

С.В. Буйкин, Е.Ю. Брагина, Л.А. Конева

РАЗРАБОТКА СТРУКТУРЫ БАЗЫ ДАННЫХ ДЛЯ БИОБАНКОВ

УДК: 616-01/09:004.652.4

Сбор, первичная обработка и последующее использование биологического материала порождают множество проблем организационного характера, которые с успехом решаются с применением биобанков. Одной из основных составляющих успешной работы биобанков является практичность, функциональность и надежность базы данных для хранения сопроводительной информации. В статье рассмотрены основные принципы и проблемы разработки баз данных для биокolleкционных ресурсов на примере Биобанка НИИ медицинской генетики СО РАМН.

Ключевые слова: биобанк, база данных, мультифакториальные заболевания, гены.

Collection, preprocessing and use of biological samples cause many organizational problems, which are successfully resolved with help of biobanks. Among the components of successful work of biobank is a practical, functional and reliable database for storage of information. In the paper, principals and problems of database development for biobanks are considered, taking as example the biobank of the Institute of Medical genetics SB RAMS.

Ключевые слова: биобанк, база данных, мультифакториальные заболевания, гены.

Введение. Интенсификация генетических исследований, посвященных изучению основ комплексных заболеваний, позволила сформулировать основные положения относительно дизайна, а также качественного и количественного состава исследуемых групп [8, 10, 12, 13]. Требования к размеру, однородности и описанию выборок становятся все строже [1, 7], это влечет за собой значительное увеличение временных и материальных затрат по регистрации, стандартизации, сохранению биологического материала и сопроводительной информации. Оптимальным решением данной проблемы являются биобанки, которые консолидируют большой массив биологических образцов и сопроводительной информации [2]. Одной из основных составляющих успешной работы биобанков является практичность, функциональность и надежность базы данных (БД) для хранения сопроводительной информации. В статье рассмотрены основные принципы и проблемы разработки баз данных для биокolleкционных ресурсов на примере БД Биобанка НИИ медицинской генетики СО РАМН.

Проектирование БД – одна из наиболее сложных и ответственных задач, связанных с созданием информационной системы. В результате ее решения должны быть определены содержание БД, способ организации данных и инструментальные средства управления данными. Процесс проектирования БД включает в себя следующие этапы: 1) концептуальное или информационно-логическое про-

ектирование; 2) логическое проектирование; 3) физическое проектирование [4]. Информационно-логическая модель представляет собой описание структуры и динамики предметной области, характера информационных потребностей пользователей системы в терминах, понятных пользователю и независимых от реализации системы. Таким образом, на этом этапе определяются: основные объекты предметной области (объекты, о которых должна храниться информация в БД), атрибуты объектов, связи между объектами, основные запросы к БД. Логическое проектирование подразумевает решение вопроса об оптимальном отображении объектов предметной области в абстрактные объекты модели данных, чтобы это отображение не противоречило семантике предметной области. Физическое проектирование призвано решить задачу обеспечения эффективности выполнения запросов к БД [5]. Все этапы должны учитывать специфику проектируемой БД.

Биобанк НИИМГ СО РАМН предназначен для хранения уникального биологического материала с подробным описанием генетической, демографической и клинической информации. Биобанк насчитывает несколько тысяч образцов ДНК и крови индивидов с различными патологиями, включая сердечно-сосудистые, иммунозависимые, инфекционные, моногенные и другие заболевания, а также ДНК индивидов из популяций Восточной, Западной Сибири и Средней Азии [2]. Сопроводительная информация по образцам включает: 1) описание анамнеза заболевания; 2) результаты физикального и лабораторного обследования индивидов; 3) результаты молекулярно-генетического анализа для большого числа генетических маркеров (результаты генотипирования, секвенирования, кариотипирования

и т.д.); 4) информацию относительно этических вопросов, связанных с возможностью использования биологических образцов для различных исследований (информированные согласия индивидов, дата и номер заключения этического комитета); 5) процедурные особенности работы с биологическими образцами (метод выделения образца ДНК, концентрация, качество, объем образца и т.д.). Таким образом, сопроводительная информация биологического материала представляет собой разноплановые по своему содержанию данные, отличающиеся для разных патологий, и хранится в виде таблиц Excel, текстовых и графических файлов.

