

УДК 597/599:004

ИНФОРМАЦИОННАЯ СИСТЕМА ДЛЯ ИЗУЧЕНИЯ ПОЗВОНОЧНЫХ: ОПЫТ СОЗДАНИЯ И ИСПОЛЬЗОВАНИЯ

А. В. Коросов¹, С. В. Бугмырин^{2*}, Т. Л. Бурдова³, М. Л. Киреева¹,
С. А. Лапина¹

¹ Петрозаводский государственный университет (пр. Ленина, 33, Петрозаводск, Республика Карелия, Россия, 185910)

² Институт биологии КарНЦ РАН, ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910), *sbugmyr@mail.ru

³ Министерство образования и спорта Республики Карелия (просп. А. Невского, 57, Петрозаводск, Республика Карелия, Россия, 185030)

Рассмотрены структура, методы пополнения и использования информационной системы, созданной для накопления данных по многолетним учетам мелких млекопитающих в окрестностях Гомсельского биологического стационара Института биологии КарНЦ РАН. Общий объем данных, полученных с 1994 по 2022 г., составляет 5404 записи по животным, 668 записей по линиям отловов. Реляционная база данных по мелким млекопитающим связана с ГИС, в которой отображается пространственное распределение местообитаний животных и расположение учетных линий. Информационная система позволяет хранить разнородные данные в отдельных связанных таблицах, обеспечивающих наглядность и простоту проверки исходных материалов. Для обмена данными между средами разных программ используется простой формат csv. Таблицы базы и ГИС связаны между собой через составное ключевое поле, которое обеспечивает надежную идентификацию объектов (особей) с уникальными пробами (линиями). Для получения определенной выборки строится многоступенчатый запрос, сочетающий простоту редактирования и возможность контроля на каждом этапе отбора и обобщения данных. Предлагаемая структура информационной системы имеет универсальный характер и может использоваться в любых случаях, когда серии образцов отбираются одновременно в отдельной локации. Система может быть использована в средах программ MS Excel, Access, LO Calc, Base, R и QGIS. Обсуждаются формы представления в таблицах данных, связи между таблицами и слоями ГИС; представлены структура и типы запросов. Рассмотрены примеры запросов, ориентированные на извлечение данных для последующей обработки количественными методами и для иллюстраций.

Ключевые слова: база данных; составной ключ; ГИС; мелкие млекопитающие

Для цитирования: Коросов А. В., Бугмырин С. В., Бурдова Т. Л., Киреева М. Л., Лапина С. А. Информационная система для изучения позвоночных: опыт создания и использования // Труды Карельского научного центра РАН. 2022. № 8. С. 123–133. doi: 10.17076/eco1614

Финансирование. Исследования выполнены в рамках темы НИР КарНЦ РАН (№ г. р. 122032100130-3).

**A. V. Korosov¹, S. V. Bugmyrin^{2*}, T. L. Burdova³, M. L. Kireeva¹, S. A. Lapina¹.
INFORMATION SYSTEM FOR STUDYING VERTEBRATES: AN EXAMPLE OF
DEVELOPMENT AND APPLICATION**

¹ *Petrozavodsk State University (33 Lenin Ave., 185910 Petrozavodsk, Karelia, Russia)*

² *Institute of Biology, Karelian Research Centre, Russian Academy of Sciences
(11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia), *sbugmyr@mail.ru*

³ *Ministry of Education and Sport, Republic of Karelia (57 Al. Nevskogo Ave., 185030
Petrozavodsk, Karelia, Russia)*

The article describes the structure and methods of filling and using the information system created to accumulate data from long-term surveys of small mammals at the Gomselga Biological Research Station of the Institute of Biology KarRC RAS. The data obtained in the period from 1994 to 2022 total 5404 records on individual animals and 668 records on sampling lines. The relational database on small mammals is linked to the GIS which maps of the spatial distribution of animal habitats and the locations of sampling lines. The information system permits storing different types of data in separate interlinked tables which make the source material more visual and easily verifiable. Data are shared between different program environments using the simple csv format. The database and GIS tables are linked via a composite key field facilitating accurate matching of objects (individuals) to unique samples (lines). To acquire a certain subset the user makes a multistage query which is both easily editable and enables control at each stage of data selection and integration. The proposed structure of the information system is universal in nature and applicable whenever series of samples are collected simultaneously in a given locality. The system can be used in the MS Excel, Access, LO Calc, Base, R, and QGIS program environments. The formats of data presentation in tables and links between tables and GIS layers are discussed; the structure and types of queries are presented. Sample queries for acquiring data for subsequent processing by quantitative methods and for illustrations are provided.

Keywords: database; composite key; GIS; small mammals

For citation: Korosov A. V., Bugmyrin S. V., Burdova T. L., Kireeva M. L., Lapina S. A. Information system for studying vertebrates: an example of development and application. *Trudy Kareli'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS*. 2022. No. 8. P. 123–133. doi: 10.17076/eco1614

Funding. The study was carried out within a KarRC RAS research theme (state registration no. 122032100130-3).

Введение

Основная биологическая цель организации работ на изучаемой территории состояла в исследовании популяций мелких млекопитающих как прокормителей личинок и нимф иксодовых клещей, переносчиков возбудителей опасных трансмиссивных заболеваний человека и животных. В последние годы лесные массивы района работ интенсивно вырубались, и к настоящему времени ландшафт представлен мозаикой разновозрастных насаждений, очень благоприятной средой для распространения как микромаммалий, так и иксодовых клещей. Для фиксации результатов учетов животных и их местообитаний информационная система должна включать средства манипуляции данными (база данных) и их пространственного представления (ГИС).

