

BearBase: БАЗА ДАННЫХ МУЗЕЙНОЙ КОЛЛЕКЦИИ ОБРАЗЦОВ БЕЛОГО МЕДВЕДЯ

© 2025 г. И.В. Бездворных*, А.А. Канапин*, А.А. Самсонова*, Е.А. Орлова*, М.В. Саблин**,
А.В. Абрамов**, В.В. Платонов**, Д. Хирата*,[#]

*Санкт-Петербургский политехнический университет Петра Великого,
Политехническая ул., 29, Санкт-Петербург, 195251, Россия

**Зоологический институт РАН, Университетская наб., 1, Санкт-Петербург, 199034, Россия

[#]E-mail: dhirata59@gmail.com

Поступила в редакцию 20.10.2024 г.

После доработки 20.10.2024 г.

Принята к публикации 06.12.2024 г.

Вклад музейных коллекций в современные научные исследования существенно вырос благодаря развитию современных омиксных технологий и методов генетического анализа. Однако, несмотря на переход к цифровым форматам хранения данных, управление ими и менеджмент коллекций по-прежнему сопряжен с трудностями. Существующие программные инструменты разработаны для решения конкретных задач и часто требуют развитых технических навыков, что ограничивает сферу их применимости. Для решения этих проблем мы разработали BearBase – удобную и портативную систему управления базами данных, не требующую продвинутой экспертизы в сфере информационных технологий. Она имеет как серверные, так и клиентские компоненты, поддерживает работу в автономном режиме, а также синхронизацию с мобильными устройствами для работы в полевых условиях. Пользователи могут легко импортировать данные из таких форматов, как MS Excel, и изменять структуру баз данных по мере необходимости. Программное обеспечение с открытым исходным кодом доступно на GitHub. В настоящее время в BearBase хранятся данные примерно о 500 образцах белого медведя (*Ursus maritimus*) из коллекции Зоологического института РАН (Санкт-Петербург), охватывающих временной интервал активных исследований и промышленного освоения арктического региона с середины XIX века по наши дни. Кроме того, система может быть использована для управления коллекциями любых биологических или медицинских образцов.

Ключевые слова: биологические коллекции, геномика, популяционная генетика, арктическая фауна, базы данных.

DOI: 10.31857/S0006302925010195, EDN: LUETCV

Музейные коллекции биологических образцов представляют собой уникальный ресурс, который в сочетании с новыми методиками работы с архивным генетическим материалом позволяет проводить исследования в различных областях, таких как популяционная и эволюционная генетика с учетом исторической перспективы. Документирование образцов в такого рода собраниях обычно велось с использованием традиционных методов, на бумажных носителях и лишь в последние десятилетия начинается переход к цифровой форме представления информации [1–3]. Разнородность информации, необходимой для полноценного описания музейных биологических образцов, также затрудняет разработку практических решений информационного менеджмента таких коллекций. Существующие программы, такие как PhenoBook, GridScore [4, 5] и другие [6, 7], имеют в основном узко специали-

зованный характер и направлены на решение конкретных практических задач, связанных, например, с фенотипированием растительных образцов. Кроме того, практически все из доступных на данный момент платформ требуют высокой профессиональной квалификации для их установки и управления, что существенно затрудняет работу с такими системами.

Разработанная нами система BearBase представляет собой простое, удобное в использовании портативное решение, с которым может работать пользователь без профессионального опыта управления базами данных. Система включает в себя серверную и клиентскую части и может быть легко развернута на локальном узле. Возможность синхронизации с мобильными устройствами для обмена информацией, а также способность работать при отсутствии постоянного

доступа в интернет позволяет осуществлять наполнение базы данных и ее редактирование в полевых условиях. Поддержка импорта различных форматов (MS Excel, tsv и других) дает возможность создавать первичные базы данных на основе уже имеющихся источников информации и впоследствии редактировать их. Исходный код содержится в репозитории открытого доступа GitHub и может быть установлен специалистами, не имеющими специальной подготовки.

