

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/302875288>

Коллекции Зоологического института РАН как информационная основа фундаментальных биологических исследований

Conference Paper · October 2015

CITATIONS

0

READS

26

5 authors, including:



Igor Sergeevich Smirnov-Allik
Russian Academy of Sciences

209 PUBLICATIONS 243 CITATIONS

[SEE PROFILE](#)



S. Yu. Sinev
Russian Academy of Sciences

41 PUBLICATIONS 94 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Коллекции Зоологического института РАН как информационная основа фундаментальных биологических исследований [View project](#)

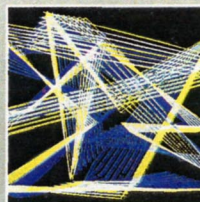
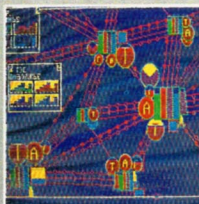
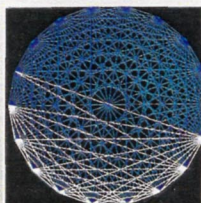


Arctic marine zoobenthos biodiversity [View project](#)

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР им. А.А. ДОРОДНИЦЫНА
РОССИЙСКОЙ АКАДЕМИИ НАУК**

**ИНФРАСТРУКТУРА
НАУЧНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ
И СИСТЕМ**

Сборник избранных научных статей



МОСКВА

81
29.4.16
ISBN 978-5-91601-127-2

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
УЧРЕЖДЕНИЕ НАУКИ
ИНСТИТУТ КОСМИЧЕСКИХ ИССЛЕДОВАНИЙ
РОССИЙСКОЙ АКАДЕМИИ НАУК

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
УЧРЕЖДЕНИЕ НАУКИ
ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР им. А.А. ДОРОДНИЦЫНА
РОССИЙСКОЙ АКАДЕМИИ НАУК

**ИНФРАСТРУКТУРА
НАУЧНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ
И СИСТЕМ**

Сборник избранных научных статей



ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР им. А.А. ДОРОДНИЦЫНА
РОССИЙСКОЙ АКАДЕМИИ НАУК
МОСКВА 2015

УДК [002:004.9] (063)
ББК [73+32.973.233]я43

Ответственные редакторы

доктор техн. наук Е.Б. КУДАШЕВ,
доктор физ.-матем. наук В.А. СЕРЕБРЯКОВ

Инфраструктура научных информационных ресурсов и систем. Сборник избранных научных статей. Труды Пятого Всероссийского симпозиума (С.-Петербург, 6–8 октября 2015 г.). Под ред. Е.В. Кудашева, В.А. Серебрякова. М.: ВЦ РАН, 2015.

Симпозиум проводится ежегодно с 2011 г. по Плану научных конференций Отделения математических наук РАН. В 2011 г. и 2012 г. Симпозиум проводился в С.-Петербурге при поддержке РФФИ, в 2013 г. – в Абхазии, г. Сухум – при поддержке Академии наук Абхазии. В 2014 г. Симпозиум проводился в С.-Петербурге при поддержке РФФИ на базе Петербургского Отделения Математического института им. В.А. Стеклова РАН – ПОМИ РАН.

Научная программа Симпозиума «**Инфраструктура научных информационных ресурсов и систем**» ориентирована на рассмотрение проблем и перспектив развития информационно-телекоммуникационных систем; методов, технологий и средств применительно к доступу, хранению и интеллектуальному анализу данных в различных областях фундаментальной науки, разработки информационных систем для научных исследований.

Основные цели Пятого Симпозиума: методы и технологии интеграции электронных коллекций; взаимодействия информационных ресурсов и формирования электронного документного пространства научных исследований и инноваций, развитие электронных библиотек.

The symposium is held annually since 2011 on the Plan of scientific conferences Department of Mathematical Sciences of RAS. In 2011 and 2012 Symposium was held in St. Petersburg and was supported by RFBR, in 2013 – in Abkhazia, Sukhum – with the support of the Academy of Sciences of Abkhazia. In 2014, the Symposium is held in St. Petersburg on the basis of the St. Petersburg Branch of the Mathematical Institute of the Academy of Sciences – PDMI RAS and is supported by RFBR.

The scientific program of the Symposium is oriented to the infrastructure of scientific information resources and systems geared to the problems and prospects of development of information and telecom-

munication systems; methods, tools and technology with respect to access, storage, and data mining in various fields of basic science, development of information systems for research.

The main objectives of the Fifth Symposium: methods and techniques of integration of digital collections; interaction of information resources and the generation of the electronic document space research and innovation, the development of digital libraries.

Рецензенты: Сметанин Ю.Г., Хорошевский В.Ф.

Научное издание

© Федеральное государственное бюджетное учреждение науки
Вычислительный центр им. А.А. Дородницына Российской
академии наук.

**Программный комитет
Пятого Всероссийского Симпозиума:**

Гордов Е.П., профессор (Институт мониторинга климатических и экологических систем СО РАН, Томск);

Кудашев Е.Б., акад. РИА (Институт космических исследований РАН / Мех.-матем. факультет МГУ им. М.В. Ломоносова);

Жижимов О.Л., профессор (Институт Вычислительных технологий СО РАН, Новосибирск);

Федотов А.М., член-корреспондент РАН (Институт вычислительных технологий СО РАН / Новосибирский государственный университет);

Серебряков В.А., профессор (ВЦ РАН, Москва);

Малков О.Ю., профессор (Институт астрономии РАН, Москва);

Вязилов Е.Д., профессор (ВНИИ гидрометеорологической информации – Мировой Центр данных), Обнинск.