Формализация медико-биологической информации, характеризующей биологические образцы биобанка, является сложной задачей. В связи с тем, что такие сведения необходимы для оценки связи того или иного генотипа (полиморфного варианта, гена) с развитием заболевания, к сбору и хранению сопроводительной информации должны предъявляться жесткие требования, позволяющие минимизировать ошибки и неточности [9].

Для обеспечения сохранности данных и повышения эффективности научных исследований была поставлена цель разработки БД. Весь массив данных был структурирован с учетом целей и задач Биобанка НИИМГ СО РАМН. При проектировании БД рассматривались не только текущие потребности исследователей, но и возможности использования биологического материала и информации в будущих медико-биологических исследованиях. Кроме того, учитывались вопросы соблюдения этических норм и правил, а также интеллектуальной собственности, связанные с хранением и представлением результатов ис-

НИИ медицинской генетики СО РАМН (г. Томск): **БУЙКИН Степан Вячеславович** – н.с., stepan_buikin@medgenetics.ru; **БРАГИНА Елена Юрьевна** – н.с., bragina@ssmu.ru; **КОНЕВА Лада Анатольевна** – н.с., Lada.Koneva@ssmu.ru.

следования образцов биологического материала биобанка [11].

БД Биобанка НИИМГ СО РАМН представляет собой реляционную базу данных, т.е. организован в виде связанных таблиц, где строки соответствуют записям, а столбцы таблиц являются атрибутами отношения. Широкое распространение реляционной модели данных объясняется в первую очередь простотой представления и формирования БД, универсальностью и удобством обработки данных [3,5].

БД Биобанка НИИМГ СО РАМН состоит из четырех основных взаимосвязанных блоков: 1) основная информация; 2) образцы биологического материала; 3) результаты молекулярно-генетического анализа; 4) пользователи БД (рисунок).

Основная информация включает в себя описание персональных данных об объекте исследования; о состоянии здоровья, условиях жизни и труда; о заболеваниях; о результатах лабораторных обследований; о родственных связях с другими индивидами в БД и дополнительные справочные таблицы.

Образцы биологического материала – содержат информацию о ДНК, образцах крови, клетках, тканях и т.д. В блок внесены сведения о месте и дате забора биологического материала, о методе выделения ДНК, условиях и месте хранения образца. Также в данном разделе регистрируется расход образцов.

Результаты молекулярно-генетического анализа представляют справочную таблицу названий и описаний изучаемых генов; уникальный номер полиморфизма по базе данных SNP (www.ncbi.nlm.nih.gov); варианты генотипов; тип полиморфизма (SNP, VNTP, del, ins и т.д.); хроматограммы и последовательности участков геномов в текстовом формате; результаты исследования мтДНК; сводные таблицы, объединяющие между собой блоки: «индивид», «образец биологического материала», «варианты генотипов».

Пользователи. Данный раздел хранит информацию о пользователях БД с указанием уровня доступа; историю запросов, для сложных запросов, производимых пользователем в БД; историю операций, для документирования производимых пользователем действий (внесение, редактирование, удаление данных и т.д.) и перечень разрешенных IP адресов.

При проектировании БД Биобанка НИИМГ СО РАМН разработаны требования к интерфейсу, отражающие специфику работы биобанка. Интерфейс БД организован в форме двух

основных частей – Пользовательской и Административной.

Административная часть необходима для внесения и изменения информации в БД пользователями, имеющими соответствующие права доступа, и содержит основной (внесение данных в БД) и вспомогательные (редактирование справочных таблиц) интерфейсы. Администратор имеет право на внесение, редактирование и удаление данных во все основные блоки, назначение прав доступа другим пользователям, имеет возможность создания резервной копии БД.

Пользовательская часть необходима для доступа к имеющимся данным, формирования выборок по заданным параметрам, отображения результатов импорта или экспорта данных и содержит: форму авторизации, форму создания запроса, форму выдачи результата, форму сохраненных запросов пользователя, форму внесения/редактирования данных базы, форму просмотра операций пользователя.

На настоящий момент завершены этап проектирования БД Биобанка НИИМГ СО РАМН. При реализации проекта для обеспечения безопасности данных на уровне информационной системы запланирован авторизованный доступ к БД, ограничение доступа пользователей (только с указанных IP адресов), защита от автоматического подбора пароля к БД (с блокировкой атакующего IP адреса), предоставление доступа к персональным данным, внесение и редактирование данных ограниченному кругу лиц. Также предусмотрено хранение истории авторизации пользователей, запросов пользователей и внесенных изменений в БД. Для БД Биобанка НИИМГ СО РАМН планируется эксплуатация в локальной компьютерной сети института, с возможностью защищенного доступа из внешней сети Internet.