Эффективная информационная система должна обеспечивать простой ввод, хранение и извлечение информации, причем в форме, нужной для непосредственного анализа – как визуального (ГИС), так и с помощью количественных методов [Костина и др., 2003]. Кроме того, система должна быть достаточно простой, чтобы работать с ней могли все занятые в работе специалисты, а не только единственный ответственный за ее обслуживание. Наша база данных, включающая в себя материалы по мелким млекопитающим и их местообитаниям, создавалась именно с этих позиций.

Материалы и методы

Изучение населения мелких млекопитающих силами лаборатории паразитологии Института биологии ИБ КарНЦ РАН началось в

1994 г. в окрестностях биологического стационара «Гомсельга», д. Малая Гомсельга (N62.07°, E33.95°). Обследуется территория площадью 3×3 км. Создание информационной системы в содружестве с кафедрой зоологии ПетрГУ началось в 1999 г., последнее пополнение базы выполнено в марте 2022 г. Собранные информация первично вводится в файлы типа **.xls**. Общий объем данных, собранный с 1994 по 2022 г., составляет 5404 записи по животным, 668 записей по линиям. Местообитания животных в районе работ изучались на территории 3×3 км. Значительная доля района работ занята водоемами (10 % площади). Малопродуктивные для мелких млекопитающих верховые сфагновые болота занимают около 4 % от общей площади, луга – 5 %. Остальная территория занята лесами (81 %). В послевоенный период была покрыта спелыми хвойными и смешанными лесами. После 1950 г. к настоящему времени эти леса вырублены на 80 % и заменились разновозрастными древостоями. Около 15 % территории все еще занимают спелые леса. Распространение разновозрастных массивов представлено в 20 слоях ГИС, последнее обновление выполнено в 2022 г.

Последовательная обработка новой информации включает использование нескольких программ. Информация по учету зверьков и описанию биотопов из журнала и с карт вводится в таблицы в среде Microsoft Excel или Libre Office Calc в файлы типа xls или csv. В среде MS Access (или LO Base, или R) информация аккумулируется и хранится в файлах типа dbf, mbd, obd, csv и с помощью запросов выводится для анализа. Появление новых вырубок отслеживается по спутниковым снимкам, представленным в открытом доступе в Интернете. Вся картографическая информация хранится и обрабатывается в среде ГИС (ранее в программе MapInfo, теперь – в QGIS) в файлах формата shp. Целевые выборки из базы данных обрабатываются в среде пакетов статистического анализа (PAST, R). Поскольку многие программы плохо «понимают» кириллицу, названия полей в базах данных представлены на латинице.

Результаты

Работу над базой данных можно разделить на три этапа – создание (и оптимизация) структуры, наполнение новыми данными (поддержка) и текущее использование. Реляционная база данных по отловам зверьков связана с ГИС, в которой отображается размещение местообитаний животных и учетных линий.

Структура базы данных

База данных построена из нескольких таблиц, связанных между собой. Главное требование к ним мы определяем как «единообразная уникальность» [Коросов, Горбач, 2021]. «Единообразие» относится к набору полей – каждая запись (строка) должна нести одинаковые характеристики, записанные в полях (столбцах). «Уникальность» относится к записям (строкам) – каждая строка должна быть уникальна, т. е. отличаться от других записей значением хотя бы в одном поле.

Данные отловов представлены в четырех связанных таблицах: **species** (с полными названиями видов), **animals** (с характеристиками зверьков), **traps** (с описанием учетных линий ловушек), **biotops** (с описанием биотопов). В название таблиц **animals** и **traps** включен год последнего обновления (рис. 1).

Вспомогательная таблица **biotops** включает в себя 7 записей с полями *nbiot* (номер биотопа), *abiot* (возраст от и до: 1–2, 3–6, 7–11, 12–17, 18–35, 35–100, 100–200), *nambiot* (название биотопа: grass, bush, grove, young, leaf, mixed, conifer) (см. ниже).

Таблица **species** (18 записей) содержит два поля – точное название вида зверька (*species*) и трехбуквенную аббревиатуру (*spc*), составленную из строчных букв: первой буквы родового названия и двух первых букв видового, например, sar – *Sorex araneus*. Сокращение используется для безошибочного ввода видового названия в среде Excel.

Таблица **traps to 2021** (668 записей) включает описания каждой линии давилок, в том числе *nlin* (номер линии), *year* (год), *mon* (месяц), *season* (сезон), составной ключ *yeseli*, ... день начала и конца отловов, ..., *traday* (число давилко-суток), *biotop* (тип биотопа). Поле *season* дополняет поле *mon*, оно введено для учета работы линий, которые были поставлены в один месяц, а сняты в другой. Для таких линий в поле *season* вносится месяц с большей продолжительностью учетов. Для каждой линии рассчитывается *уникальный* составной ключ: $yeseli = year * 10\ 000 + season * 100 + nlin$. Например, 3 линия 8 сезона 2017 года получит значение: $2017 * 10\ 000 + 8 * 100 + 3 = 20170803$.