Председатели Программного комитета Симпозиума:
академик Российской инженерной академии **Е.Б. Кудашев**
и профессор **В.А. Серебряков**.

Оргкомитет Симпозиума

Председатель Оргкомитета

Серебряков В.А., профессор (ВЦ РАН, Москва);

Е.Б. Кудашев (ИКИ РАН), академик РИА;

Теймуразов К.Б. (ВЦ РАН, Москва);

Молородов Ю.И., (Томский государственный университет, Томск);

Малков О.Ю., проф. (ИНАСАН, Москва);

Фазлиев А.З. (Институт оптики атмосферы СО РАН, Томск);

ОГЛАВЛЕНИЕ

Предисловие.....9

Д.А.Ковалева, О.Ю.Малков, А.С.Позаненко

Большие данные в астрономии.....12

Барабанов С.И., Верецагин С.В., Чупина Н.В.,
Фионов А. С.

Структура данных наблюдений Звенигородской астрономической
обсерватории ИНАСАН.....24

*Efim Kudashev, Alexander Belov, Natalya Kalenova,
Eugene Kudashev*

BIG DATA FROM SPACE.....32

E.B. Kudashev, M.A. Popov

Remote Sensing and International Co-operation in the Earth Science domain.....49

Мочалов В.А., Мочалова А.В., Никитина Н.Н., Шутов А.А.,
Маряхина А.А.

Некоторые вопросы применения открытых систем распределенных вычислений и обработки больших данных62

Н.А. Авдеева, К. К. Боярский

Метрики оценки тематической сегментации текстов.....74

Каневский Е.А.

К вопросу о пополнении словаря для синтаксического анализа
корпуса (на примере парсера SemSin).....81

Кошечева С.С.

Выделение двуязычных терминов и терминологических
словосочетаний на основе параллельного англо-русского
корпуса по языкам программирования.....93

<i>Кузнецов В.А. , Мочалов В.А. , Мочалова А.В.</i>	
Распределенная программная реализация упрощенного онтологически-семантического анализатора.....	99
<i>Леонтьева Н.Н.</i>	
Семантические тупики в системе АПТ.....	109
<i>Митрофанова О.А.</i>	
Построение тематических моделей корпусов энциклопедических текстов	122
<i>Мочалова А.В.</i>	
Вопросно-ответная система, основанная на сравнении семантических графов.....	128
<i>Ананьева Н.Б., Пугачев О.Н., Халиков Р.Г., Синев С.Ю., Смирнов И.С.</i>	
Коллекции Зоологического института РАН как информационная основа фундаментальных биологических исследований	140
<i>Косолапов К.А., Серебряков В.А., Теймуразов К.Б., Шорин О.Н.</i>	
Методы и средства интеграции и обогащения библиотечных данных.....	153
<i>Зеленчук А.М., Прядухин И.Ю., Молородов Ю.И., Фазлиев А.З.</i>	
Прикладная онтология по теплофизике. Проблема сведения.....	165
<i>Е.Д.Вязилов, Д.А.Мельников, Н.В. Чуняев, А.Е.Кобелев</i>	
Эксплуатация, модернизация и развитие новых форм отображения данных с помощью средств ЕСИМО.....	180
<i>А.В. Кошкарёв</i>	
Европейская программа INSPIRE и российские инициативы в области ИПД: критический анализ.....	194

*Гарцман Б.И., Красношеев С.М., Шулькин Е.В.,
Пашинский С.С., Меновицкова Т.С.*
Сервис ситуационного информирования о состоянии
водных объектов и развитии чрезвычайных
ситуаций в Приморском крае.....208

*Паниди Е.А., Казаков Э.Э., Капралов Е.Г.,
Терехов А.В.*
Внедрение вычислений на стороне клиента при реали-
зации веб-сервисов обработки пространственных дан-
ных на основе стандарта OGC WPS.....223

Смирнов С.В.
Построение системы массового распознавания
архивных документов с автоматической
корректировкой результатов.....239

Лурье И.К., Аляутдинов А.Р., Семин В.Н.
Разработка и использование сетевых информацион-
ных ресурсов пространственных данных для реше-
ния научных и прикладных задач.....246

Молородов Ю.И., Рюмкин С.И.
Структура информационно-вычислительной системы
для мониторинга зараженности территорий патоге-
нами вирусов переносимыми клещами....255

ПРЕДИСЛОВИЕ

В настоящем Сборнике представлены доклады Пятого Всероссийского симпозиума «ИНФРАСТРУКТУРА НАУЧНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ И СИСТЕМ», организованного Вычислительным центром им. А.А. Дородницына РАН и Институтом космических исследований РАН в рамках Плана научных конференций РАН в 2015 году. Заседания Симпозиума проходили 6–8 октября 2015 г. в Санкт-Петербурге в Санкт-Петербургском отделении Математического института им. В.А. Стеклова РАН.

