

Эволюция стандарта ZOOCOD – концепции отражения зоологических иерархических классификаций в плоских таблицах реляционных баз данных*

© А.Л. Лобанов, И.С. Смирнов, М.Б. Дианов, А.А. Голиков, Р.Г. Халиков

Зоологический институт РАН
smiris@zin.ru

Аннотация

Биологические базы данных опираются на таксономические таблицы, в которых должны полно отражаться сложные многоуровневые иерархические схемы классификации таксонов. Отражение иерархий в плоских таблицах реляционных баз данных является нетривиальной задачей. Несколько способов решения этой проблемы известны в теории информационных систем. В Зоологическом институте РАН (ЗИН) большинство этих способов было разработано независимо в ходе многолетних работ по созданию зоологических баз данных. Все эти разработки отражены в стандарте ZOOCOD, многократно описанном в ряде публикаций, успешно используемом уже более 20 лет и распространенном сейчас уже и за пределами ЗИН (БИН РАН, ИПЭЭ РАН, МГУ, МГПИ, Нижегородский госуниверситет и др.). На основе этого стандарта создана и развивается национальная информационная система по биоразнообразию России (BIODIV). Концепция представления многоуровневых иерархий в реляционных базах данных успешно перенесена с таксономии на географические данные, сведения по морфологии и экологии животных.

1 Появление концепции ZOOCOD

Крупнейшие в России и одни из крупнейших в мире фондовые коллекции Зоологического института РАН насчитывают более 60 млн единиц хранения, представляющих почти все биологическое разнообразие России, в значительной степени биоразнообразие Антарктики и биоразнообразие отдельных регионов мира.

Огромные масштабы коллекций и рутинные операции по их поддержанию потребовали разработки компьютерных методов ведения зоологических коллекций [16, 17].

В свою очередь, все зоологические и ботанические исследования опираются на строго иерархические системы (классификации) животных и растений. Самым простым и полным представлением таких классификаций являются их графические изображения в виде разветвленных схем (в теории графов именуемых деревьями) (рис. 1). Однако классификации, включающие сотни и тысячи таксонов, изображать таким способом неудобно. Поэтому биологами давно изобретен не имеющий количественных ограничений способ – текстовые систематические списки. В этих списках ранги и соподчинение таксонов отражаются сдвигом строк списка относительно друг друга (табл. 1). Каждому таксономическому рангу соответствует определенный сдвиг строки от левого края страницы (т.е. определенное число пробелов). Подчиненные таксоны идут в списке за таксоном, к которому они относятся (т.е. за родительским таксоном, если использовать терминологию логики). Для удобства чтения перед каждым таксоном иногда указывают его ранг.

С появлением компьютеров систематические списки стали составлять с помощью текстовых редакторов (а позднее – текстовых процессоров типа MS Word). И если такие типичные операции по ведению систематических списков, как перестановки, вставки и удаление таксонов, в текстовом редакторе выполняются просто, то соблюдение одинакового числа пробелов перед таксонами одного ранга в длинном списке выдержать уже трудно. Кроме того, текстовые редакторы практически исключают: получение из списка разнообразной статистики, автоматическую генерацию из большого списка подписков по каким-либо аспектам, логическую проверку содержания списка (в том числе логической целостности классификации и отсутствия пропусков обязательных рангов, диктуемых соответствующим кодексом; в зоологии это: тип, класс, отряд, семейство, род и вид) и использование словарей для единообразного написания повторяющихся слов (в первую очередь – это фамилии авторов таксонов).

Труды 10-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2008, Дубна, Россия, 2008.