Таким образом, в Биобанке НИИМГ СО РАМН разработана структура



Структура организации БД Биобанка НИИМГ СО РАМН

уникального, функционального, специализированного электронного ресурса (БД), отвечающая требованиям специалистов, работающих в различных областях медицинской генетики и соответствующая требованиям федерального закона об обеспечении безопасности персональных данных [6]. БД Биобанка НИИМГ СО РАМН может использоваться в научно-исследовательских центрах медицинского и общепрофильного профиля, для обеспечения сохранности биологических коллекций и повышения эффективности научных исследований.

Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009 – 2013 годы (гос. контракт №П722).

Литература

- Брагина Е.Ю. Влияние численности контрольной выборки на значимость ассоциаций генетических маркеров с развитием мультифакториальных заболеваний / Е.Ю. Брагина, С.В. Буйкин // Якутский медицинский журнал. – 2009. – №2(26). – Стр. 142-144.
- Bragina E.J. Influence of control sample size on significance of associations of genetic markers with development of multifactorial diseases / E.J. Bragina, S.V. Buikin // Yakut Medical Journal. – 2009. – №2 (26). – P. 142-144.
- Брагина Е.Ю. Биологические банки: проблемы и перспективы их использования в исследованиях генетических аспектов комплексных заболеваний человека / Е.Ю. Брагина, С.В. Буйкин, В.П. Пузырев // Медицинская генетика. – 2009. – №3. – С. 20-27.
- Bragina E.J. Biobanks: problems and perspectives of their use in genetic studies of

human common diseases / E.J. Bragina, S.V. Buikin, V.P. Puzyrev // Medical Genetics. – 2009. – №3. – P. 20-27.

3. Игнасимуту С. Основы биоинформатики / С. Игнасимуту. – М., 2007 – Ижевск: НИЦ «Регулярная и хаотическая динамика», Институт компьютерных исследований. – 320 с.

Ingasimutu S. Principles of bioinformatics. – Izhevsk: SRC «Regular and chaotic dynamics», Institute of computer research, 2007. – 320 p.

4. Введение в базы данных: учебное пособие / И.П. Карпова. – М: Издательство МГИЭМ, 2005. – 96 с.

Introduction to databases. The manual / I.P. Karpova. – M: Publishing house MGIEEM, 2005. – 96 p.

5. Кузнецов С.Д. Основы современных баз данных, информационно-аналитические материалы / С.Д. Кузнецов // <http://citforum.ru/database/osbd/contents.shtml>. – 2008.

Kusnetsov S.D. Basics of modern databases, information-analytical materials / S.D. Kusnetsov // <http://citforum.ru/database/osbd/contents.shtml>. – 2008.

6. Федеральный закон от 27.07.2006 N 152-ФЗ (ред. от 27.12.2009) "О персональных данных" // Российская газета. – 2006. – № 165.

The Federal law 7/27/2006 N 152-FZ (red. by 12/27/2009) "About personal data" // The Russian newspaper. – 2006. – №165.

7. Feero W.G., Guttmacher A.E. Genomewide Association Studies and assessment of the risk of disease // N Engl J Med. – 2010. – V. 363. – P. 166-176.

8. Frazier L., Sparks E., Sanner J.E. Biobanks and biomarker research in cardiovascular disease // J Cardiovasc Nurs. – 2008. – V. 23(2). – P. 153-158.

9. Gottweis H., Zatloukal K. Biobank governan-

ce: trends and perspectives // Pathobiology. – 2007. – V.74(4). – P. 206-211.

10. Little J., Higgins J.P.T., Ioannidis J.P.A. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE Statement // Hum.Genet. – 2009. – V. 125. – P. 131-151.

11. Shickle D. The consent problem within DNA biobanks // Stud Hist Philos Biol Biomed Sci. – 2006. – V.37(3). – P. 503-519.

12. Swede H., Stone C.L., Norwood A.R. National population-based biobanks for genetic research // Genetics in Medicine. – 2007. – V.9(3). – P. 141-149.