Симпозиум проводится ежегодно в С.-Петербурге при поддержке РФФИ, начиная с 2011 г., что отражает возросшее значение и особую актуальность научных проблем, связанных с созданием и исследованием современных информационных систем для научных исследований. С учетом того, что в последние годы созданы принципиально новые технологии обработки данных, которые позволяют оперативно получать информацию о различных объектах и явлениях на очень больших территориях, необходимо отметить широкий интерес к фундаментальным и прикладным научным исследованиям в области развития инфраструктуры научных информационных ресурсов, а также к использованию геопроостранственных данных.

В связи с лавинообразным ростом объема цифровых научных данных, подходы к длительному хранению и непрерывному доступу к научным информационным ресурсам вызывают особый интерес. На Симпозиуме традиционно обсуждаются исследования цифровых e-Infrastructures с целью формирования распределенных научных информационных ресурсов, развития взаимосвязанных каталогов и создания сети интегрированных интероперабельных баз данных.

Важным научным направлением состоявшегося Симпозиума является обсуждение результатов исследований в области создания e-Science инфраструктуры научных данных и информационных систем, разработки программных средств по созданию такой инфраструктуры, включающих в себя средства сбора, моделирования и представления данных на основе международных стандартов, исследований информационных технологий электронного правительства и управления государственными информационными

ми системами, развития общества знаний на основе применения информационных технологий в социально-гуманитарной сфере.

В докладах на Симпозиуме отмечалось, что научные данные рассредоточены, их использование ограничено зачастую рамками того проекта, где они созданы, затруднен или невозможен поиск существующих данных и доступ к ним, не налажен обмен данными. Причина этого – отсутствие эффективной системы управления пространственными данными. Ее создание позволило бы интегрировать данные и знания о территории, строить и использовать модели природных и социально-экономических явлений и процессов, их взаимодействия в системе «общество – природная среда», использовать методы пространственного анализа, обеспечивать территориальное планирование и управление.

Развитие e-Science Infrastructures должно стать основой формирующихся систем коллективной работы исследователей на основе виртуального объединения информационных и вычислительных ресурсов. В нашей стране накоплен большой опыт использования геоинформационных технологий, реализованы многочисленные геоинформационные проекты, созданы информационные системы научных ресурсов.

На Симпозиуме были заслушаны доклады из различных регионов России: это Институт географии РАН, Зоологический институт РАН, Институт астрономии РАН, Институт космических исследований РАН, Санкт-Петербургский информационно-аналитический центр, Всероссийский научно – исследовательский институт гидрометеорологической информации – Мировой центр данных, Экономико-математический институт РАН, Институт космофизических исследований и распространения радиоволн ДВО РАН, Санкт-Петербургский государственный университет, Географический факультет Московского Государственного Тихоокеанский институт географии ДВО РАН, Приморское управление по гидрометеорологии и мониторингу окружающей среды, Вычислительный центр им. А.А. Дородницына РАН, Российская национальная библиотека, Московский государственный университет, Институт вычислительных технологий СО РАН, Сибирский государственный университет телекоммуникаций и информатики, Институт прикладных математических исследований КарНЦ РАН, Новосибирский государственный университет, Томский государ-

ственный университет, Институт оптики атмосферы СО РАН, Петрозаводский государственный университет, НИВЦ МГУ имени М.В.Ломоносова.

Основным направлением работы Пятого Симпозима были вопросы применения современных подходов в технологии развития информационных систем к задачам информационной поддержки научных исследований. Важным и интенсивно развивающимся направлением информационной поддержки научных исследований являются электронные библиотеки, технологии которых обеспечивают интеграцию разнородных данных. Исходя из целей ЭБ и анализа существующих систем, направленных на поддержку научных исследований, сформулированы следующие функциональные требования к модели ЭБ по научному наследию: надежное долговременное и защищенное от исчезновения хранение информации; актуальность, полнота, достоверность происхождения документов; историчность информации; географическая привязка информации; наличие большого числа словарей-классификаторов (справочников), для обеспечения идентификации и классификации ресурсов; поддержка неоднородных и слабо структурированных информационных ресурсов; поддержка взаимосвязей информационных ресурсов; предоставление информации пользователю в виде, выбранном пользователем; наличие интеллектуальных служб обслуживания запросов пользователя; наличие программных интерфейсов для поддержки аналитической работы пользователя с помощью программных приложений; поддержка требований интероперабельности как на программном, так и на семантическом уровне; поддержка работы с внешними источниками.

Редакторы и авторы Сборника выражают свою благодарность и признательность Российскому фонду фундаментальных исследований за поддержку исследований в области развития инфраструктуры научных информационных ресурсов и предоставленную возможность издания их результатов.

Е.Б. Кудашев, В.А. Серебряков

Коллекции Зоологического института РАН как информационная основа фундаментальных биологических исследований (грант РФФИ 15-29-02457).

*Ананьева Н.Б., Пугачев О.Н., Халиков Р.Г., Синева С.Ю.,
Смирнов И.С.*

Зоологический институт РАН, smiris@zin.ru

Аннотация

Начиная с 1985 года в Зоологическом институте РАН, ведутся работы по компьютеризации зоологических исследований и созданию информационно-поисковой системы по классификации и коллекциям животных. Статья посвящена истории создания системы и современному этапу разработки, ориентированному на оцифровку (дигитализацию) уникальных коллекций.