Таблица **animals to 2021** (5404 записи) содержит описание отловленных особей разных видов; одна запись несет 35 значений (полей). Основными ключевыми полями для шаблонных запросов по оценкам численности являются *npp* (номер особи), *spc* (вид), *year* (год), *mon* (месяц), *season* (сезон), *nlin* (номер линии), *yeseli* (составной ключ). Для каждой записи

animals to 2021.xlsx - LibreOffice Calc

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	npp	dnum	spc	day	mon	year	season	sex	fer	age	YMeLi	yeseli	nlin
2	1	1	mag	16	6	1994	6	F		9	199406001	199406001	1
3	2	2	clg	16	6	1994	6	M			199406001	199406001	1
4	3	3	mag	17	6	1994	6	M			199406001	199406001	1
5	4	4	clg	17	6	1994	6	F	juv	1	199406001	199406001	1
6	5	5	mag	17	6	1994	6	M			199406001	199406001	1
7	6	6	sbe	17	6	1994	6	M			199406001	199406001	1
8	7	7	clg	17	6	1994	6	F	juv	1	199406001	199406001	1

traps to 2022.xlsx - LibreOffice Calc

	A	B	C	D	E	F	G	H	I	N	O	P
1	Npp	nlin	NlineT	YSeKa	yeseli	season	year	mon	day	Traps	traday	biotop
2	159	1		199400000	199405001	5	1994	6	15	50	450	meadow
3	228	3		199400000	199405003	5	1994	6	15	50	450	mixed
4	160	1		199400000	199406001	6	1994	6	26	50	200	meadow
5	229	3		199400000	199406003	6	1994	6	26	50	200	mixed
6	161	1		199400000	199410001	10	1994	10	5	50	200	meadow
7	230	3		199400000	199410003	10	1994	10	5	50	200	mixed
8	162	1		199500000	199505001	5	1995	6	5	50	200	meadow
9	163	1		199500000	199505001	5	1995	6	11	50	200	meadow

Рис. 1. Заголовки основных таблиц базы данных
 Fig. 1. Headers of the main database tables

рассчитывается составной ключ $yeseli = year * 10\ 000 + season * 100 + nlin$. Одинаковые значения этого ключа несут все особи, попавшие в ловушки данной линии. Ключевыми полями для запросов по выборкам зверьков, кроме указанных, служат поля *sex* (пол), *fer* (зрелость), *age* (возраст). Каждая строка содержит многие другие характеристики особи: масса тела и органов, размеры тела и органов, описание половой системы и пр. (*w*, *lt*, *lc*, *lp*...).

Связи между таблицами выполняются по ключевым полям. Таблицы **species** и **animals** связаны по полю *spc* (подстановка). Таблицы **animals** и **traps** связаны по составному ключу *yeseli*.

Структура ГИС

В ГИС включены пять групп слоев.

Картографические слои (**Карты**) – это растры отсканированных карт (формат geotiff), которые имеют географическую привязку к району исследований – местам постановки линий отлова и контурам биотопов.

Спутниковые слои (**Снимки**) – привязанные к местности растровые снимки (формат geotiff), служащие для дешифрирования и выделения биотопов. Снимки взяты из открытых

источников (<https://earthexplorer.usgs.gov/> и <https://earth.google.com/web/>). Анализ снимков за прошедшие годы позволил отобрать для работы 19 снимков, выполненных с промежутком 2–5 лет (с 1969 по 2021 г.).

Слои с линиями ловушек (**Линии**) – векторные слои линейных объектов (формат shp), на которые нанесены локализации линий ловушек (**lines_1994**, ... **lines_2021**). Такая карта строится в конце полевого сезона. Таблица несет четыре поля: *nlin* (номер линии ловушек), *mon* (месяц), *season* (сезон), *year* (год отлова); *yeseli* (составной ключ) рассчитывается как $yeseli = year * 10\ 000 + season * 100 + nlin$. Аналогичный ключ рассчитывается в таблицах базы данных **traps** и **animals**. По мере проведения новых отловов в ГИС вносятся расположение и характеристики новых линий. Число записей в слое **lines_** совпадает с числом записей в таблице **traps to....** для одних и тех же лет.

Слои биотопических выделов (**Биотопы**) – это векторные слои площадных объектов (формат shp), построенных по контурам выделенных биотопов. В нашем случае в качестве биотопа принят такой тип местности, который надежно идентифицируется по снимку и которому соответствует определенный возраст. Хорошо заметным *болотам* задан возраст 1000 лет,

лугам – 300 лет. Свежая вырубка хорошо идентифицируется на снимке по контурам и получает возраст 0. С каждым годом ее возраст увеличивается, и тип биотопа меняется. Для идентификации биотопов выполнены многочисленные их описания по упрощенной геоботанической схеме [Коросов и др., 2003]. Анализ карт, аэро- и космических снимков (с 1950 по 2021 г.) и полевых описаний показал, что в изучаемом районе существенные по размеру биотопы отличаются только возрастом. Исходная реконструированная карта (для условного 1950 г.) включала всего пять явно различающихся типов биотопов (луг, болото, лиственный лес, смешанный лес, хвойный лес). Сплошные рубки, проводимые в последующие годы, сделали ландшафт мозаичным. На каждой делянке идет типичная для юга Карелии сукцессия, не раз описанная в литературе и обобщенная нами для региона исследований [Гусева и др., 2014]. Это исследование позволило выделенным вырубкам определенного возраста назначить соответствующий ему тип биотопа: 0–3 лет – свежая вырубка (grass), 4–7 – молодая вырубка (bush), 8–11 – заростающая вырубка (grove), 12–18 – молодой лес, жердняк, роща (young), 19–35 – лиственный лес (leaf), 36–100 – смешанный лес (mixed), более 100 лет – хвойный лес (conifer) с преобладанием сосны; чистых ельников нет. Эта информация зафиксирована во вспомогательной таблице **biotops**. Для составления карт на космических снимках за разные годы выявили све-