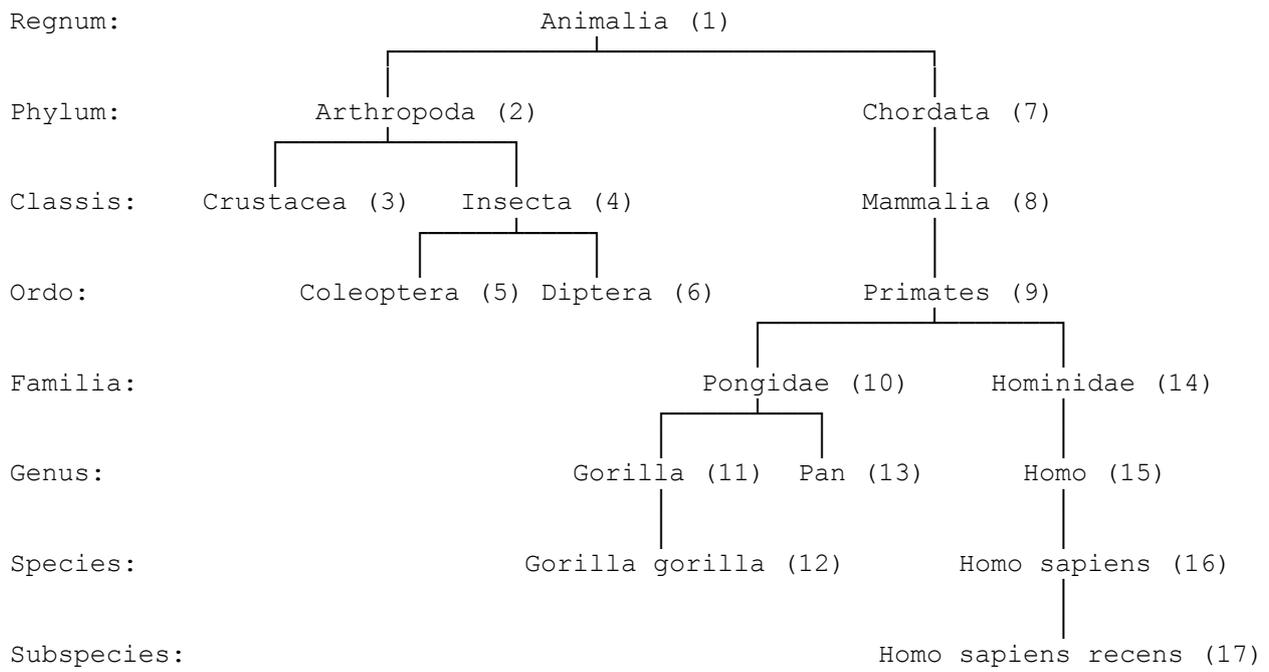


Рис. 1. Фрагмент древа зоологической классификации (номера в скобках показывают порядок обхода при левосписковом описании древа (см. в тексте далее) и при создании систематического списка)

Таблица 1
Фрагмент систематического списка, соответствующий древу классификации на рис. 1

- Regnum: Animalia
- Phylum: Arthropoda
 - Classis: Crustacea
 - Classis: Insecta
 - Ordo: Coleoptera
 - Ordo: Diptera
- Phylum: Chordata
 - Classis: Mammalia
 - Ordo: Primates
 - Familia: Pongidae
 - Genus: Gorilla
 - Species: Gorilla gorilla
 - Genus: Pan
 - Familia: Hominidae
 - Genus: Homo
 - Species: Homo sapiens
 - Subspecies: Homo sapiens recens

Поэтому следующим этапом использования компьютеров для составления и дальнейшего ведения систематических списков стали попытки применения более совершенных программ - табличных процессоров (типа Excel) и систем управления базами данных (СУБД) самых разных моделей. Наиболее распространены три модели СУБД: иерархические, сетевые и реляционные.