Ключевые слова: зоологические коллекции, Зоологический институт, изучение и сохранение биоразнообразия, банки ДНК, информационные системы, виртуальные (цифровые) музеи

Abstract

Since 1985 in the Zoological Institute of the Russian Academy of Science, works on a computerization of zoological researches and creation of an information retrieval system on classification and collections of animals are conducted. The article is devoted to history of creation of the system and the present stage of development focused on digitalization of the unique collections.

Keywords: zoological collections, Zoological Institute of RAS, studying and saving of biological diversity, banks of DNA, information systems, virtual (digital) museum

Зоологический институт РАН (ЗИН) обладает одной из крупнейших в мире зоологических коллекций, все элементы которой (единицы хранения, научные материалы, образцы, экспонаты, зоологические объекты) представляют собой основу для создания информационно-поисковой системы по уникальным сборам животных, насчитывающим сотни лет

(<http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2006/part4/SLPAV>) [24].

К настоящему времени коллекции ЗИН превышают цифру в 60 миллионов единиц хранения, которые зарегистрированы в книгах поступлений и бумажных каталогах (http://www.zin.ru/anim_r.htm) [4-11]. В состав коллекции ЗИН РАН входят и необычные структурные элементы: коллекции живых организмов, живых культур и банк ДНК животных самых различных таксонов [13, 14].

Изучение биологического разнообразия животных во всех его проявлениях – это фундаментальная научная задача, решаемая в рамках междисциплинарного комплекса систематических, зоогеографических, экологических, молекулярно-генетических исследований и возрастающего с каждым годом применения информационных технологий. Важнейшим инструментом и информационной основой такого рода исследований служат зоологические коллекции. В настоящее время эти уникальные материалы требуют активного вовлечения в научный оборот через создание интегрированных информационных систем по биоразнообразию, разрабатываемых специалистами-зоологами в процессе научного курирования ими коллекций по основным группам животного мира.

Одна из актуальнейших научных проблем, а также необходимое условие модернизации зоологических коллекций и алгоритмов их исследования – развитие информационных систем по биоразнообразию и разработка основ создания структуры этих систем с последующим интегрированием в международные распределенные информационно-поисковые системы [3, 23].

Работа по компьютеризации основных процессов накопления и анализа коллекционных, экспертных и литературных данных по систематике и фаунистике различных таксонов животных – основы изучения биоразнообразия – была начата в Зоологическом институте в 1985 г. на терминале БЭСМ-6, а в 1988 г. продолжена на СМ-1420. Приобретение персональных компьютеров (Amstrad-1640) в 1989-90 гг. ускорило разработку прикладных программ для зоологов. В 1991 г. родилась идея создания пакета программ, Зоологической ИНТЕгрированной системы (ЗООИНТ), которая могла бы обеспечить ввод информации о коллекциях и одновременно позволила бы анализировать накопленные данные с помо-

щью различных запросов [18]. В 1993 г. эта идея ИПС по зоологии получила поддержку Российского Фонда Фундаментальных Исследований (проект N 93-04-21216 "Система компьютерной интегрированной обработки данных по биоразнообразию животных (ЗООИНТ)"). В 1996 г. разработка проекта была продолжена при поддержке РФФИ (грант N 96-07-89086). Проект получил название "Развитие интегрированной системы ZOOINT для использования зоологических баз данных". В силу сложившихся в зоологических учреждениях России предпочтений на первом этапе создания ИПС нами была выбрана СУБД dBase III+, которая уступила место FoxBase, и, наконец, последняя трансформировалась в FoxPro, сначала под MS DOS, затем под Windows. К сожалению, система FoxPro, приобретенная корпорацией Microsoft, практически прекратила свое развитие. Тем не менее, освоение этих относительно несложных СУБД позволило зоологам, не привлекая сторонних специалистов и полагаясь только на свои силы, приступить к проектированию и созданию зоологических БД. При низком уровне заработной платы, который существует в академических учреждениях, привлечь со стороны высококвалифицированных программистов оказалось практически невозможно. В ходе выполнения проект немного трансформировался и дал результаты, несколько отличающиеся от планировавшихся ранее, но даже более ценные.

После появления первой пилотной версии ИПС возникла настоятельная потребность в связывании в единую сеть, поначалу компьютеров разработчиков, а затем и институтских подразделений, сотрудники которых могли бы пополнять БД и использовать уже накопленные данные в своих целях. Это был первый толчок к дальнейшему развитию системы ЗООИНТ. Вторым мощным толчком к развитию компьютерных зоологических систем стало появление мультимедийных технологий, которые резко расширили возможности накопления, хранения и использования зоологической информации (особенно изображений). К 1999 г. система ЗООИНТ поддерживалась СУБД FoxPro для DOS в локальной сети из 7 компьютеров. Был также создан сайт об ИПС ЗООИНТ (http://www.zin.ru/projects/zooint_r/). В 1999 г., по итогам конкурса РФФИ поддержал проект "Информационно-поисковая система ЗООИНТ по биоразнообразию животных" N 99-07-90315. Основной целью проекта служило продолжение разработки комплекса

программных средств для работы с классификацией животного мира, которые стали базой для создания конкретных БД и ИПС систематического, экологического и коллекционного характера. В проекте 1999 г. делалась ставка на интенсивное развитие локальной общепрофессиональной сети, которая позволила бы объединить усилия специалистов-зоологов по вводу данных по разным группам животных и, главное, воспользоваться уже накопленной в БД информацией.

