

*Д.В. Колдаков***О подходах к созданию комплексного источника по истории населенных пунктов Алтайского края с применением технологии баз данных***D.V. Koldakov***Different Approaches to the Creation of the Complex Source on the History of Altai Region Settlements with an Application of Database Technologies**

Статья посвящена вопросам формирования комплексного источника по истории населенных пунктов Алтайского края с применением технологий баз данных. Рассматриваются методики создания баз данных на основе различных категорий источников, в том числе электронных.

Ключевые слова: базы данных, населенные пункты.

The article is devoted to the problem of creation of the complex source on the history of Altai region settlements with an application of database technologies. Some methods to create databases on the basis of different kinds of sources including electronic are examined.

Key words: databases, settlements.

Компьютеризация исторических исследований, особенно в последние три десятилетия, дает немало примеров эффективного применения технологий построения информационных систем, методов математического моделирования исторических процессов, моделирования пространственных процессов и явлений. В рамках выполнения научно-исследовательских работ уже накоплены значительные электронные массивы данных. Появление и рост числа оцифрованных копий источников в сети Интернет также позволяет значительно расширить базу исследований, поэтому роль машиночитаемых данных в настоящее время трудно переоценить.

Цель настоящей статьи – рассмотрение методик формирования с применением технологии баз данных комплексного источника по истории населенных пунктов Алтайского края.

Начало коллекции машиночитаемых данных по истории населенных пунктов Алтая было положено созданием в конце 1990-х гг. первой версии базы данных «Населенные пункты Алтайского края». В ее основу лег список, составленный известным исследователем и краеведом Ю.С. Булыгиным в 1997–1999 гг. Эта работа была поддержана РГНФ (грант №01-97-00415). В указанном списке перечислены все учтенные населенные пункты, существующие или существовавшие на территории Алтайского края в административных границах на начало 1990-х гг. Он включает сведения о современном (для исчезнувших – последнем)

названии населенного пункта и его статусе (виде); встречающиеся в источниках более ранние названия и их варианты (также с указанием статуса населенного пункта); привязку населенного пункта к гидросети; выявленный по источникам период существования, представленный датами образования (или наиболее раннего достоверного свидетельства о существовании) и датой наиболее позднего учета или упоминания в источниках. Отдельно отмечены населенные пункты, месторасположение которых на территории современного Алтайского края установить не удалось. Список составлен на основе данных как архивных материалов, так и сведений из опубликованных источников, однако сами сведения об источниках не приведены. Достаточно хорошая структурированность материала позволила автору настоящей статьи перевести его в формат реляционной базы данных. Реляционные базы данных широко используются для создания машиночитаемых источников на основе материалов, изначально имеющих четкую структуру и относящихся к так называемым highly structured historical sources (структурированным историческим источникам): материалам переписей, личным делам и карточкам, анкетам, справочникам, которые содержат разнотипную информацию (текстовую, числовую и т.д.) на уровне отдельных объектов, соединенной в одном формуляре. Формуляр такого источника часто представляет практически готовую структуру базы данных [1, с. 95–96]. Тем не менее потребовалось переструктурировать спи-

сок средствами макропрограммирования текстового процессора MS Word, чтобы потом его можно было импортировать в таблицы базы данных, созданной в СУБД MS FoxPro 2.5 for Windows.

Указанная база данных была рассчитана не только на выполнение справочно-информационных запросов. Она создавалась для проведения пространственного анализа материала с целью выявления основных закономерностей образования населенных пунктов на территории Алтайского края средствами геоинформационных систем (ГИС). Основным средством для этого стала широко применяемая в историко-пространственных исследованиях географическая информационная система Atlas GIS, зарекомендовавшая себя как надежное средство пространственно-географического анализа исторических данных. Полученные результаты исследования нашли отражение в ряде публикаций [2, с. 25–44; 3, с. 21–27].