Иерархическая модель очень подходит по своим особенностям для работы с иерархическими систематическими списками, но программы для ее поддержания являются экзотикой и отсутствуют на массовом компьютерном рынке. Сетевая модель

прошла небольшой период популярности, но из-за ряда присущих ей недостатков она была вытеснена с коммерческого рынка реляционной моделью [1]. В настоящее время практически все широко используемые и активно развиваемые производителями СУБД опираются на реляционную модель, которая помимо многих положительных качеств имеет под собой строгую математическую теорию и обоснованные ею правила нормализации баз данных, которые позволяют пользователям избегать ошибок, ввода излишней повторяющейся информации и дают возможность облегчить впоследствии работу по созданию программ, реализующих запросы к накопленной информации [3].

В итоге перед нами встала задача разработки способа, позволяющего легко и полно отражать в плоских реляционных базах данных сложные многоуровневые иерархические схемы классификации таксонов.

В основу первых решений этой задачи было положено моделирование в плоской таблице традиционного систематического списка [4, 7]. Такие списки существуют несколько сот лет, но не потеряли своего значения и сейчас. Они официально признаны и информатикой, где именуется «левосписковыми описаниями иерархического древа» [2]. Оказалось, что полную информацию об иерархии таксонов (их названия хранятся в поле LATNAM, куда заносится всегда только одно слово: для биноменов и тринименов – лишь последнее) можно передать всего двумя полями таблицы – цифровым кодом таксономического ранга (RANCOD) и систематическим кодом было (SYSCOD), который описывает порядок таксонов в списке (он может быть цифровым или символьным). Для возможности сравнения разных таблиц необходимо, чтобы RANCOD имел в

них одинаковый смысл. Первым шагом к стандарту было составление таблицы, включающей практически все применяющиеся в зоологии ранги (их оказалось более 40) (табл. 2).

Таблица 2

Стандартные зоологические таксономические категории (ранги) и их цифровые коды (значения для поля RANCOD)

Код ранга	Русское название	Латинское название
1	Царство	Regnum
2	Подцарство	Subregnum
4	Надотдел	Superdivisio
5	Отдел	Divisio
6	Подотдел	Subdivisio
8	Надтип	Superphylum
10	Тип	Phylum
12	Подтип	Subphylum
13	Аппендикс	Appendix
14	Инфратип	Infraphylum
18	Надкласс	Superclassis
19	Раздел	Part
20	Класс	Classis
24	Подкласс	Subclassis
26	Инфракласс	Infraclassis
30	Легион	Legio
32	Подлегион	Sublegio
36	Когорта	Cohors
37	Подкогорта	Subcohors
38	Надотряд	Superordo
40	Отряд	Ordo
42	Подотряд	Subordo
43	Инфраотряд	Infraordo
46	Серия семейств	Series
47	Кладус	Cladus
48	Надсемейство	Superfamilia
49	Группа семейств	Famil. grex
50	Семейство	Familia
52	Подсемейство	Subfamilia
58	Надтриба	Supertribus
60	Триба	Tribus
62	Подтриба	Subtribus
68	Надрод	Supergenus
70	Род	Genus
72	Подрод	Subgenus
80	Секция	Sectio
82	Подсекция	Subsectio
84	Группа видов	Spec. grex
88	Надвид	Superspecies
90	Вид	Species
92	Группа подвидов	Subsp. grex
93	Гибрид	Hybridus
94	Подвид	Subspecies
95	Племя	Natio
96	Вариетет	Varietas
97	Форма	Forma
98	Морфа	Morpha
99	Аберрация	Aberratio

Впервые новый подход был использован в работе над кадастром животных СССР [13]. Таксономические классификаторы – специализированные данные таксономические базы данных, получили широкое применение, и новый подход постоянно совершенствовался.

Принципиально важным было добавление поля SYN, в котором помечались синонимичные таксоны (они приобретали одинаковое значение SYSCOD со своими валидными таксонами), и поля ABBR, которое на первых порах играло роль уникального кода таксона (в нем программно генерировалась уникальная буквенная аббревиатура, имевшая максимальную мнемоничность). В основу концепции, получившей название «стандарт ZOOCOD», легли описание типичной структуры таксономической таблицы и правила присвоения кодов и порядка заполнения таблиц (табл. 3). Под этим названием было опубликовано первое подробное описание концепции и примеры ее применения в систематике млекопитающих [6].