жие рубки, которые нанесли на 19 отдельных слоев (**cut_xxxx**) для следующих лет: 1960, 1965, 1970, 1973, 1975, 1978, 1984, 1986, 1989, 1993, 2001, 2003, 2005, 2009, 2013, 2015, 2017, 2019, 2021 гг. Для тех же лет были построены 20 слоев биотопов **biotop_1950, ... biotop_2021**. Слой **biotop_1950** включал всего пять типов биотопов. На слои для последующих лет добавлялись области (region) вырубленных лесов (свежие вырубки). Слой **biotop_2021** несет уже 10 типов выделов (рис. 2). Таблица для каждого слоя **biotop_** несет поля: *npp* (номер записи), *abiot* (возраст биотопа), *nambiot* (имя или тип биотопа).

Связи между разными слоями обеспечиваются привязкой к одной системе координат (в нашем случае WGS84 UTM Zone 36N), что обеспечивает возможность пространственных совмещений. Так организована связь между слоями **cut_x**, **biotop_x** и **lines_x**, в котором они стояли. Связь между картой **lines** и таблицей **traps** среды Access задается по составному ключевому полю *yeseli*.

Поддержание ГИС

Обновление ГИС и баз данных осуществляется по окончании полевого сезона и призвано отобразить как изменения, произошедшие в природе, так и итоги полевых работ.

Дополнение слоя **Биотопы** – сложная исследовательская задача, поскольку с течением времени типы биотопов меняются по двум причинам. Во-первых, в регионе постоянно

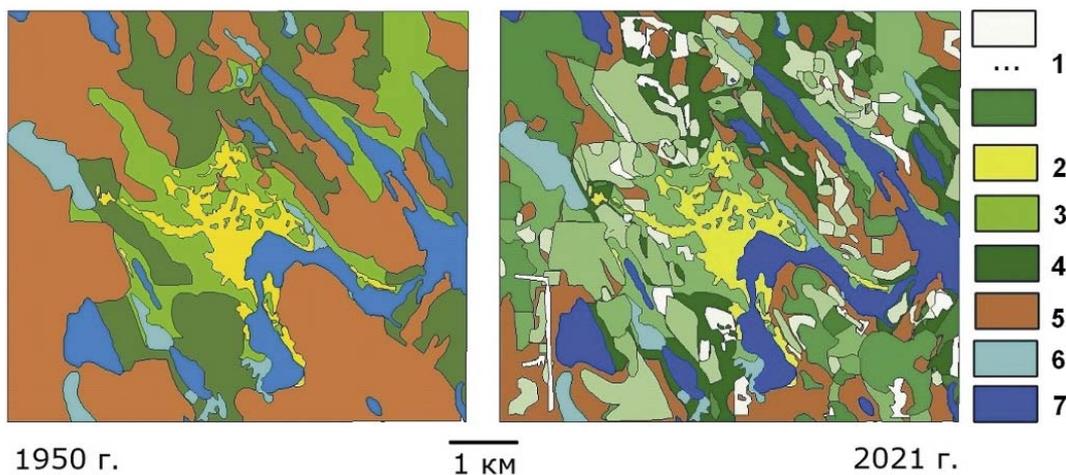


Рис. 2. Карта биотопов для 1950 и 2021 гг.:

1 – вырубки разного возраста; 2 – луга; 3 – лиственные леса; 4 – смешанные леса; 5 – хвойные леса; 6 – болота; 7 – озера

Fig. 2. Map of biotopes for 1950 and 2021:

1 – felling sites of various ages; 2 – meadows; 3 – deciduous forests; 4 – mixed forests; 5 – coniferous forests; 6 – swamps; 7 – lakes

появляются новые рубки, во-вторых, возраст биотопов увеличивается на 1 год. Для учета свежих вырубок выполняются следующие манипуляции (например, для слоя **biotop_1989**) (рис. 3): 1) загрузка нового космического снимка для текущего года, 2) создание нового векторного слоя, отображающего свежие вырубки (с возрастом 0) (**cut_1989**), 3) вырезание из слоя **biotop_1986** слоя **cut_1989** (спелые леса сокращают площадь), 4) добавление в слой **biotop_1989** слоя **cut_1989** (замещение вырубленных выделов спелых лесов выделами рубок), 5) сохранение слоя (как **biotop_1989**). Для пересчета возраста биотопов в обновленной таблице слоя **biotop_1989** с помощью калькулятора полей к возрасту биотопа (*abiot*) прибавляется единица. Затем полученный возраст визуально сверяется со шкалой серии биотопов по возрастам (таблица **biotops**), и при достижении порогового возраста в поле *nambiot* вручную вписывается (копируется из таблицы **biotops**) новое название биотопа, слой **biotop_1989** сохраняется. Процедура выполняется ответственным за обслуживание базы.

Поддержание базы данных

При дополнении основных таблиц новой информацией работа идет с тремя таблицами – старой, вспомогательной и обновленной. Информация о полевых работах (например, за 2021 г.) заносится в две таблицы базы данных формата xls, csv.