В настоящее время база данных, развернутая на сайте shirokuma.compbio.ru/db, содержит детальную информацию об образцах белого медведя (*Ursus maritimus*), содержащихся в коллекции Зоологического института РАН (Санкт-Петербург), и охватывает около 500 животных, добытых в ходе исследовательских экспедиций и промысловых работ с середины XIX века по настоящее время.

Созданный нами программный комплекс также может быть адаптирован для работы с практически любыми коллекциями биологических или медицинских образцов.

МАТЕРИАЛЫ И МЕТОДЫ

Источники данных. База данных BearBase содержит информацию об образцах белого медведя, хранящихся в коллекции Зоологического института РАН. Общее число документированных образцов составляет 506, время добычи животных находится в интервале от 1840 до 2000 года и представлено преимущественно черепами различной степени сохранности. Каталогизация образцов проводилась с использованием фотофиксации и маркирования каждого уникальным идентификатором и QR-кодом для облегчения поиска и идентификации в последующем.

Основным источником информации об образцах служили карточки каталога, содержащие информацию о таких параметрах, как тип образца (череп, скелет), место хранения (номер шкафа и полки), возраст и пол животного, место и время сбора образца, время фотофиксации и описания. Помимо каталогизации музейных образцов, база данных служит основой для реализации проекта по геномике белого медведя и дополнительно включает в себя информацию о выделении ДНК из костных тканей или зубов, качестве выделенной ДНК и последующем секвенировании.

Первоначально данные о музейных образцах заносили в электронную таблицу в формате MS Excel, которая затем была конвертирована в реляционную базу данных (платформа SQLite, см. ниже), что позволило создать основу данного ресурса.

Реализация системы хранения данных. Реляционная база данных, составляющая основу

BearBase, функционирует на платформе SQLite (версия 3.45.0 или выше). Элементами хранения в базе данных являются музейные образцы, пользователю доступно представление информации как в формате карточки отдельного образца, так и списка образцов. В данный момент для каждого образца доступно 24 информационных поля в таблицах базы данных, содержащих различную информацию, в том числе текстовую, цифровую, данные геолокации и изображения.

В основу схемы базы данных положена концепция виртуальных полей, позволяющая создавать произвольные таблицы для представления различных типов данных. Пример создания виртуальной таблицы приведен на рис. 1. В данном примере доступный пользователю список образцов представлен как виртуальная таблица (выделена зеленым цветом), содержащая изменяемый список виртуальных полей. Физические таблицы не доступны пользователю (выделены на схеме желтым) и являются источником данных для создания виртуальной таблицы. Таблица "Sample" является ключевым коннектором для других таблиц и содержит внутренние ключи базы данных, однозначно идентифицирующие образец и являющиеся идентификаторами для других таблиц. Таблица "Data" используется для хранения реальных данных (например, фотографий, данных геолокации, фенотипических параметров и т. д.). Каждый элемент в этой таблице связан с уникальным идентификатором образца и может быть представлен в различных форматах, численном (например, данные геолокации, такие как широта) или строковом (например, путь к файлу изображения для графических данных). В таблице "Struct" хранятся описания всех полей виртуальных таблиц, видимых пользователю, а также описание типов данных и полей. Наконец, таблица "SampleStructData" связывает вышеупомянутые таблицы через внутренние и внешние ключи и обеспечивает доступ к данным для выполнения поисковых запросов.

В приведенном примере Sample_ID ссылается на три различных типа данных в таблице "Struct": 2 – Галерея, 3 – Остеологический идентификатор и 4 – Коллектор. Поле Data_ID из таблицы "SampleStructData" ссылается на три поля в таблице "Data", содержащие реальные данные различных типов (а именно, путь к графическим файлам в виде строки, данные геолокации в виде чисел с плавающей точкой и имя коллектора в виде строки).

