

Международные проекты по созданию электронных коллекций морских животных: первые результаты*

© И.С.Смирнов, А.Ф.Алимов, А.Г.Кирейчук, Е.П.Воронина, А.Л.Лобанов

Зоологический институт РАН
smiris@zin.ru

Аннотация

В работе описываются международные проекты по созданию коллекционных баз данных и информационных систем. Обсуждаются вопросы разработки стандартов описания электронных биологических коллекций и проводится сравнение различных структур баз данных и программ по их заполнению. Отмечается важность коллекционных баз данных в изучении и сохранении биоразнообразия, контроле над вредными видами-вселенцами и при создании систем экологического мониторинга.

1. Изучение морского биоразнообразия и компьютерные технологии

Создание биологических (зоологических, ботанических) информационно-поисковых систем - способствует быстрой и качественной экологической экспертизе, как в интересах рационального морского природопользования, так и при обеспечении экологической безопасности при освоении различных месторождений полезных ископаемых на дне российских морей. Недавнее международное совещание в Гамбурге (ноябрь 2004 г.; <http://www.vliz.be/obi/>) показало существенную востребованность информационно-поисковых систем по биоразнообразию и географическому распространению морских организмов [1].

К сожалению, единая информационно-поисковая система (ИПС) по биоразнообразию российских морей до сих пор отсутствует, несмотря на большое число проектов, в том числе и международных [2,3]. До сих пор неизвестно общее число видов различных животных, обитающих в морях России.

По предварительным данным, полученным в Лаборатории морских исследований Зоологического института (ЗИИ) для арктических морей установлены следующие величины: Баренцево море

населяют 3245 видов беспозвоночных; беломорская фауна, или обедненная баренцевоморская фауна, включает 1817 видов; далее количество видов устойчиво снижается в восточном направлении: 1671 - для Карского моря, 1472 - для моря Лаптевых, 1011 - для Восточно-Сибирского, и 1168 разновидностей для Чукотского моря. Известно 837 видов из глубоководной части Центрального Арктического Бассейна, смежной с евразийскими морями [9].

Видовой состав населения морей необходимо знать не только из чисто биологического интереса. Очень остро последнее время стоит проблема видов-вселенцев, которые способны серьезно нарушать экологическую обстановку и вести к колоссальным экономическим потерям. Эта проблема может быть решена, только тогда, когда хорошо известны местные фауна и флора, т.е. состав населения водоемов, и тогда, когда на протяжении многих лет можно проследить динамику появления чужеродного вида и захват им новых акваторий и мест обитания. Без современных информационных технологий механизм подобного мониторинга создать невозможно, учитывая современные средства передвижения, способствующие быстрому пассивному переносу нежелательных чужеродных видов.

Общей проблемой создания единой ИПС для морей России является отсутствие эффективной координации. Успешно разрабатываемая Единая система информации о Мировом океане (ЕСИМО) ориентирована в основном на океанографические данные. К тому же, относительно небольшие деньги, выделяемые на развитие программы, не позволяют привлекать широкий круг специалистов из разных ведомств - отсюда определенная однонаправленность и предсказуемая неполнота данных.

Труды 7^{ой} Всероссийской научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции" - RCDL'2005, Ярославль, Россия, 2005.

Очень серьезной проблемой для развития ИПС, в том числе и географического профиля, является разработка иерархической системы географических названий и терминов (формализация соподчинения различных терминов, особенно для морских акваторий: океан – море – залив – пролив – бухта, составление электронных тезаурусов и т.д.).

С конца 80-х годов в Зоологическом институте начали разрабатываться электронные базы данных сначала для морских беспозвоночных, а затем, с 1991 г., для рыб. Создана ИПС «ОКЕАН», которая была ориентирована на коллекционные данные экземпляров морских животных, хранящиеся в ЗИНе [4,5,6,7,8].

2. Международные проекты по созданию электронных коллекций

В последнее время все большее развитие получают международные проекты по созданию баз данных и информационно-поисковых систем по биоразнообразию. Это такие известные проекты как “Species 2000” (<http://www.sp2000.org/>), GBIF (<http://www.gbif.org/>), FishBase (<http://www.fishbase.org/home.htm>) и др.