Характеристика отловленных зверьков вносится во вспомогательную таблицу **animals2021**; в качестве шаблона для полей используется первая строка общей таблицы **animals**. После

заполнения этой вспомогательной таблицы рассчитываются значения составного поля *yeseli*. Затем все записи переносятся в таблицу **animals to 2020**, которая сохраняется с новым именем **animals to 2021**.

Описание линий ловушек (поставленных, например, в 2021 г.) вносится в таблицу **traps2021**, после заполнения рассчитываются значения составного поля *yeseli*. Каждой новой линии таблицы **traps2021** присваивается тип биотопа, в котором выполнялся отлов. Для этого в среде ГИС открываются два слоя: **biotop_2021** и **lines_2021**. Объекты слоя **biotop_2021** подписываются из поля *nambiot*, объекты слоя **lines_2021** подписываются из поля *nlin*. На карте становится видно, в каком типе биотопа была установлена каждая данная линия. Это название биотопа вручную записывается (копируется из таблицы **biotops**) в поле *nambiot* таблицы **traps2021**. После заполнения таблицы все записи копируются в таблицу **traps to 2020**, которая сохраняется с новым именем **traps to 2021**.

Ввод данных в среде Excel (Calc) не требует специальной подготовки. При возникновении проблемы комментарием можно записать в текущей строке в поле для комментариев, чтобы к ней вернуться во время проверки ответственным за базу.

Если для ведения базы используется программа Access (Base), то блоки новых данных с листа Excel (Calc) копируются в буфер обмена и добавляются к соответствующей таблице в среде Access. Работа выполняется любым сотрудником, но контролируется ответственным за базу. При манипуляции с базой данных в среде R используются только файлы xls (или csv).

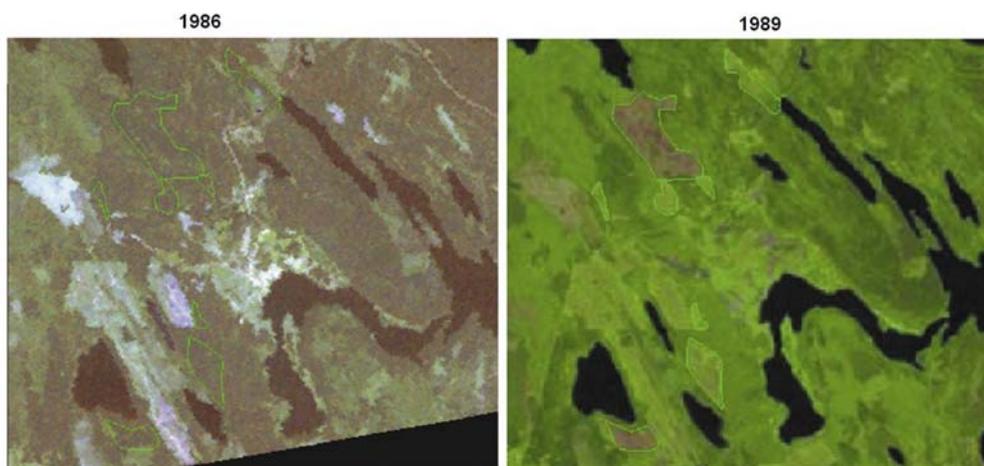


Рис. 3. Выявление новых рубок на снимке 1989 г. при сравнении со снимком 1986 г.
Fig. 3. Identification of new fellings in the 1989 image compared with the 1986 image

Использование базы данных. Запросы

Запрос – это и процедура, и результат ее выполнения в виде новой таблицы (выборки), в которой определенным образом комбинируются данные из исходных таблиц базы данных или других запросов. Запрос (шаблон для процедуры) сначала нужно построить, затем им можно многократно пользоваться (для создания разных выборок). В нашей базе построением шаблонов запросов занимается только ответственный за базу. В нашей практике с помощью запросов строятся выборки морфологических характеристик самих зверьков и выборки по учету численности зверьков. Запросы можно создавать в средах Access и Base. Но для краткости записи даны примеры запросов в среде R.

Запрос на выборку для популяционно-морфологической характеристики – это простые запросы, они оперируют только одной таблицей, **animals**. В запросе указывается имя таблицы, задается список ключевых полей, а также список полей с данными для обработки. После этого вводятся значения – критерии для отбора особей определенного статуса (вид, пол, зрелость, возраст). В результате формируется выборка, подлежащая обработке. В примере – формирование выборки и построение распределения по массе тела лесной мышовки (рис. 4):

```
>mm<-read.csv(«animals to 2021.csv»)
>hist(mm[mm$spc=='sbe',17])
```