Благодаря усилиям администрации, Институт, наконец, получил постоянный и достаточно быстрый доступ в Интернет. Была осуществлена реализация удаленного доступа к таксономической информации, с возможностью работать с базами данных ИПС ЗООИНТ в режиме on-line. Это потребовало не только обновления компьютерного парка разработчиков и пользователей, но и освоения новых программных средств: языка HTML, операционной системы Windows NT, технологии Active Server Pages (ASP) и др. В 2002 г. разработка веб-страницы проекта ИПС "ЗООИНТ" получила поддержку РФФИ (N 02-07-90217). По мере разработки проекта отдельные аспекты ЗООИНТ превратились в самостоятельные направления [19]. Многолетняя работа над коллекционными БД дала целое дерево ответвлений [20]. Центральный ствол был представлен интегрированной системой ЗООИНТ. Мощное ответвление представлено системой по мировой фауне блох PARHOST, развиваемой С.Г. Медведевым с коллегами. Другой проект, получивший поддержку РФФИ - информационная система для энтомологов - проект ZInsecta (РФФИ 01-07-90284). В Интернете открыта (<http://www.zin.ru/projects/zinsecta/index.html>) и действует двуязычная (рус/англ) информационная система. В ней объединены БД по систематике насекомых, по их распространению и по составу энтомологических коллекций с представлением в Интернете перечней таксонов, записей этикеток и изображений коллекционных экземпляров. Наиболее полно система охватывает отряды: жесткокрылых, сетчатокрылых, перепончатокрылых, стрекоз и двукрылых. Метод раскрывающихся папок позволил подключать к информационной системе через посредство организованных кодов неограниченное число БД. Характер присоединяемых баз данных не ограничен ни таксономически (можно использовать таксоны любого ранга), ни территориально (отдельные блоки ИПС вклю-

чают сведения по фауне Карелии, России, Палеарктики и всего мира). Наполнение БД отражает, в первую очередь, задачи, которые стояли при создании БД (перечни типовых экземпляров коллекции ЗИН, каталог материалов изучаемого таксона, хранящихся в музеях мира и т.д.) [17].

В 2001 году Институт выиграл по конкурсу государственную поддержку по программе "Информационная система по биоразнообразию" (Гос. контракт 01.20.0301251), 2002-2004 гг. В работе над проектом было задействовано 6 академических институтов. В ИПС BioDiv (<http://www.zin.ru/BioDiv/index.html>) в едином Интернет-пространстве проведено объединение иллюстрированных БД и самостоятельных информационных систем по биологическому разнообразию России. Как сами базы данных, так и их совмещение с использованием единой поисковой системы "Биоразнообразие России" выполнены высококлассными специалистами - учеными ведущих биологических институтов страны и разработчиками компьютерной версии системы. В 2002 г. ЗИН заключил государственный контракт на выполнение научно-исследовательских и опытно-конструкторских работ по теме "Информационная система по биоразнообразию". Соисполнителями стали Институт проблем экологии и эволюции РАН (ИПЭЭ), Ботанический институт РАН (БИН), Институт физиологии микроорганизмов РАН (ИФМ) и Институт цитологии и генетики СО РАН (ИЦиГ). Основной целью проекта "Информационная система по биоразнообразию" (ИСБР) было создание комплекса программных средств и БД для работы с классификацией животного и растительного мира, которые будут служить основой формирования информационно-поисковой системы по биоразнообразию России, поддерживающей неоднородные коллекции распределенных информационных ресурсов, содержащих сведения систематического, коллекционного и экологического характера. Основные задачи ИСБР состояли в разработке стандартов, форматов и методологии создания единой национальной распределенной БД по биологическому разнообразию в рамках российского сегмента глобальной сети Интернет, и создание пилотной информационной сети по видовому разнообразию России, объединяющей БД, поддерживаемые институтами-соисполнителями, по видовому составу всех таксонов (микроорганизмов, растений и животных) и данные о

коллекциях, хранящиеся в этих учреждениях. Исходными материалами для проведения работы служили результаты исследований структурных подразделений ЗИН, БИН, ИПЭЭ, ИФМ и ИЦиГ в области биологии, в том числе уже созданные и частично опубликованные ими научные продукты в области систематики животных и растений.

Объектом исследования участников проекта служили названия бактерий, простейших, грибов, растений и животных, обитающие на территории России и сопредельных территориях. В процесс работы над проектом были проведены теоретические и прикладные исследования, а также разработаны и усовершенствованы прикладные программы, БД, информационные системы и Интернет-сайты по различным таксономическим группам живых организмов [21, 22].

Дальнейшим шагом в реализации изучения биоразнообразия является вовлечение в исследовательский процесс на основе информационных подходов уникальных коллекций, которыми обладает ряд зоологических учреждений России.

Одна из актуальнейших научных проблем, а также необходимое условие модернизации зоологических коллекций и алгоритмов их исследования – развитие информационных систем по биоразнообразию и разработка основ создания структуры этих систем с последующим интегрированием в международные распределенные информационно-поисковые системы [1-3, 12, 15, 17].