2 Развитие концепции ZOOCOD

Использование концепции ZOOCOD и ее дальнейшее развитие проходило в дальнейшем чаще всего при исследованиях по систематике беспозвоночных – группы, отличающейся большим числом видов и надвидовых таксонов [5, 8, 10, 20, 21] и др. На основе концепции ZOOCOD были проведены важные обобщающие работы ЗИН РАН по биоразнообразию [14, 15] и созданы функционирующие много лет информационные системы ZOOINT [22] и OCEAN [23]. Следующим важным шагом в совершенствовании типовой структуры таблицы явилось введение поля GENUS, в котором специальной программой собирались все начальные части полного названия таксона видового (или еще более низкого) ранга. Соединение содержания полей GENUS+LATNAM образует полное название таксона. Для валидных таксонов это поле избыточно (сборку полного названия можно сделать поднимаясь в таблице вверх по полю LATNAM). Но для синонимичных комбинаций (отмеченных в поле SYN) только в поле GENUS можно было найти названия родов, в которых ранее фигурировал вид или подвид.

Естественным было появление поля ALTGEN, отметка в котором означала первичное описание вида в другом роде и необходимость заключения в круглые скобки фамилии автора таксона и года описания (в поля AUTHOR и YEAR скобки никогда не заносились). Со временем претерпело изменения поле SYSCOD, которое тоже стало частично отражать иерархию, но пока только для обязательных таксономических уровней. Усовершенствованный стандарт получил название ZOOCOD2.

Фрагмент таблицы в стандарте ZOOCOD, соответствующий систематическому списку в таблице 1

GENUS	LATNAM	SYN	RANCOD	ABBR	SYSCOD
	Animalia		1	AN	100
	Arthropoda		10	AR	110
	Crustacea		20	CR	120
	Insecta		20	IN	130
	Coleoptera		40	INCO	13010
	Diptera		40	INDI	13013
	Chordata		10	CH	140
	Mammalia		20	MA	150
	Primates		40	MAPR	15010
	Pongidae		50	MAPRPO	15010100
	Hylobatidae	=	50	MAPRHY	15010100
	Gorilla		70	MAPRPOGOR	150101001000
Gorilla	gorilla		90	MAPRPOGORGOR	1501010010001000
	Pan		70	MAPRPOPAN	150101001010
	Hominidae		50	MAPRHO	15010105
	Homo		70	MAPRHOHOM	150101051000
Homo	sapiens		90	MAPRHOHOMSAP	1501010510001000
Homo sapiens	recens		94	MAPRHOHOMSAPRE	150101051000100010

В последние годы, в связи со спецификой разработок, использующих стандартные управляющие элементы и диалоговые панели Windows, концепция ZOOCOD претерпела значительные изменения и появилась 3-я версия – ZOOCOD3 (табл. 4). К описанному ранее набору полей теперь добавлены поля UNICOD (уникальный код таксона, делающий ненужным поле ABBR) и UPPCOD (значение UNICOD для ближайшего родительского таксона; для синонима таким является его валидный таксон). Эти поля позволяют пользоваться TreeView – специфическим

элементом Windows, изображающим иерархию таксонов в виде дерева. Хотя эти два поля полностью отражают многоуровневую иерархию, они неудобны для быстрого поиска потомков одного таксона. Для решения этой проблемы изменены правила заполнения поля SYSCOD (в новом виде оно называется HIERCOD). Его значениям приданы строгая и полная иерархичность, ширина поля делается равной максимальному числу уровней иерархии в таблице, а для заполнения используются не только цифры, но и все идущие за ними символы.