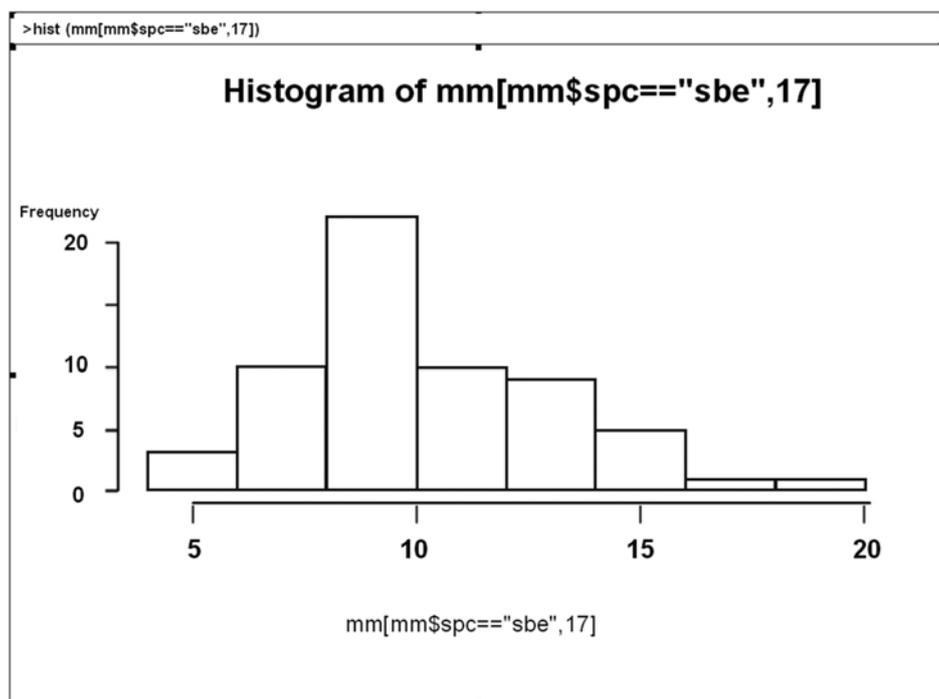


Рис. 4. Распределения значений массы тела лесной мышовки (*Sicista betulina*)
Fig. 4. Distributions of body mass values of the northern birch mouse (*Sicista betulina*)

Запрос на выборку для оценки относительной численности основан на двух таблицах – **animals** и **traps**. Основной смысл информационной системы состоит в том, чтобы рассчитывать значения относительной численности зверьков, то есть подсчитывать число зверьков (N), попавших в определенное число ловушек (n), и найти частное между этими числами (N/n). При этом необходимо сохранить и те учеты, когда в линии не попалось

ни одного зверька, т. е. для которых нет записей в таблице **animals**. Чтобы не допустить такие потери, ведущей таблицей при подсчетах относительной численности является таблица регистрации линий **traps**. С помощью оценок численности требуется описать динамику по годам, сезонам, районам, биотопам, видам и пр.

Общая архитектура построения запроса для определения относительной численно-

сти складывается из 5 ступеней, или этапов, причем независимо от того, в какой среде они выполняются – в Access, Base или R. Для краткости мы рассмотрим пример построения запросов в среде R. Пример запроса на выборку рассмотрен для характеристики *сезонной динамики относительной численности обыкновенной бурозубки* (*spc = 'sar'*) в *хвойных* биотопах (*biotop = 'coniferous'*).

На первом этапе строится *простой запрос* относительно зверьков – из исходной таблицы **animals** отбираются все зверьки определенного статуса – вида, возраста, зрелости. В конце концов именно для этой группы живот-

ных и будут рассчитываться характеристики относительной численности. При построении такого запроса (выборки) сокращается не только количество записей (строк), но и количество полей. Можно оставить только те поля, что понадобятся для последующей обработки и для контроля правильности отбора: *yeseli*, Год (*year*), Сезон (*season*), Биотоп (*biotop*), Номер линии (*nlin*). В одну линию могут попасться несколько зверьков, значит, в первом запросе одинаковые значения ключа *yeseli* могут повторяться. Для обеспечения устойчивости всей работы этот запрос (выборка) сохраняется как отдельная (рабочая) таблица с именем **query01**.

```
>mm<-read.csv(«animals to 2021.csv»)
>head(mm[,1:13],1)
  npp dnum spc day mon year Season sex fer age YMeLi yeseli line
1    1    1 mag  16   6 1994     6   F    9 199406001 199406001  1
>query01<- mm[mm$spc=='sar',]
>head(query01[,1:13],1)
  npp dnum spc day mon year Season sex fer age YMeLi yeseli line
11   11   11 sar  20   6 1994     6   M  ad  1 199406001 199406001  1
```

На втором этапе строится *итоговый запрос*, который подсчитывает число зверьков заданного статуса, попавших в каждую отдельную линию. Ключом для обобщения служит поле *yeseli*, полем для подсчета может быть любое, например, то же ключевое *yeseli*. Запрос несет

новое поле *count*, содержащее число попавшихся в линию зверьков. Каждая запись в этой вторичной выборке стала уникальна (по ключу *yeseli*) и соответствует только одной записи (линии) из таблицы **traps**. Запрос получает имя **query02**.

```
>query02 <- aggregate(query01$yeseli,list(query01$yeseli),length)
>names(query02)<-c('yeseli','count')
>tail(query02,3)
  yeseli count
688 201609001    11
689 201609002     4
690 201609003     4
```

На третьем этапе объединяются данные по зверькам и ловушкам, т. е. создается *простой запрос* с объединением таблицы **traps** и запроса **query02**. При этом к каждой линии таблицы **traps** присоединяется запись из запро-

са **query02**, т. е. число попавшихся зверьков. Поскольку в некоторые линии зверьки не попадались, вместо числа особей к этой записи добавляется значение NULL (или NA). Запрос сохраняется с именем **query03**.

```
>tra<-read.csv(«traps to 2021.csv»)
>head(tra,1)
  Npp nlin yeseli season year mon day yearEnd monEnd dayEnd period Traps traday biotop
1 159   1 199405001     5 1994  6  15  1994     6    24   NA    50  450 meadow
>tr<-tra[,c(3,4,5,13,14)]
>query03<-merge(tr,query02,by='yeseli')
>head(query03,1)
  yeseli season year traday biotop count
1 199406001     6  1994    200 meadow     3
```