В настоящее время Зоологический институт Российской Академии Наук приступил к реализации конкретной фундаментальной задачи, которая состоит в разработке алгоритма дигитализации материалов фондовой коллекции Зоологического института РАН, учитывающего специфику их хранения для отдельных систематических групп. Учитывая гигантский объем этой коллекции (более 60 миллионов единиц хранения и десятки тысяч экземпляров первичных типов), на первом этапе работы целесообразно ограничиться рядом модельных таксонов, относящихся к основным классам животного мира. Основное внимание предполагается сосредоточить на эталонных (типовых) экземплярах, являющихся носителями научных названий видов и подвидов. Поиск таких экземпляров, их правильное обозначение и исследование с применением самых современных методов представляет собой научную

задачу первостепенной важности, поскольку любые неточности, допущенные при типификации таксонов, могут привести к цепи ошибочных заключений на всех уровнях изучения биологического разнообразия и существенно исказить, а то и совершенно обесценить полученные результаты. Особую актуальность имеет изучение типовых экземпляров видов, установленных в XVIII-XIX веках еще до создания Международного кодекса зоологической номенклатуры, зачастую недостаточно полно описанных и специально не обозначенных, что создает серьезные проблемы в плане идентификации таксонов и правильном применении их названий. В качестве модельных групп для проведения работ по согласованию структуры каталога образцов, созданию интегрированных информационных систем образцов фондовых коллекций будут использованы следующие группы наземных и водных беспозвоночных: для типа членистоногие (Arthropoda) модельные группы выделены среди обширного (более 1 миллиона видов) класса насекомых (Insecta) - отряд блохи (Siphonaptera), отряд чешуекрылые (Lepidoptera), отряд жесткокрылые, или жуки (Coleoptera), семейство комаров (Culicidae) отряда Diptera; среди класса ракообразных (Crustacea) - бокоплавы, или амфиподы, или разноногие раки (Amphipoda) надотряда Peracarida; морские беспозвоночные типа иглокожих (Echinodermata) и класса погонофор (Pogonophora), а также два класса позвоночных животных: земноводные (Amphibia) и млекопитающие (Mammalia). Информационная система должна включать в себя таксономические, типовые, коллекционные, зоогеографические и библиографические сведения.

Применяемые программные алгоритмы основаны на передовых отечественных (стандарт ZOOCOD для иерархических классификаторов) и мировых разработках (стандарты геореференсирования точек находок National Science Foundation, прикладные интерфейсы он-лайн картирования Google Maps API, алгоритмы он-лайн публикации изображений сверхвысокого разрешения на базе сервера изображений IPIImage Server), что подразумевает возможность последующей интеграции создаваемой информационной системы в распределенные зоологические он-лайн ресурсы.

С использованием имеющейся серверной инфраструктуры ЗИН РАН и создаваемой информационной системы коллекционных образцов, станет возможным развертывание точки публи-

ции данных GBIF (Integrated Publishing Toolkit, IPT, <http://www.gbif.org/publishingdata/howtopublish>) и последующей выборочной публикации коллекционных данных ЗИН РАН на портале Глобального информационного средства по биологическому разнообразию (GBIF) - международной организации, которая стремится сделать доступными мировые научные данные по биоразнообразию с помощью Интернета для всеобщей выгоды и научного, и мирового сообщества. В 2013 году в сети GBIF были задействованы 462 поставщика данных из 40 стран мира. Полученные данные были объединены в 9960 баз данных.

Инфологическая структура информационной системы построена по технологии «клиент-сервер» и включает: 1) уровень данных (уровень «сервера») – таксономический классификатор в стандарте ZOOCOD 4, коллекционные каталоги образцов, географический классификатор, вспомогательные и связочные словари, таблицы и пр.; 2) уровень представления данных («клиентский» уровень) – набор пользовательских интерфейсов для работы с информационной системой кураторов коллекций в среде сети интранет и публичного доступа посредством сети Интернет. Программно информационная система реализуется посредством клиент-серверной СУБД корпоративного уровня Microsoft SQL Server 2012 и универсального веб-интерфейса (технология ASP, кросс-платформенный скриптовый язык JavaScript, технология Ajax). Согласованная инфологическая структура каталога коллекционных образцов будет реализована в виде основных и вспомогательных таблиц в среде СУБД Microsoft SQL Server 2012. Среди традиционных для таблиц коллекционных данных полей (таких как, каталожный и полевой номера, оригинальная этикетка, экологическая характеристика биотопа, характеристика субстрата, типовой статус, коллектор, дата сбора и др.) предусмотрены поля, содержащие графическую информацию (изображения оригинальной этикетки, собственно коллекционного образца, дополнительных материалов и др.). Особое значение уделяется сохранению как информации в аутентичной форме (данные, как они есть в первоисточнике), так и сведений об обработке (интерпретации) данных разными экспертами и разными методами. Инфологическая структура создаваемой информационной системы будет выполнена с учетом перспективы использования стандартов геореференсирова-