Таблица 4

Полный (максимальный) перечень полей таксономической таблицы стандарта ZOOCOD3

Имя поля	Тип	Длина	Назначение
GENUS	C	60	недостающие до полного названия вышестоящие термины (род – для вида, род и вид – для подвида, род, вид и подвид – для разновидности или формы)
LATNAM	C	25	латинское название (одно слово для данного ранга)
AUTHOR	C	45	автор таксона
YEAR	C	4	год описания таксона
RANCOD	N	2	код ранга таксона
SYN	C	1	код синонима
TAXON	C	85	полное название таксона (без автора и года описания)
ALTGEN	C	1	отметка описания в другом роде
ALTGENUS	C	25	название рода, в котором вид был впервые описан
SYSCOD	C	15	систематический код (необязателен при наличии HIERCOD)
HIERCOD	C	32	полностью иерархический систематический код (длина поля изменяется в зависимости от числа таксономических рангов в классификаторе)
UNICOD	C	10	уникальный код названия (генерируется автоматически)
UPPCOD	C	10	значение UNICOD у вышестоящего таксона
RANK	C	15	словесное обозначение ранга таксона (дублирование RANCOD для неподготовленного пользователя)
ABBR	C	14	оставлено исторически, ибо во многих старых таблицах играет роль UNICOD

Поле HIERCOD заполняется специальной утилитой, но зато поиск потомков заданного таксона выполняется молниеносно и предельно просто – наложением фильтра с полностью оптимизируемым условием. Интересно, что мы в своих разработках самостоятельно пришли ко всем способам представления иерархии в реляционных таблицах, известным в информатике [19, 24]. К полям GENUS и LATNAM добавлено поле TAXON, в котором хранится полное название таксона (т.е. биномен для видов и триномен для подвидовых таксонов). Для возможности отказа от использования поля GENUS введено более короткое поле ALTGENUS, в которое заносится названия рода, в котором был впервые описан вид. По техническим причинам поле YEAR переименовано в YEARD.

Этот стандарт многократно описан в публикациях [9] и сейчас распространен и за пределами ЗИИ (БИН, ИПЭЭ, МГУ, МГПИ, Нижегородский госуниверситет и др.). На базе этого стандарта была создана основа национальной информационной системы по биоразнообразию России [12].

Концепция представления многоуровневых иерархий в реляционных базах данных перенесена с таксономии на географические данные, сведения по морфологии насекомых и другие аспекты зоологической информации [11].

3 Правила унификации биологических баз данных

В итоге работ над концепцией ZOOCOD были сформулированы некоторые основные положения построения и унификации биологических (таксономических) баз данных, которые должны облегчить общение биологов и создание интегрированных систем:

1. Должна использоваться реляционная модель баз данных.

2. В соответствии с требованиями нормализации баз данных информационные системы должны включать отдельные таблицы по каждому аспекту сведений о животных или растениях.

3. Каждый элемент сведений должен храниться в конкретной системе только в одной записи одной таблицы. Это правило влечет за собой широкое использование словарей и классификаторов (или тезаурусов – словарей с иерархической структурой и родо-видовыми отношениями между понятиями).

4. В основе системы должны лежать классификаторы животных, растений, грибов, бактерий и других царств живых организмов.

5. Часто используемые биологами данные о географическом распространении организмов должны быть строго разделены на отдельные аспекты (административно-территориальный, физико-географический, биогеографический, ландшафтный, природно-зональный и т.п.). Термины каждого аспекта должны быть зафиксированы в классификаторах.

6. Добиться ведения разных баз данных в одной СУБД невозможно даже внутри одного института и

к этому не нужно стремиться. Важно только использовать профессиональные коммерческие СУБД, имеющие проверенные механизмы импорта и экспорта. В качестве критерия пригодности СУБД предлагается использовать формат DBF (dBaseIII). Если СУБД не способна импортировать или экспортировать этот формат, то ее не следует использовать.