Такой модульный подход позволяет модифицировать существующие типы данных и добавлять практически неограниченное количество новых полей. Виртуальное поле также дает возможность расширить стандартный список типов данных, обычно представленных в традиционных

Физические таблицы

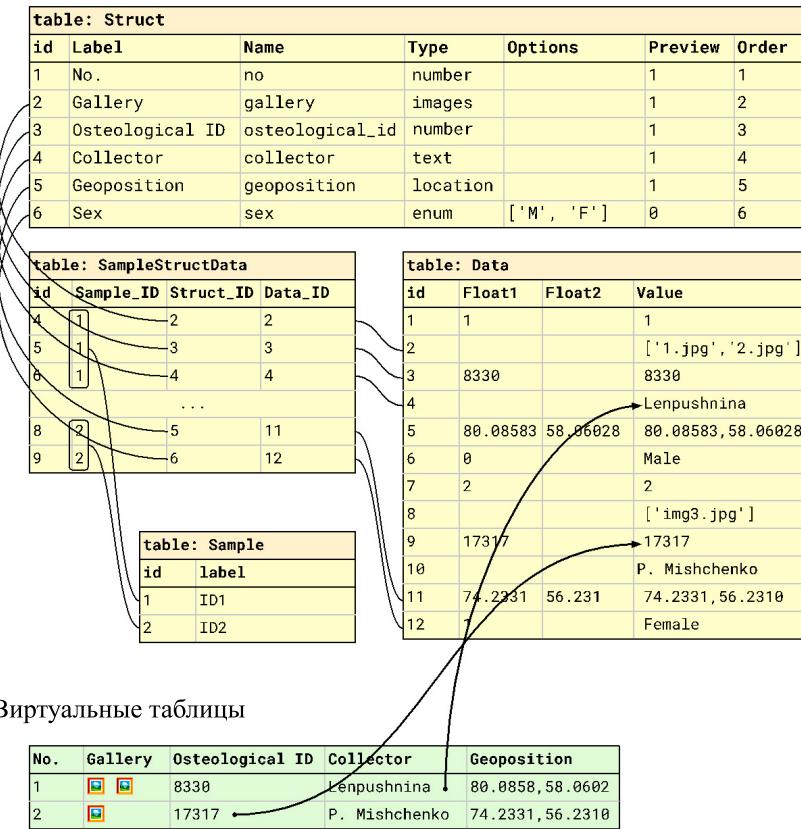


Рис. 1. Схематическое представление концепции виртуальных таблиц. Виртуальная таблица выделена зеленым цветом, физические таблицы – желтым.

реляционных СУБД. В частности, в BearBase используются такие типы, как геолокационные данные (с возможностью визуализации на карте) и изображения.

Платформа поддерживает следующие стандартные типы данных:

- string – простой тип стандарта CHAR;
- числовые – целые числа и числа с плавающей точкой;
- date and time – реализованы в SQL-стандартах DATE и DATETIME;
- geodata – система включает проверку долготы и широты и визуализацию карты;
- picture – графические данные с возможностью предварительного просмотра;
- enum – выбор значения из фиксированного списка или контролируемого словаря.