ния, разработанных для международных проектов National Science Foundation по созданию распределенных информационных систем (<http://manisnet.org/GeorefGuide.html>). Это дает гибкость для дальнейшего развития системы и возможность расширения ее функционала без существенного изменения внутренней структуры. Получаемые координаты и дополнительные данные будут храниться в специализированных таблицах геопроектирования точек находок, смогут предоставляться он-лайн посредством программного интерфейса Google Maps API (<https://developers.google.com/maps/>), а в перспективе служить для интеграции в распределенные зоологические информационные системы. Большое внимание будет уделено вопросам производительности и масштабируемости внедряемых решений. Предусмотрено выполнение всесторонних тестов масштабируемости как на серверном уровне «данных» с последующей оптимизацией специализированных элементов программирования SQL (скалярные и табличные функции, хранимые процедуры, триггеры) в СУБД Microsoft SQL Server 2012, так и на клиентском уровне интерфейса с последующей оптимизацией JavaScript-сценариев и подключаемых модулей. Удобный и функциональный административный пользовательский веб-интерфейс информационной системы по образцам фондовых коллекций – ключевой аспект работы с ней и единственно возможный в рамках клиент-серверной технологии способ взаимодействия конечного пользователя-куратора коллекций с унифицированным серверным хранилищем коллекционных данных. Создаваемая информационная система по образцам фондовых коллекций, в первую очередь типовым, уже на этапе ее размещения в публичном доступе на веб-портале ЗИН РАН не будет иметь ограничений для работы в напряженном многопользовательском режиме с разделяемым доступом к данным. Особое значение оцифровка приобретает для типовых образцов, многократно повышая их доступность для научной общественности. Создаваемая в рамках проекта информационная система изначально предусматривает наличие библиотеки изображений коллекционных образцов и сопутствующих материалов. Для получения высококачественных изображений при этом будут широко применяться самые различные методы дигитализации. К их числу относятся сканирование коллекционных образцов и сопутствующих материалов (оригинальные этикетки, каталожные

карточки, конверты, рентгеновские снимки и т. п.) на планшетных CCD-сканерах высокого разрешения, а также специализированная фотосъемка (макросъемка) с применением полупрофессиональных цифровых зеркальных фотокамер в комплекте со специальными макрообъективами и системами импульсных источников света (макровспышками) [15, 16].

Организация коллекции генетического материала, пригодного для исследования различными методами в настоящее время и в будущем, чрезвычайно важна для разработки и реализации стратегии сохранения биоразнообразия, как и информационно-поисковая система о хранящихся экземплярах. Будет использован международный опыт в организации хранения, документирования и создания информационных систем данных ваучерных экземпляров коллекций различных таксономических групп беспозвоночных и позвоночных животных, включая образцы ДНК хранящихся в ЗИН экземпляров и оцифровку изображений. Планируемая интегрированная информационная система позволит собрать полную информацию об имеющемся биоразнообразии России, где каждая составная часть — коллекции, банк ДНК и информационный банк данных об экземплярах будет нести свои специфические функции и сможет адекватно обслуживать потребности биологической науки и технологии в новом тысячелетии в соответствии с современными принципами и стандартами мировых зоологических коллекций.

Литература

1. Ананьева Н.Б., Голынский Е.А. Анализ распространения горной кольцехвостой туркестанской агамы *Paralaudakia lehmanni* (Nikolsky, 1896): использование программы Максент // Труды Зоологического института РАН. 2013. Т. 317, № 4, С. 54–84.
2. Ананьева Н.Б., Туниев Б.С., Орлов Н.Л., Туниев С.Б. Опыт использования критериев редлистинга МСОП в создании глобальной базы данных по амфибиям и рептилиям мировой фауны и Кавказского Экорегиона // В кн.: Вопросы герпетологии, 2011. С. Петербург, С. 17–27.
3. Khalikov R.G. ZooDiv – an information storage and retrieval

- biodiversity system and its implementation for varied zoological projects. Российско-китайский семинар «Исследование и охрана амфибий и рептилий Евразии: результаты и перспективы сотрудничества», 29 июля - 3 августа, 2009, Санкт-Петербург, С. 19.
4. Синев С.Ю. Типовые экземпляры узкокрылых огневок (*Lepidoptera*, *Phycitidae*), хранящиеся в коллекции Зоологического института АН СССР. 1 // Энтомологическое обозрение. 1990а. Т. 69, вып. 1. С. 118-133.
 5. Синев С.Ю. Типовые экземпляры узкокрылых огневок (*Lepidoptera*, *Phycitidae*), хранящиеся в коллекции Зоологического института АН СССР. 2 // Энтомологическое обозрение. 1990б. Т. 69, вып. 2. С. 419-431.
 6. Синев С.Ю. Каталог роскошных узкокрылых молей (*Lepidoptera*, *Phycitidae*) мировой фауны. Санкт-Петербург: ЗИН РАН, 2002. 184 с.
 7. Алимов А.Ф., Пугачев О.Н., Синев С.Ю. Пополнение и сохранение зоологических коллекций // В сб.: Наука и техника: вопросы истории и теории. Вып. 18. Санкт-Петербург, 2002. С. 76-78.
 8. Синев С.Ю. (ред.). Каталог чешуекрылых (*Lepidoptera*) России. Санкт-Петербург-Москва: Товарищество научных изданий КМК, 2008. 424 с.
 9. Синев С.Ю. О соотношении классических и новейших методов исследования в систематике насекомых // Энтомологическое обозрение. 2011. Т. 90, вып. 4. С. 821-832.
 10. Синев С.Ю. Каталог серых молей (*Lepidoptera*, *Phycitidae*) мировой фауны. Санкт-Петербург: ЗИН РАН, 2014а. 108 с.
 11. Синев С.Ю. Коллекция Э.А. Эверсмана и ее значение для работ по систематике палеарктических чешуекрылых // Труды Казанского отделения Русского энтомологического общества. 2014б. Вып. 3. С. 78-83.
 12. Ananjeva N.B., Iohanssen L.K., Golynsky Y.A. Progress and harmonization in reptile database in Zoological Institute, Russian Academy of Sciences. Российско-китайский семинар «Исследование и охрана амфибий и рептилий Евразии: результаты и перспективы сотрудничества», 29 июля - 3 августа, 2009, Санкт-Петербург: 13.