7. Для обеспечения простоты обмена данными между разными СУБД, работающими в разных версиях операционных систем, следует ограничить длину имен файлов 8 символами, длин расширений имен 3 символами, а названия полей баз данных 10 символами; при этом использовать во всех этих именах только латинские буквы, цифры и символ подчеркивания (при этом не использовать разницу между заглавными и строчными буквами, если какая-то система такую возможность поддерживает). В системах, где длина символьного поля не ограничивается или может быть очень большой, не делать длину поля более 254 символов.

8. Если для представления иерархии не используется стандарт ZOOCOD, то все же нужно стремиться к соблюдению некоторых правил, облегчающих переход к этому стандарту:

а) для указания ранга таксона или понятия использовать только одно поле таблицы, приводя в нем коды или полные названия рангов;

б) иметь поле с уникальным кодом таксона и использовать эти коды для связи с таблицами фактов о таксонах;

в) в отдельном поле отмечать факт синонимии, указывая уникальный код валидного названия для каждого синонимичного;

г) иметь поле для указания систематического порядка записей, ибо по правилам нормализации толкование таблицы не должно зависеть от физического порядка записей (упорядочение по этому полю должно создавать аналог «систематического списка», который привычен и необходим большинству биологов);

д) наличие двух полей, отражающих иерархию способом, принятым теперь в приложениях для Windows (поля для отражения иерархии методом TreeView), ни в коей мере не отменяет необходимость пунктов а) и г), так как эти два поля не отражают истинные ранги таксонов и нужный порядок среди дочерних таксонов одного родительского.

4 Концепция ZOOCOD 3 и ее развитие

Стандарт ZOOCOD3 продолжает эволюционировать, приспособиваясь к новым требованиям специалистов-систематиков и изменениям в компьютерных технологиях.

Легко заметить, что перечисленный набор полей (табл. 4) обладает значительной избыточностью и формирование окончательной структуры таблицы – дело вкуса составителя (поле GENUS лишнее при наличии поля TAXON; поля UNICOD и UPPCOD могут быть сгенерированы по значениям RANCOD

и HIERCOD; и т.д.). Такая избыточность оставлена намеренно, ибо окончательная структура должна выбираться с учетом предпочтений программиста, который будет писать программы работы с базой данных, обеспечивающие изменения классификации, дополнения, выполнение типовых запросов, генерацию вторичных списков и таблиц, страниц HTML и ASP, проведение различных проверок и анализа по разным аспектам.

Особенно важным нам кажется резюмировать решения, связанные с главной задачей - представлением иерархии. Обобщая сказанное выше можно выделить следующие способы представления сложной иерархии в плоских реляционных таблицах стандарта ZOOCOD3:

1. Использование совокупности полей: SYSCOD + RANCOD;

2. Использование совокупности полей: HIERCOD + RANCOD;

3. Использование совокупности полей: UNICOD + UPPCOD + RANCOD + SYSCOD.

Эти способы существенно отличаются при написании программ обработки таксономических таблиц.

Все изложенное выше относится к первым 20-ти годам эволюции концепции ZOOCOD, проходившей в основном с использованием СУБД FoxPro и Access. Но используемый FoxPro формат DBF все хуже и хуже воспринимается современными программами, работающими под управлением Windows. Кроме того, стандарт ZOOCOD3 выполняет не все требования нормализации. Поэтому в 2007–2008 гг. была предпринята еще одна серьезная ревизия стандарта: отказ от полей, избыточных в новой модели данных (ABBR, GENUS, HIERCOD); использование строгой реляционной модели.