Международные проекты по созданию электронных коллекций по морским животным только сейчас начинают активно разрабатываться. Одним из перспективных проектов, получившим поддержку Фонда Альфреда Слоана (Sloan Foundation), является глобальный Интернет-проект «Перепись морской жизни» (Census of Marine Life или CoML; <http://www.coml.org/>). Информационная составляющая проекта представлена “Океанской Биогеографической информационной системой” (Ocean Biogeographic Information System или OBIS; <http://www.coml.org/descrip/obis.htm> или <http://www.iobis.org/Welcome.htm>).

Американским Национальным океанографическим центром данных (NODC; <http://www.nodc.noaa.gov/General/NODC-cdrom.html>) с помощью российских специалистов выпущено несколько лазерных дисков с информацией о российских исследованиях и данных, в том числе и по биологии, в частности, планктону: Зоопланктон арктических морей (Zooplankton of the Arctic Seas 2002, International Ocean Atlas Series: Volume 6); 36-летние данные по зоопланктону Белого моря (36-Year (1963-1998) Time Series of Zooplankton, Temperature, and Salinity in the White Sea, International Ocean Atlas Series: Volume 7) и данные по планктону Баренцева и Карского морей (Biological Atlas of the Arctic Seas 2000: Plankton of the Barents and Kara Seas, International Ocean Atlas Series: Volume 2).

Несомненен прогресс и в области баз данных для донных организмов морей и океанов. При поиске в Интернете сайтов, на которых могут быть представлены коллекции морских донных беспозвоночных, вы получаете уже на десятки тысяч ссылок больше, по сравнению с

предыдущими годами. Но истинно работающих электронных коллекций все же мало. По-прежнему, лидирующее положение в презентации своих коллекций морских беспозвоночных занимает американский Национальный музей Естественной истории в Вашингтоне, округ Колумбия, США (NMNH; <http://goode.si.edu/webnew/pages/nmnh/iz/Query.php>).

Причем наряду с пополнением базы данных о находках видов, растет коллекция изображений оригинальных этикеток и собственно изображений коллекционных экземпляров. К сожалению, изображения логически не связаны с коллекционной базой данных, что затрудняет анализ этой весьма обширной электронной коллекции.

3. Создание стандартов разработки и представления данных по морским электронным коллекциям

Особое место при создании баз данных и информационных систем занимает разработка стандартов для представления данных и описания баз данных и ИПС.

Началом истории компьютеризации и стандартизации коллекционных данных можно считать создание Рабочей группы по Таксономическим Бадам данных. Первое совещание Рабочей Группы по Таксономическим Бадам данных (Taxonomic Databases Working Group или TDWG) состоялось в Женеве, в 1985 году.

Международная Рабочая Группа (The International Working Group on Taxonomic Databases) – некоммерческая научная и образовательная организация, входящая в Международный Союз Биологических наук (IUBS), призванная развивать международное сотрудничество в области биологических баз данных. Первоначально она была ориентирована на разработку стандартов для описания таксономических баз данных ботанических учреждений и проектов. В 1994 году Группа включила в круг своих объектов базы данных по животным и микроорганизмам.

В сентябре 2005 г. Группа по Таксономическим Бадам данных проведет свое рабочее совещание в стенах одного из крупнейших мировых хранилищ, в Зоологическом институте РАН в Санкт-Петербурге, где будут обсуждаться вопросы стандартизации и интеграции информации по биоразнообразию.

В 1992 году Комитет по компьютеризации и сетям Ассоциации систематических коллекций (ASC) провел рабочее совещание по стандартизации данных биологических коллекций, на котором рассматривались различные информационные модели коллекционных данных и предлагалось для описания коллекций использовать 55 элементов метаданных или сущностей.

Дарвинское ядро набора метаданных (Darwin Core), специально разрабатываемое для биологических данных, содержит от 22 до 36

элементов. Относительная простота набора метаданных делает в некоторых случаях Дарвинское ядро неадекватным для описания данных. Предшествующие версии Дарвинского ядра-2 были объединены с поисковым протоколом DiGIR - Distributed Generic Information Retrieval (<http://darwincore.calacademy.org/>).