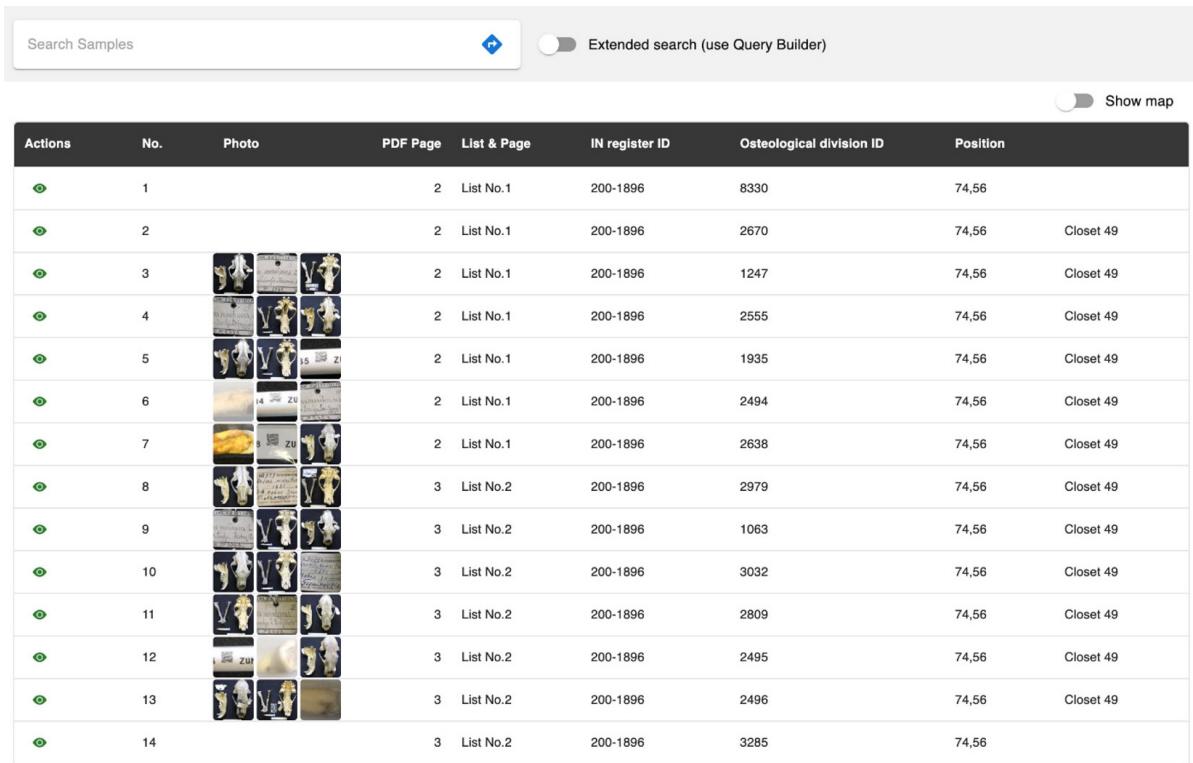
Развёртывание базы данных и пользовательского интерфейса. Исходный код системы доступен в репозитории открытого доступа (github.com/comp-bio/BearBase); используются только необходимые зависимости, что упрощает установку и поддержку. Программное обеспечение может

быть легко установлено без использования административного (root) доступа к устройствам.

BearBase включает в себя серверную и клиентскую части с веб-интерфейсом. Данные хранятся в локальных каталогах, которые не зависят от клиент-серверного интерфейса и могут легко обмениваться между различными рабочими группами. Такой подход обеспечивает высокую гибкость и простоту управления данными, включая резервное копирование и контроль версий.

Серверная часть реализована на python3 с использованием фреймворков Flask и SQLAlchemy, что обеспечивает стабильность, масштабируемость и эффективную производительность. Для управления и обработки изображений используется библиотека Pillow. Пользовательский интерфейс на стороне клиента разработан с помощью инструментов программирования, поддерживающих динамические веб-приложения, а именно React, Webpack, Babel, Sass и Axio.

Дополнительная поддержка мобильных устройств с оптимизацией интерфейса делает BearBase удобным и доступным инструментом вне зависимости от типа устройства. Мобильное



The screenshot shows a table with 14 rows, each representing a sample entry. The columns are: Actions, No., Photo, PDF Page, List & Page, IN register ID, Osteological division ID, and Position. The 'Photo' column contains small thumbnail images of the samples. The 'List & Page' column indicates the page number (2 or 3) and list number (e.g., List No.1, List No.2). The 'Position' column shows coordinates (e.g., 74,56, Closet 49).