Литература

- [1] Алимов А.Ф., Лобанов А.Л., Пугачев О.Н. Сравнительный анализ реляционного и сетевого подходов к созданию банков данных по систематике, экологии и географическому распространению животных // Журнал общей биологии. – 1993, т. 54, № 1. – С. 96–103.
- [2] Глушков В.М. Основы безбумажной информатики. – М. : Наука. 1982. – 552 с.
- [3] Дейт К. Дж. Введение в системы баз данных : пер. с англ. – 6-е издание. – Киев – М. : Диалектика, 1998. – 784 с.
- [4] Лобанов А.Л. Линейно-иерархическая структура баз данных о таксонах животных // Принципы и методы экоинформатики. – М., 1986. – С. 293–295.
- [5] Лобанов А.Л., Дианов М.Б., Смирнов И.С. Результаты разработок и использования зоологических информационно-поисковых систем // Информационно-поисковые системы в зоологии и ботанике : Труды Зоологического ин-та РАН. – Т. 278. – СПб., 1999. – С. 54–55.
- [6] Лобанов А.Л., Зайцев М.В. Создание компьютерных баз данных по систематике млекопитающих на основе классификатора названий

животных «ZOOCOD» // Вопросы систематики, фаунистики и палеонтологии мелких млекопитающих : Труды Зоологического ин-та РАН. – Т. 243. – СПб., 1991. – С. 180–198.

- [7] Лобанов А.Л., Сергеев Г.Е. Проект классификатора названий животных и принцип представления информации об распространении в структуре биологических баз данных // Принципы и методы экоинформатики. – М., 1986. – С. 214–215.
- [8] Лобанов А.Л., Смирнов И.С. Принципы построения и использования классификаторов животных в стандарте ZOOCOD // Базы данных и компьютерная графика в зоологических исследованиях : Труды Зоологического ин-та РАН. – Т. 269. – СПб., 1997. – С. 66–75.
- [9] Лобанов А.Л., Смирнов И.С. Место и роль информационных технологий в исследованиях Зоологического института РАН // Фундаментальные зоологические исследования. Теория и методы. – М. – СПб. : Изд-во КМК, 2004. – С. 283–318.
- [10] Лобанов А.Л., Смирнов И.С., Дианов М.Б. ZOOCOD – концепция представления зоологических иерархических классификаций в реляционных базах данных // Информационно-поисковые системы в зоологии и ботанике (Тезисы международного симпозиума, май 1999) : Труды Зоологического ин-та РАН. – 1999, Vol. 278. – Р. 66.
- [11] Медведев С.Г., Лобанов А.Л. Информационно-аналитическая система по мировой фауне блох (Siphonaptera): результаты и перспективы развития // Энтомологическое обозрение. – 1999, 78(3). – С. 732–748.
- [12] Пугачев О.Н., Дианов М.Б., Лобанов А.Л., Смирнов И.С., Халиков Р.Г., Голиков А.А. Итоги разработки проекта «Информационная система по биоразнообразию животных России» (ZooDiv) // Отчетная научная сессия по итогам работ 2007 г., 8–10 апреля 2008 : Тезисы докладов. – СПб. : Изд-во Зоологического ин-та, 2008. – С. 42–44.
- [13] Скарлато О.А., Алимов А.Ф., Лобанов А.Л., Умнов А.А. Машинные банки данных – подход к кадастру животного мира // Всесоюзное совещание по проблеме кадастра и учета животного мира : Научно-информац. материалы к совещанию. – Уфа, 1989. – С. 56–64.
- [14] Скарлато О.А., Старобогатов Я.И., Лобанов А.Л., Смирнов И.С. Базы данных по зоологической систематике и сведения о высших таксонах животных // Зоологический журн. – 1994а, Т. 73, Вып. 12. – С. 100–116.
- [15] Скарлато О.А., Старобогатов Я.И., Лобанов А.Л., Смирнов И.С. Биоразнообразие и возможности его анализа с применением компьютерных банков данных // Биоразнообразие. Степень таксономической изученности. – М. : Наука, 1994б. – С. 20–41.
- [16] Смирнов И.С., А.Л. Лобанов, А.Ф. Алимов, В.А. Кривоухатский. Электронные коллекции

Зоологического института РАН // Электронные библиотеки: перспективные методы и технологии, электронные коллекции : Труды Пятой Всероссийской научной конференции RCDL'2003 (Санкт-Петербург, 29–31 октября 2003 г.). – СПб. : НИИ Химии СПбГУ, 2003. – С. 275–278.