Постепенно начинает вырисовываться определенный стандарт описаний коллекционных баз данных.

К сожалению, при всем обилии совещаний и публикаций, в том числе и в Интернете [<http://iobis.org/publications/>; <http://www.tdwg.org/otherlinks.html>] до сих пор отсутствует схема пригодная для практического составления приемлемого набора данных и описания ИПС.

Лаборатория ихтиологии нашего института участвует в международном проекте по созданию баз данных по рыбам (проект FishBASE). Для заполнения баз данных используется программа ввода «Artdian». Нами была предпринята попытка, сравнить эту программу ввода и программу ввода из ИПС «ОКЕАН», чтобы выявить преимущества и недостатки обеих.

Преимущества программы из ИПС «ОКЕАН»:

1. Стационарный уникальный код генерируется автоматически.

2. Важные для анализа поля могут заполняться по-русски (судно и др.) Это важно, так как при переводе или транслитерации возможно потеря данных для последующего поиска.

3. Пробелы удаляются программой автоматически.

4. Имеются поля для орудий лова, станций, заливов (для морских экспедиций это является иногда единственной зацепкой при поиске неполных данных с повторных этикеток).

5. Дата определения имеет текстовый формат (часто полная дата отсутствует и при использовании поля даты информация пропадает).

6. Имеется поле экспедиций (коллекторы или сборщики могут быть разные).

7. Заполнение всех географических позиций осуществляется с помощью словарей.

8. Число экземпляров не является обязательным полем и т.д.

Преимущества программы "Artdian" из проекта FishBASE:

1. Поля для пресноводных станций более детализованы (особенно, что касается суши).

2. По состоянию объекта заполняется большее число полей (производится более полное описание объекта – коллекционного экземпляра).

3. Блок полей «Коллекторы» состоит из нескольких полей.

4. Классификатор названий встроен внутрь программы ввода.

5. Проведена стандартизация в соответствии с Darwin Core - для полей координат, в делении

станций на суше (по бассейнам озер, рек, штатам, провинциям) и т.п.

Проведенное сравнение показывает наличие определенных расхождений в подходах и предполагает изучение опыта различных команд с дальнейшей попыткой создать не идеальную, но оптимальную систему для ввода и анализа данных по морским животным, главным образом, донным беспозвоночным и рыбам [8].

Заключение

Важным моментом развития коллекционных баз данных и ИПС является создание на их основе полноценных справочных пособий и определителей [10,11,12].

Своевременная идентификация тех же чужеродных вселенцев и принятие соответствующих мер предотвращающих их нежелательное распространение может сэкономить колоссальные финансовые и человеческие ресурсы.

Система «ОКЕАН», рассчитанная на ведение электронных коллекций Лабораторий морских исследований и ихтиологии Зоологического института, развивалась из необходимости обобщения большого фактического материала, который невозможно адекватно анализировать с помощью бумажных технологий. Не получая прямой поддержки (несмотря на неоднократные попытки таковую получить) ИПС «ОКЕАН» имела распространение только в стенах Зоологического института. В этом году Российский фонд фундаментальных исследований поддержал идею создания виртуального музея на Интернет-портале Зоологического института. Это создает хорошую перспективу для участия сотрудников ЗИНа не только в данном проекте, но и в других активно развиваемых отечественных и международных программах по созданию электронных коллекций и музеев, а также для разработки реально работающего сегмента виртуального музея по морским донным животным.

Литература

- [1] Ocean Biodiversity Informatics International Conference on Marine Biodiversity Data Management Hamburg, Germany: 29 November to 1 December 2004: <http://www.vliz.be/obi/>
- [2] I.S. Smirnov, E.P. Voronina, A.L. Lobanov & A.V. Neyelov. 2004. The information system of the marine animals collection (fish and invertebrates) in the Zoological Institute Russian Academy of Sciences. Ocean Biodiversity Informatics. International Conference on Marine Biodiversity Data Management. Hamburg, Germany: 29/11-1/12/2004: 27.
- [3] Воронина Е.П., Смирнов И.С., Голиков А.А. Компьютерные методы в ихтиологических исследованиях Зоологического института // Информ.-поисковые системы в зоологии и