Actions	No.	Photo	PDF Page	List & Page	IN register ID	Osteological division ID	Position
	1		2	List No.1	200-1896	8330	74,56
	2		2	List No.1	200-1896	2670	74,56, Closet 49
	3		2	List No.1	200-1896	1247	74,56, Closet 49
	4		2	List No.1	200-1896	2555	74,56, Closet 49
	5		2	List No.1	200-1896	1935	74,56, Closet 49
	6		2	List No.1	200-1896	2494	74,56, Closet 49
	7		2	List No.1	200-1896	2638	74,56, Closet 49
	8		3	List No.2	200-1896	2979	74,56, Closet 49
	9		3	List No.2	200-1896	1063	74,56, Closet 49
	10		3	List No.2	200-1896	3032	74,56, Closet 49
	11		3	List No.2	200-1896	2809	74,56, Closet 49
	12		3	List No.2	200-1896	2495	74,56, Closet 49
	13		3	List No.2	200-1896	2496	74,56, Closet 49
	14		3	List No.2	200-1896	3285	74,56

Рис. 2. Общий вид базы данных BearBase.

приложение может работать в автономном режиме и накапливать данные в локальном хранилище устройства, а затем синхронизировать их либо с центральным сервером, либо с локальным сервером, который может быть развернут без доступа в интернет.

Менеджмент данных. Ввод данных в BearBase осуществляется посредством импорта из основных форматов: текстовых файлов с разделителями, электронных таблиц MS Excel и файлов MS Access. Первоначальное создание таблиц может быть выполнено простым импортом данных из вышеупомянутых источников (см. пример на GitHub). В ходе этой процедуры создается полный набор таблиц в соответствии с информацией, представленной в файлах источника, и производится заполнение их данными. Поля базы данных и их формат могут быть в дальнейшем изменены.

Работа в BearBase может осуществляться в двух режимах:

– Редактор/менеджер – режим создания и изменения схемы базы данных. При этом доступно добавление, редактирование или удаление полей данных в таблицах и курирование всей информации об образцах;

– Внешний пользователь – режим доступа только для чтения, обеспечивающий безопасное

представление данных об образцах в открытой сети без возможности модификации базы данных.

Кроме того, платформа включает в себя набор опций для представления и работы с геолокационными данными. В частности, для каждого образца с геолокационными данными генерируется фрагмент карты для отображения точки сбора и облегчения сортировки и группировки образцов. Также может быть построена обзорная карта, на которой отображаются все образцы с геолокацией.

Поисковые инструменты включают простой текстовый поиск по базе данных и более сложный конструктор запросов с реляционными операторами AND, OR, NOT, CONTAINS и т. д.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Главная страница базы данных (shirokuma.compbio.ru/db) представлена в виде списка всех образцов в режиме просмотра «внешний пользователь» (рис. 2). Данный список включает в себя настраиваемый набор информационных полей для сортировки, возможность простого текстового поиска, конструктор запросов и управляющий элемент для создания общей карты всех образцов, для которых доступна информация по геолокации. Крайний левый столбец