На четвертом этапе осуществляется отбор тех характеристик линий, которые нужны для фор-

мирования обобщающей таблицы. Могут потребоваться обобщения по разным годам (меж-

годовая динамика), по сезонам (сезонная динамика), по биотопам (биотопическое размещение) или в разных комбинациях этих критериев. С этой целью на основании **query03** строится

простой запрос с именем **query04**, в который (в режиме конструктора вручную) вводятся критерии отбора. В результате формируется сокращенная таблица отловов в конкретные линии.

```
>query04<-query03[query03$biotop=='conifer',]
>nrow(query04)
[1] 60
```

На пятом этапе подводятся итоги расчетов, то есть суммируется количество зверьков и количество давилко-суток по выделенным критериям, а затем отыскиваются отношения этих значений. Общая формула расчетов такова: $100 * (\text{сумма зверьков за период отлова}) / (\text{сумма давилко-суток, работавших в данном месте в течение всего времени})$, единицы – число особей на 100 давилко-суток. В среде Access

новый запрос **query05** строится в форме *перекрестной таблицы* на основе **query04**. В режиме конструктора в него можно включить разные поля из запроса **query04**. Результатом является обобщенная двупольная таблица, в которой фигурируют оценки относительной численности в данном биотопе в данный период времени для зверьков заданного статуса (например, конкретного вида).

число зверьков:

```
>(comm<-tapply(query04$count,query04$season,sum))
 1  5  6  7  8  9 10
27  3 14 82 173 4 22
```

число давилко-суток:

```
>(cotr<-tapply(query04$traday,query04$season,sum))
 1  5  6  7  8  9 10
506 400 840 2450 2939 150 585
```

численность (число особей на 100 давилко-суток):

```
>(query05<-round(100*comm /cotr ,2))
 1  5  6  7  8  9 10
5.34 0.75 1.67 3.35 5.89 2.67 3.76
```

Расчеты показали, что попадаемость бурозубок в течение теплого сезона возрастает с 0,8 до 6 экз. на 100 давилко-суток, а

осенью падает.

Второй пример – распределение тех же бурозубок по биотопам:

```
> query04<-query03[query03$season==8,]
> (comm<-tapply(query04$count,query04$biotop,sum))
 bog bush conifer grass grove leaf meadow mixed young
 8 84 173 90 93 110 115 117 199
> (cotr<-tapply(query04$traday,query04$biotop,sum))
 bog bush conifer grass grove leaf meadow mixed young
 350 920 2939 717 935 1179 2643 2211 1796
> (query05<-round(100*comm /cotr ,2))
 bog bush conifer grass grove leaf meadow mixed young
 2.29 9.13 5.89 12.55 9.95 9.33 4.35 5.29 11.08
```

Расчеты показали, что попадаемость бурозубок в августе во вторичных биотопах (разнородные вырубки) выше, чем в первичных (болота, луга, леса).

Обсуждение

Остановимся на основных проблемах, которые пришлось решать при построении информационной системы.

Существенным моментом, на наш взгляд, является отказ как от «стандартной нумерации», так и от «сквозной нумерации» линий ловушек («проб», «точек отбора проб»). В начале работ некоторые линии получили «стабильные» номера 1, 2, ... 7. Однако в силу разных обстоятельств линии оказывались разными по размеру – то 25, то 50 давилко-суток, часть линий из-за удаленности стали недоступными, а линии на вырубках через 20 лет оказались в лесу.

Номер линии перестал соответствовать свойствам биотопов. Наши организационные попытки придерживаться однотипной сквозной нумерации тоже потерпели неудачу. В разные годы животных отлавливали от 1 до 4 групп разных исполнителей. И все усилия, потраченные в конце сезона на согласование номеров линий (в хронологическом порядке), становились напрасными, когда появлялись новые данные от другой группы. В конце концов опыт подсказал, что сообщить индивидуальность каждой записи в таблице линий может только уникальный составной ключ, надежно идентифицирующий каждую линию. Поскольку отловы велись не чаще одного раза в месяц, указание в составном ключе (*yeseli*) года, месяца и рабочего номера линии оказалось достаточным, чтобы отличать линии.

Большой проблемой оставалась оценка принадлежности каждой линии к тому или иному типу биотопа. За последние 70 лет в районе работы вырублено 80 % хвойных лесов, на месте которых формируются вторичные леса. За те 30 лет, что здесь проводятся наблюдения за мелкими млекопитающими, многие территории существенно сменили свой статус. При постановке линии каждый раз приходилось заново называть биотоп, что разные исполнители могли делать по-разному. Путаница с назначением типа биотопа закончилась только тогда, когда была принята возрастная шкала для назначения типа биотопа и построены ГИС-слои, идентифицирующие возраст биотопа, позволяющий их однозначно классифицировать. Таким образом, ГИС играет ключевую роль в формировании представлений о пространственном распределении микромаммалий.

Наш опыт подсказывает, что планирование любого экологического исследования должно включать в себя создание и обслуживание информационной системы – базы данных и ГИС. Сразу создать оптимальную структуру вряд ли получится в связи с еще устойчивой традицией ведения записи первичной информации в журналы. Не отрицая необходимость журналов, формат записи в них нужно трансформировать с учетом удобства последующего ввода данных в компьютер.