13. Абрамсон Н.И., Ананьева Н.Б., Подлипаев С.А., Пугачев О.Н. Коллекции ДНК - новый путь изучения и сохранения биологического разнообразия // *Фундаментальные зоологические исследования*. 2004. Теория и методы. КМК, М.-СПб., С. 273-282.
14. Ананьева Н.Б., Подлипаев С.А., Пугачев О.Н. От кунсткамеры до банка генетических ресурсов. *Природа*. 2002. № 8. С.17.
15. Халиков Р.Г., Н.Л. Орлов. Современные компьютерные технологии обработки изображений: опыт практически применения в зоологических исследованиях // *Материалы Первого съезда Герпетологического общества им. А.Н. Никольского*. 2001. Пущино-Москва. С. 309-312.
16. Khalikov R.G., Orlov N.L. Experience of practical use of modern digital imaging techniques in zoological research // *Annual Reports of ZIN RAS*. 2001. P. 47-50.
17. Кривохатский В.А., Лобанов А.Л., Медведев Г.С., Белокобыльский С.А., Дианов М.Б., Смирнов И.С., Халиков Р.Г. Информационная система по энтомологическим коллекциям в Интернете // *Труды Русского энтомологического общества*. 2003. Т. 74. СПб. С. 59-70.
18. Lobanov A., Sokolov E., Smirnov I. ZOOINT - an integrated system for zoological data bases. - In: *Abstracts of International workshop on advances in databases and information systems (ADBIS'94)*. Moscow, May 23-26, 1994. P. 52,53.
19. Смирнов И.С., А.Л.Лобанов, А.Ф.Алимов, С.Г.Медведев, А.А.Голиков. Итоги развития проекта ЗООИНТ и его дальнейшие перспективы // *Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды Четвертой Всероссийской научной конференции RCDL'2002*, (Дубна, 15-17 октября 2002 г.): в 2 т. – Дубна: ОИЯИ, 2002. Т. 2: 308-315. (на русском языке с английским резюме)
20. Лобанов А.Л., Смирнов И.С. Место и роль информационных технологий в исследованиях Зоологического института РАН // *Фундаментальные зоологические исследования: Теория и методы*. (По материалам Международной конференции «Юбилейные чтения, посвященные 170-летию Зоо-

- логического института РАН», проходившей 23-25 октября 2002 г.), М.-СПб: Товарищество научных изданий КМК. 2004: 283-318 (резюме на англ. яз.).
21. Пугачев О. Н., Алимов А. Ф., Лобанов А. Л., Кривохатский В. А., Смирнов И. С. Первые итоги разработки информационной системы по биоразнообразию России (BIODIV-ZOODIV) // "Информационные системы и web-порталы по разнообразию видов и экосистем". Материалы международн. симпозиума. Борок, 28 ноября - 1 декабря 2006 г., Москва, Тов-во науч. изд. КМК, 2006. С. 170-173.
 22. Смирнов И.С., Пугачев О.Н., Кирейчук А.Г., Дианов М.Б., Лобанов А.Л., Халиков Р.Г., Голиков А.А., Кривохатский В.А. Итоги и перспективы информационной системы по биоразнообразию животных России (ZOODIV – BIODIV) // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды Двенадцатой Всероссийской научной конференции «RCDL’2010» (Казань, Россия, 13-17 октября 2010 г.): – Казань: Казанский университет, 2010: 461-464. (на русском языке с английским резюме)
 23. Лобанов А.Л., Кирейчук А.Г., Пугачев О.Н., Смирнов И.С. Зоологические коллекции, интерактивные определители и Интернет // Зоологические коллекции России в XVIII-XXI веках: социально-политический и научный контекст. / Ред.-сост. Н.В. Слепкова. СПб: Изд-во СПбГЭТУ «ЛЭТИ», 2012: 123-132.
 24. Смирнов И.С., Лобанов А.Л., Пугачев О.Н., Алимов А.Ф., Воронина Е.П. Электронные коллекции в зоологии и электронные библиотеки // Электронные библиотеки, 9 (4). 2006
<http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2006/part4/SLPAV>) ISSN 1562-5419 Институт развития информационного общества (ИРИО) института РАН как информационная основа фундаментальных биологических исследований (грант РФФИ 15-29-02457).

Инфраструктура научных информационных ресурсов и систем
Сборник избранных научных статей

Подписано в печать 28.05.2015

Формат бумаги 60×84 1/16

Уч.-изд. л. 15. Усл.-печ. л. 16,25

Тираж 120 экз. Заказ 37

Отпечатано на ротапринтах
в Федеральном государственном бюджетном учреждении науки
Вычислительный центр им. А.А. Дородницына Российской академии наук
119333, Москва, ул. Вавилова, 40



**Федеральное государственное
бюджетное учреждение науки
Вычислительный центр им. А.А. Дородницына
Российской академии наук
119333, Москва, ул. Вавилова, 40**

ISBN 978-5-91601-127-2



9 785916 011272