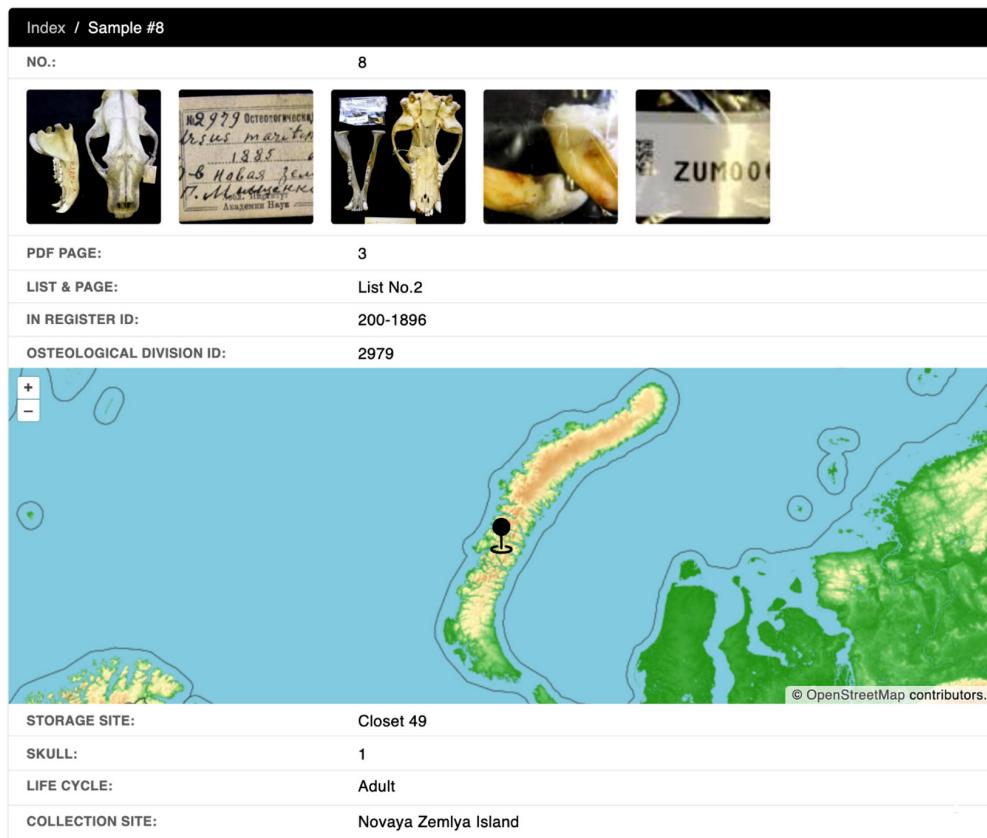


Рис. 3. Вид карточки индивидуального образца (фрагмент).

списка содержит ссылку для просмотра индивидуальных образцов (рис. 3). Карточка индивидуального образца включает в себя уменьшенные копии фотографий музеиного материала, карточки учета и метки с QR-кодом. Далее отображаются все информационные поля, соответствующие данному животному, включая карту с местом сбоя образца (при наличии данных геолокации).

В настоящее время существует множество систем управления информацией, включая традиционные СУБД и базы данных на основе графов [8, 9], предлагающих широкий выбор для научного сообщества. Однако лишь некоторые из них могут быть доступны неопытным пользователям без подключения к интернету. Этот фактор серьезно препятствует внедрению систем управления сбором образцов в практику исследовательских экспедиций и полевых выездов. Наша платформа призвана восполнить этот пробел и ускорить внедрение систем управления информацией в такие традиционные области биологии, как зоология и ботаника, предоставив исследователям простую и надежную систему для работы с гетерогенными данными с удобным графическим интерфейсом, внутренней проверкой данных, развитой системой поиска и поддержкой графических форматов.

В частности, традиционные реляционные СУБД используют определенную схему и не позволяют легко модифицировать базу данных «на лету». Каждый объект в базе данных должен быть детально описан набором определенных атрибутов, что требует перепроектирования схемы базы данных при изменении атрибутов и отношений между ними. Используемая нами концепция виртуальных таблиц и полей позволяет легко настраивать схему базы данных в соответствии с требованиями конкретной задачи. Кроме того, в отличие от имеющихся аналогов, требующих прав суперпользователя для установки и, в большинстве случаев, функционала docker, наша база данных имеет простую процедуру настройки с минимальными требованиями к дополнительным библиотекам и пакетам. BearBase сочетает в себе преимущества традиционных реляционных баз данных и гибкость новых подходов (графовые базы данных, распределенные базы данных).

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена при финансовой поддержке Российского научного фонда (грант № 22-74-00038).