Созданная нами структура и наполнение информационной системы помогали подготавливать материал для расчетов и публикаций (число в скобках), посвященных особенностям динамики микромаммалий на севере (3), рассматривающих пространственное и биотопическое размещение зверьков (5) и их паразитов (7), анализирующих причины динамики численности отдельных видов (обыкновенной

бурозубки – 4, рыжей полевки – 2), а также рассматривающих явление скейлинга для местных сообществ микромаммалий (1). Опыт сопряжения баз данных с ГИС вошел в учебные пособия по технике ведения ГИС (2) и практике обработки биологических данных (5).

Заключение

Построенная нами информационная система на основе реляционной базы данных и ГИС по результатам учетов мелких млекопитающих позволяет хранить разнородные данные в отдельных связанных таблицах, обеспечивающих наглядность и простоту проверки исходных материалов. Для обмена данными между средами разных программ используется простой формат csv. Прозрачная структура информационной системы позволяет обрабатывать данные в средах разных программ – Access, Base, R. Связи (relations) между таблицами базы и ГИС обеспечиваются через составное ключевое поле *yeseli*, которое обеспечивает надежную идентификацию объектов (особей) с уникальными пробами (линиями). Для получения определенной выборки данных, необходимых для обработки или иллюстрации, строится многоступенчатый запрос, сочетающий простоту редактирования, возможность строгого контроля на каждом этапе отбора и обобщения данных и любую степень сложности решаемой задачи.

Предлагаемая структура информационной системы имеет универсальный характер и может использоваться в любых случаях, когда серии образцов отбираются одновременно в отдельной локации. Построенная нами база данных помогла подготовить и опубликовать 24 статьи в реферируемых изданиях, 5 учебных пособий, защитить 3 кандидатских диссертации.

Литература

Гусева Т. Л., Коросов А. В., Беспятова Л. А., Анисанова В. С. Многолетняя динамика биотопического размещения обыкновенной бурозубки (*Sorex araneus* Linnaeus, 1758) в мозаичных ландшафтах Карелии // Ученые записки Петрозаводского государственного университета. 2014. № 8(145). С. 13–20.

Коросов А. В., Горбач В. В. Практическое введение в среду R. Петрозаводск: ПетрГУ, 2021. 117 с. URL: <https://disk.yandex.ru/i/skOj2DT4UTIWGQ> (дата обращения: 20.05.2022).

Коросов А. В., Матросова Ю. М., Бугмырин С. В., Анисанова В. С., Беспятова Л. А. Опыт реконструкции территориального размещения рыжей полевки в мозаичном ландшафте южной Карелии // Труды

Карельского научного центра РАН. 2003. Вып. 4. Биогеография Карелии (флора и фауна таежных экосистем). С. 204–212.

Костина Н. В., Розенберг Г. С., Шитиков В. К. Экспертная система экологического состояния бассейна крупной реки // Известия Самарского НЦ РАН. 2003. Т. 5, № 2. С. 284–294.

References

Guseva T. L., Korosov A. V., Bespyatova L. A., Anikanova V. S. Long-term dynamics of biotopical distribution of a common shrew (*Sorex araneus* Linnaeus, 1758) in Karelian mosaic landscape. *Proceedings of Petrozavodsk State University*. 2014;8(145):13–20. (In Russ.)

Korosov A. V., Gorbach V. V. A practical introduction to R. Petrozavodsk: PetrSU; 2021. 117 p. URL: <https://disk.yandex.ru/i/skOj2DT4UTIWGQ> (accessed: 20.05.2022). (In Russ.)

Korosov A. V., Matrosova Yu. M., Bugmyrin S. V., Anikanova V. S., Bespyatova L. A. An attempt of the reconstruction of bank vole (*Clithrionomys glareolus*) territorial distribution in mosaic landscapes of Southern Karelia. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS*. 2003;4:204–211. (In Russ.)

Kostina N. V., Rozenberg G. S., Shitikov V. K. Expert system of ecological status of basin of larger river. *Izvestiya Samarskogo NC RAN = Izvestia of Samara Scientific Center RAS*. 2003;5(2):284–294. (In Russ.)

Поступила в редакцию / received: 31.05.2022; принята к публикации / accepted: 02.12.2022.
Авторы заявляют об отсутствии конфликта интересов / The authors declare no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРАХ:

Коросов Андрей Викторович

д-р биол. наук, профессор кафедры зоологии и экологии Института биологии, экологии и агротехнологий

e-mail: korosov@petsu.ru

Бугмырин Сергей Владимирович

канд. биол. наук, заведующий лабораторией паразитологии животных и растений

e-mail: sbugmyr@mail.ru

Бурдова Татьяна Леонидовна

ведущий специалист отдела лицензирования Управления государственного контроля (надзора) в сфере образования

e-mail: burdova@minedu.karelia.ru

Киреева Марина Леонидовна

научный сотрудник лаборатории функциональной зоологии

e-mail: kireeva89@yandex.ru

Лапина Софья Александровна

студентка Института биологии, экологии и агротехнологий

e-mail: akella.keidg@yandex.ru

CONTRIBUTORS:

Korosov, Andrey

Dr. Sci. (Biol.), Professor

Bugmyrin, Sergey

Cand. Sci. (Biol.), Head of Laboratory

Burdova, Tatyana

Leading Specialist

Kireeva, Marina

Researcher

Lapina, Sofya

Student