- [17] Смирнов И.С., А.Л. Лобанов, А.Ф. Алимов, А.Г. Кирейчук, А.Е. Вахитов. Электронные зоологические коллекции и технологии электронных библиотек // Электронные библиотеки: перспективные методы и технологии, электронные коллекции : Сб. тезисов постерных докладов Восьмой Всероссийской научной конференции (RCDL'2006). Суздаль, 17–19 октября 2006 г. – Ярославль : Ярославский государственный ун-т им. П.Г. Демидова, 2006. – С. 60–61.
- [18] Старобогатов Я.И. О проблемах номенклатуры высших таксономических категорий // Справ. по систематике ископаемых организмов. – М. : Наука, 1984. – С. 174–187.
- [19] Щеваев П.А. Способы хранения иерархических структур в реляционных базах данных // Новые информационные технологии и системы : Труды VI Междунар. научно-технической конф. Ч. 2. – Пенза : ПГУ, 2004. – С. 226–233.
- [20] Lobanov A.L., Smirnov I.S., Dianov M.B. ZOOCOD – conception of representation of zoological hierarchical classifications in relational databases // Information Retrieval Systems in Biodiversity Research (Abstracts of the International Symposium. May 1999). Proceedings of the Zoological Institute RAS. 1999. Vol. 278. P. 65.
- [21] Smirnov I. 1993. Working out a databank on marine invertebrates. – In: Abstracts of 8th International Echinoderm Conference. Dijon, September 6–10, 1993. P. 111.
- [22] Smirnov I.S., Lobanov A.L., Alimov A.F., Dianov M.B., Golikov A.A., Stepanjants S.D. Integrated Information System ZOOINT: the Present State and Prospects // Proceedings of the First East-European Symposium on Advances in Databases and Information Systems (ADBIS'97). St.-Petersburg, September 2-5, 1997. V. 2, 1997. P. 112–114.
- [23] Smirnov I., Lobanov A., Dianov M. 1994. Information system for Antarctic marine animals. - In: Abstracts of SCAR SIXTH BIOLOGY SYMPOSIUM: Antarctic Communities: Species, Structure and Survival. Venice, 30 May – 3 June, 1994. P. 247.
- [24] VanTulder Gijs, 2003. Storing hierarchical data in a DB: <http://www.sitepoint.com/print/hierarchical-data-database>

Evolution of the standard ZOOCOD – concept of reflection of zoological hierarchical classifications in the plane tables of relation databases

A.L. Lobanov, I.S. Smirnov, M.B. Dianov,
A.A. Golikov, R.G. Khalikov

Biological databases lean on taxonomic tables in which complex multilevel hierarchical schemes of taxa classification should be reflected full. Reflection of hierarchies in flat tables of relation databases is not trivial problem. Some ways of the decision of this problem are known in the theory of intelligence systems. In Zoological Institute of the Russian Academy of Science (ZIN) the majority of these ways has been developed independently during long-term works on creation of zoological databases. All these development are reflected in standard ZOOCOD repeatedly described in a number of publications. Already more 20 years it successfully is used and widespread outside of ZIN (BIN of the RAS, IPEE of the RAS, the Moscow State University, MSPI, the Nizhniy Novgorod state university, etc.). On the basis of this standard the national intelligence system on a biodiversity of Russia (BIODIV) was created and developed. The concept of representation of multilevel hierarchies in relation databases is successfully transferred from taxonomy on geographical data, data on morphology and ecologies of animals.

* Гранты РФФИ № 05-07-90179-в, 05-07-90354-в и 06-04-08020, проект «Антарктика» (ФЦП «Мировой океан»), программа «Биоразнообразию».