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит описания исследований с использованием людей и животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

1. Benham P. M. and Bowie R. C. K. Natural history collections as a resource for conservation genomics: Understanding the past to preserve the future. *J. Heredity*, **114** (4), 367–384 (2023).
DOI: 10.1093/jhered/esac066, PMID: 36512345
2. Card D. C., Shapiro B., Giribet G., Moritz C., and Edwards S. V. Museum genomics. *Annu. Rev. Genet.*, **55**, 633–659 (2021).
DOI: 10.1146/annurev-genet-071719-020506, PMID: 34555285
3. Raxworthy C. J. and Smith B. T. Mining museums for historical DNA: advances and challenges in museomics. *Trends Ecol Evol.*, **36** (11), 1049–1060 (2021).
DOI: 10.1016/j.tree.2021.07.009, PMID: 34456066
4. Crescente J. M., Guidobaldi F., Demichelis M., Formica M. B., Helguera M., and Vanzetti L. S. Phenobook: an open-source software for phenotypic data collection. *Gigascience*, **6** (4), 1–5 (2017).
DOI: 10.1093/gigascience/giw019, PMID: 28327910
5. Raubach S., Schreiber M., and Shaw P. D. GridScore: a tool for accurate, cross-platform phenotypic data collection and visualization. *BMC Bioinformatics*, **23** (1), 214 (2022).
DOI: 10.1186/s12859-022-04755-2, PMID: 35668357
6. Bán M., Boné G. M., Bérces S., Barta Z., Kovács I., Ecsedi K., and Sipos K. OpenBioMaps – self-hosted data management platform and distributed service for biodiversity related data. *Earth Sci. Inform.*, **15**, 2007–2016 (2022). DOI: 10.1007/s12145-022-00818-3
7. Kimble M., Allers S., Campbell K., Chen C., Jackson L. M., King B. L., Silverbrand S., York G., and Beard K. medna-metadata: an open-source data management system for tracking environmental DNA samples and metadata. *Bioinformatics*, **38** (19), 4589–4597 (2022). DOI: 10.1093/bioinformatics/btac556, PMID: 35960154
8. Timón-Reina S., Rincón M., and Martínez-Tomás R. An overview of graph databases and their applications in the biomedical domain. *Database (Oxford)*, **2021**, baab026 (2021).
DOI: 10.1093/database/baab026, PMID: 34003247
9. Page R. D. M. Ozymandias: a biodiversity knowledge graph. *PeerJ*, **7**, e6739 (2019).
DOI: 10.7717/peerj.6739, PMID: 30993051

BearBase: Database of the Museum Collection of Polar Bear Specimens

I.V. Bezdvornykh*, A.A. Kanapin*, A.A. Samsonova*, E.A. Orlova*, M.V. Sablin, A.V. Abramov**, V.V. Platonov**, and D. Hirata***

*Peter the Great St. Petersburg Polytechnic University, Polytekhnicheskaya ul. 29, St. Petersburg, 195251 Russia

**Zoological Institute, Russian Academy of Sciences, Universitetskaya nab. 1, St. Petersburg, 199034 Russia

Museum collections of biological specimens are invaluable for research, particularly when paired with modern genetic analysis and omics technology. Although documentation has evolved from paper to digital formats, managing the diverse data requirements of these collections continues to pose challenges. Current software tools are designed for specific tasks and often require advanced technical skills, which can limit accessibility. To tackle these challenges, we developed BearBase, a user-friendly and portable database system that does not necessitate professional expertise. It features both server and client components, supports offline usage, and can sync with mobile devices for fieldwork. Users can easily import data from formats like MS Excel and modify database structure as needed. The open-source code is available on GitHub. Currently, BearBase hosts data on approximately 500 polar bear (*Ursus maritimus*) specimens from the collection of the Zoological Institute of the Russian Academy of Sciences (St. Petersburg), encompassing those harvested during research and commercial activities dating back to the 19th century. Moreover, the system can be customized to manage other collections of biological or medical specimens.

Keywords: biological collections, genomics, population genetics, Arctic fauna, databases