- ботанике. Тез. докл. Межд. симп. Тр. ЗИН РАН Т. 278. СПб, 1999. С.116-117.
- [4] И.С.Смирнов, А.В.Смирнов. "ОКЕАН" - банк данных по коллекциям морских беспозвоночных Зоологического института РАН // Базы данных и компьютерная графика в зоологических исследованиях, Труды Зоологического института, т. 269, Санкт-Петербург, 1997: 133-135.
- [5] А. А. Голиков, И. С. Смирнов, А. Л. Лобанов, М. Б. Дианов. Итоги развития ИПС "ОКЕАН" и банка данных по коллекциям морских беспозвоночных Зоологического института РАН. // Информационно-поисковые системы в зоологии и ботанике, Труды Зоологического института РАН, т. 278, Санкт-Петербург, 1999: 103-104.
- [6] И.С.Смирнов, А.Л.Лобанов, А.А.Голиков, А.В.Неелов, Е.П.Воронина. Информационно-поисковая система «Океан» по коллекциям беспозвоночных и рыб Зоологического института РАН. Информационные ресурсы об океане – актуальные проблемы формирования, распространения и использования в научных исследованиях и в морской деятельности. Тезисы докладов конференции (8-10 октября 2002 г., г. Обнинск). – Изд. отдел ВНИГМИ-МЦД, 2002. с. 73, 74.
- [7] И.С. Смирнов, А.Л. Лобанов, А.А. Голиков, Е.П. Воронина, А.В. Неелов. Информационно-поисковая система «ОКЕАН» // Информационные системы по биоразнообразию видов и экосистем. Научная программа и тезисы 4-го Международного симпозиума. СПб. 2003. с. 15, 16.
- [8] И.С.Смирнов, Е.П.Воронина, А.Л.Лобанов, А.А.Голиков, А.В.Неелов. Создание информационно-поисковых систем по коллекциям морских животных (рыб и беспозвоночных) в Зоологическом институте РАН // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды Шестой Всероссийской научной конференции RCDL'2004, (Пушино, 29 сентября - 1 октября 2004 г.): – Москва, типография ООО «Мультипринт», 2004: 30-33. (на русском языке с английским резюме)
- [9] List of species of free-living invertebrates of Eurasian Arctic seas and adjacent deep waters (ed. by V.I.Sirenko). Explorations of the fauna of the seas. V. 51 (59). 2001. 132 pp.
- [10] Лобанов А.Л., Рысс А.Ю. Компьютерные идентификационные системы в зоологии и ботанике: современное состояние и перспективы // Информационно-поисковые системы в зоологии и ботанике (Тезисы международного симпозиума, май 1999). Труды Зоологического института РАН. 1999. Т. 278. с. 17-19.
- [11] Lobanov A.L. Biological identification with computers: 30 years of evolution // Компьютерные

базы данных в ботанических исследованиях. Сборник научных трудов. 1997. С. 51-55.

- [12] 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 Зоологического института РАН», проходившей 23-25 октября 2002 г.), М.-СПб.: Товарищество научных изданий КМК. 2004: 283-318 (резюме на англ. яз.).

The international projects on creation of digital collections of marine animals: the first results

I.S. Smirnov, A.F. Alimov, A.G. Kirejtshuk,
E.P. Voronina, A.L. Lobanov

The international projects on creation of collection data bases and information systems are described. The problems on development of the standards of the description of electronic biological collections are discussed and the comparison of various structures of data bases and programs on their filling is carried out. The importance of collection data bases in study and preservation of biodiversity, monitoring above harmful alien invasive species and for creation of systems of ecological monitoring is marked.

* Работа по теме осуществляется при поддержке грантов РФФИ N 05-07-90179-в и 05-07-90354-в, проекта N11 "Исследование Антарктики. Проведение комплексного изучения антарктической биоты", Федеральной Программы "Мировой Океан", программ «Биоразнообразию», Arctic Census of Marine Life (ArcCoML) и контракта № 43.700.11.0002.