13. The Wellcome trust Case Control Consortium. Genome-wide association study of 14000 cases of seven common diseases and 3000 shared controls // Nature. – 2007. – V. 447. – P. 661-678.

ОРГАНИЗАЦИЯ ЗДРАВООХРАНЕНИЯ, МЕДИЦИНСКОЙ НАУКИ И ОБРАЗОВАНИЯ

Т.Е. Уварова, Т.Е. Бурцева

ДИНАМИКА ОСНОВНЫХ ПОКАЗАТЕЛЕЙ МЕДИЦИНСКОЙ ОБЕСПЕЧЕННОСТИ НАСЕЛЕНИЯ В РАЙОНАХ ПРОЖИВАНИЯ КОРЕННЫХ МАЛОЧИСЛЕННЫХ НАРОДОВ СЕВЕРА

УДК 616-053.2(571.56)

Проведен анализ основных показателей медицинской обеспеченности населения в районах проживания коренных малочисленных народов Севера РС (Я) за 1990, 1995, 1999-2003 гг. Выделены основные тенденции: сокращение числа больничных учреждений, фельдшерско-акушерских пунктов, коечного фонда, численности кадровых ресурсов.

Ключевые слова: больничные учреждения, больничные койки, врачебные амбулаторно-поликлинические учреждения, фельдшерско-акушерские пункты.

The analysis of the basic indexes of health care system of the population in residence of the native small in numbers people of the North of Republic Sakha (Yakutia) for 1990, 1995, 1999-2003 is carried out. The basic tendencies such as decrease of number of medical institutions, feldsher's-obstetrical stations, bed complement and number of personnel resources are detected.

Keywords: medical institutions, hospital beds, medical out-patient-polyclinic institutions, feldsher's-obstetrical stations.

В сложные годы социально-экономических преобразований ключевым вопросом для здравоохранения Российской Федерации стало повышение структурной эффективности отрасли. Реструктуризация лечебно-профилактической сети в стране предусматривала сокращение коечного фонда, централизацию и реорганизацию больничных учреждений и диспансеров. В РС (Я) за 1990-2005 гг. число больничных учреждений снизилось на 20%, больничных коек – на 31, фельдшерско-акушерских пунктов – на 25%. За вышеуказанный период число врачей всех специальностей повысилось на 2,6, число средних медицинских работников сократилось на 2,0% [3]. Предметом исследований, представленных в данной статье, стали основные показатели медицинской

обеспеченности населения в районах проживания коренных малочисленных народов Севера (КМНС) РС (Я) по материалам Территориального органа Федеральной службы государственной статистики по РС (Я) за 1990, 1995, 1999-2003 гг.

За 1990-2003 гг. число больничных учреждений в районах проживания КМНС снизилось с 51 до 48 (на 5,9%). Удельный вес их в общей численности больничных учреждений республики снизился с 16,8% в 1990 г. до 16,6 в 2003 г. (табл. 1). Число больничных коек сократилось с 1009 в 1990 г. до 852 в 2003 г. (на 15,6%). Между тем за тот же период показатель обеспеченности больничными койками возрос с 139 до 151 на 10000 населения (на 8,6%) (табл. 1). Таким образом, приведенные

выше данные свидетельствуют о снижении числа больничных учреждений и больничных коек в районах проживания КМНС, что соответствовало аналогичной общереспубликанской тенденции данного периода. Рост показателя

Таблица 1

Число больничных учреждений и коек в районах проживания коренных малочисленных народов Севера

	1990	1995	2002	2003	2003 г. в % к 1990 г.
Число больничных учреждений					
РС (Я)	303	365	291	290	- 4,3
Районы проживания КМНС	51	58	55	48	- 5,9
Удельный вес в РС (Я), %	16,8	15,9	18,9	16,6	
Число больничных коек					
Всего РС (Я)	17471	16080	14349	14213	- 18,6
Районы проживания КМНС	1009	1029	894	852	- 15,6
На 10000 населения РС(Я)	156	155	151,3	150	- 3,8
Районы проживания КМНС на 10000 населения	139	151	158	151	+ 8,6

УВАРОВА Татьяна Егоровна – к.м.н., и.о. зав. лаб. ЯНЦ КМП СО РАМН; **БУРЦЕВА Татьяна Егоровна** – к.м.н., зам. директора по науке ЯНЦ КМП СО РАМН, bourtsevat@rambler.ru.