. 26-28.

Гельтман Д.В., Росков Ю.Р. (под ред.) 1997. Компьютерные базы данных в ботанических исследованиях : сб. науч. тр. С-Петербург: Ботанический институт РАН. 113 с.

Давыдов Д.А., Мелехин А.В., Константинова Н.А., Боровичев Е.А. 2017. Возможности информационной системы Стуртогамический Русский Информационный Систем // Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях». Апатиты, Мурманская область 28-31 марта 2017 г.: Тезисы докладов. Апатиты. С. 32-33.

Иванова Н.В., Шашков М.П. 2017. Информационная система о распространении охраняемого лишайника *Lobaria pulmonaria* на основе стандартов GBIF // Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях». Апатиты, Мурманская область 28-31 марта 2017 г.: Тезисы докладов. Апатиты. С. 48-49.

Иванова Н.В., Шашков М.П. 2014. Перспективы создания открытого всероссийского информационного ресурса по биоразнообразию на основе международного стандарта GBIF // Математическая биология и биоинформатика. Т. 9. С. 396-405.

Иванова Н.В., Шашков М.П., Щигель Д.С. 2016. Глобальная информационная система по биоразнообразию GBIF: перспективы развития в России // Современные концепции экологии биосистем и их роль в решении проблем сохранения природы и природопользования: материалы всерос. (с междунар. участием) науч. шк.-конф., посвящ. 115-летию со дня рождения А.А. Уранова (Г. Пенза, 10-14 Мая 2016 Г.). Пенза: Изд-во ПГУ. С. 368-369.

Иванова, Н.В., Шашков, М.П. 2015. Использование обменного формата Darwin Core для создания web-ориентированной геоинформационной системы по распространению редких видов, совместимых с международной базой GBIF // Материалы четвертой национальной научной конференции с международным участием «Математическое моделирование в экологии» 18-22 Мая 2015, Пушкино. С. 82-83.

Мелехин А.В. 2017. CRIS как рабочая модель межрегиональной информационной системы для криптогамных ботаников // Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях». Апатиты, Мурманская область 28-31 марта 2017 г.: Тезисы докладов. Апатиты. С. 88-90.

Мелехин А.В., Давыдов Д.А., Шалыгин С.С., Боровичев Е.А. 2013. Общедоступная информационная система по Биоразнообразию цианопрокариот и лишайников CRIS (Cryptogamic Russian Information System) // Бюл. Моск. о-ва испытателей природы. Отд. Биол. Т. 118. Вып. 6. С. 51-56.

Павлов Д.С., Дребудзе Ю.Ю., Петросян В.Г. (под ред.) 2006. Информационные системы и WEB-Порталы по разнообразию видов и экосистем. Материалы международного симпозиума. М.: Т-во научных изданий КМК. 261 с.

Рысс А.Ю., Минтер Д. (под ред.) 2001. Информационные и телекоммуникационные ресурсы в зоологии и ботанике. Тезисы 2-го международного симпозиума. СПб.

Рысс А.Ю., Смирнов И.С. (под ред.) 1999. Информационно-поисковые системы в зоологии и ботанике. Труды Зоологического института РАН. Т. 287.

Смирнов И.С., Лобанов А.Л., Пугачев О.Н. Гельтман Д.В. (под ред.) 2006. К истории совещаний по биологическим (зоологическим и ботаническим) базам данных и информационно-поисковым системам в СССР и России // Информационные системы и WEB-Порталы по разнообразию видов и экосистем. Материалы международного симпозиума / Павлов Д.С. др. (под ред.). М.: Т-во научных изданий КМК. С. 18-19.

Степаньянц С.Д., Лобанов А.Л., Дианов М.Б. (под ред.) 1997. Базы данных и компьютерная графика в зоологических исследованиях. Труды Зоологического института РАН. Т. 269.

Тетерюк Л.В., Чадин И.Ф., Федорова Е.В., Мади Е.Г., Кирпичев А.Н., Оплеснина Н.А. 2017. Электронный каталог и БД Гербария института биологии Коми НЦ УРО РАН (SYKO) // Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях». Апатиты, Мурманская область 28-31 марта 2017 г.: Тезисы докладов. Апатиты. С. 119-120.

Филиппов И.В. 2012. Итоговый отчет по договору № 038-2012: «Информационная система по объектам животного и растительного мира Ханты-Мансийского автономного округа». Москва. 31 с.

Чадин И.Ф., Далькэ И.В., Захожий И.Г., Малышева Р.В., Мади Е.Г., Кузванова О.А., Кириллов Д.В., Елсаков В.В. 2017. Информационная система «Распространение инвазивных видов растений»: опыт сбора данных о распространении борщевика сосновского (*Hercleum sosnowskyi*) и их публикации в GBIF // Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях». Апатиты, Мурманская область 28-31 марта 2017 г.: Тезисы докладов. Апатиты. С. 127-128.

Шашков М.П., Иванова Н.В. 2017. Российские данные в глобальной информационной системе по биоразнообразию - GBIF // Международная научно-практическая конференция «Использование современных информационных технологий в ботанических исследованиях». Апатиты, Мурманская область 28-31 марта 2017 г.: Тезисы докладов. Апатиты. С. 132-133.

Шмаков А.И., Смирнов С.В., Яковлев Р.В., Ваганов А.В. 2009. Формирование баз данных по биоразнообразию - опыт, проблемы, решения: материалы Международной научно-практической конференции (Барнаул, 13-15 мая 2009 г.). Барнаул: Изд-во «АРИКА». 284 с.

BIODIVERSITY INFORMATICS: GLOBAL TRENDS, NATIONAL PERSPECTIVE
AND REGIONAL PROGRESS IN KHANTY-MANSI AUTONOMOUS OKGUG

Filippova N.V.¹, Filippov I.V.¹, Schigel D.S.², Ivanova N.V.³, Shashkov M.P.^{3,4}

1) Yugra State University, Khanty-Mansiysk

2) Global Biodiversity Information Facility, Secretariat, Copenhagen, Denmark

3) Institute of Mathematical Problems of Biology, Pushchino, Moscow Region

4) Institute of Physicochemical and Biological Problems in Soil Science, Pushchino, Moscow Region

This paper briefly summarizes the history and the contemporary trends in biodiversity informatics, as well as the standards used for biodiversity databases construction. We provide the examples of modern global and local projects on data integration in biodiversity in Russia and worldwide. Finally we discuss the developing of the regional system in the Khanty-Mansi autonomous region - Yugra, where biological collections hasn't been digitized and their data hasn't been shared through global repositories such as GBIF until recently.

Keywords: information systems, data bases, biodiversity, biological collections, GBIF

Citation: Filippova N.V., Filippov I.V., Schigel D.S., Ivanova N.V., Shashkov M.P. 2017. Biodiversity informatics: global trends, national perspective and regional progress in Khanty-Mansi Autonomous Okrug // Environmental dynamics and global climate change. V. 8. No. 2. P. 46-56.

Поступила в редакцию: 15.09.2017
Переработанный вариант: 17.11.2017