Несмотря на то, что исходный информационный массив был представлен в электронной форме и его предварительная обработка была в значительной степени автоматизирована, полностью избежать ручной работы не удалось. Предварительный анализ списка позволил выявить наличие порой совершенно идентичных фрагментов информации. Так, неоднократно встречались населенные пункты с одинаковыми названиями, существовавшие на территории нескольких или даже одного современного административного района. Период их существования мог различаться или совпадать. Например, на территории Хабарского района на 1926 г. зафиксировано три поселка Васильевка. На территории, подчиненной Барнаулу, в 1939 г. на Оби находились две бакенские будки под номером 1 и две – под номером 2, и учитывались они только в этом году. Налицо совпадение всех значений атрибутов объекта. Согласно же теории реляционных баз данных, каждая запись (кортеж) в отношении (таблице) должна быть уникальна, т.е. однозначно идентифицироваться отдельным атрибутом или их набором. В последнем приведенном случае это реально только при введении в структуру таблицы поля с суррогатным ключом, который должен генерироваться СУБД или приложением, либо самим пользователем при занесении данных. Частично это удалось решить еще на этапе подготовки данных к импорту в базу данных, а затем для совпадающих названий, особенно весьма распространенных (Ивановка, Березовка и т.п.), провести корректировку. Вручную также пришлось устранять опечатки, в некоторых случаях унифицировать написание названий населенных пунктов.

Разрабатывая эту версию базы данных, мы учитывали то обстоятельство, что вполне возможна дальнейшая корректировка ее структуры в связи с расширением круга решаемых задач. Практическое наполнение компонентов этой достаточной обширной структуры

могло быть как параллельным, так и отдаленным друг от друга по времени.

Появление достаточно дружественных к пользователю средств разработки баз данных, таких как СУБД MS Access, предоставляющих, в частности, удобные средства для разработки запросов, позволило перенести данные из СУБД MS FoxPro 2.5 в новую среду MS Access 2000 и приступить к созданию второй версии БД «Населенные пункты Алтайского края». Немаловажным обстоятельством, повлиявшим на такой выбор, явилась тесная интеграция этой СУБД с другими приложениями пакета Microsoft Office, что дает возможность использования дополнительного инструментария при работе с данными. СУБД MS Access пользуется заслуженной популярностью у исследователей-историков России и стран СНГ, о чем, например, свидетельствует частота упоминаний этого средства разработки баз данных в публикациях Ассоциации «История и компьютер». В качестве еще одного примера можно привести информационную систему «Города и села Среднего Урала в XX веке», разработанную исследователями Уральского государственного университета [4, с. 168–183].

Во второй версии база данных подверглась дальнейшим трансформациям. В отдельные таблицы были выделены сведения о старых названиях (всего 5714 записей для 3423 населенных пунктов), о ближайшем объекте гидросети.

Для дальнейшего изучения процессов заселения территории Алтая с применением геоинформационной системы MapInfo 7.8 была произведена процедура геокодирования объектов базы данных. Мы использовали цифровую карту Алтайского края в масштабе 1:500000, переданную для научных целей Институтом водных и экологических проблем СО РАН. Геокодирование объектов выполнил к.и.н. М.Е. Чибисов.

Публикация в начале 2000-х гг. нового справочника по административно-территориальным изменениям [5] позволила уточнить списочный состав населенных пунктов, а также внести определенные корректировки в базу данных, связанные с датировкой снятия с учета. В настоящее время база данных «Населенные пункты Алтайского края» (свидетельство РФ №2007620001 от 09.01.2007, авторы Д.В. Колдаков, В.Н. Владимиров, И.Г. Силина, М.Е. Чибисов) содержит сведения о 8434 объектах поселенческой сети Алтая XVIII–XX вв. На «Алтайском региональном историческом портале» представлена ее версия по существующим населенным пунктам края [6].

В связи с дальнейшим совершенствованием базы данных возникла необходимость уточнения того круга источников, данные которых использовал при составлении своего списка Ю.С. Булыгин. Конечно, часть источников можно определить а priori: так, сведения о населенных пунктах, первые упоминания которых датируются 1926 г., определенно взяты из «Спис-

ков населенных мест Сибирского края», датировки 1950–1980-х гг. – из справочников административного и статистического учета и т.д. [7–11]. Но датировки XVIII – первой половины XIX в. сделаны по архивным документам, и наличие сведений о них в базе данных крайне желательно. В связи с этим база данных «Населенные пункты Алтайского края» была подвергнута дальнейшей реструктуризации, и в ее состав были добавлены таблицы «Архивные источники о населенных пунктах» и «Опубликованные источники о населенных пунктах».

Таблица «Архивные источники о НП» была заполнена данными по материалам личного фонда Ю.С. Булыгина, хранящегося в государственном архиве Алтайского края (Ф. ФР-1765. Д. 119–133), а также сведениями, которые удалось обнаружить в ряде публикаций, имеющих отношение к истории районов Алтайского края и его населенных пунктов. Решено было занести не только ссылки на первичные документы, но и ссылки на те картотеки исследователя, где они были обнаружены, а также представляющие интерес пометки Ю.С. Булыгина относительно этих данных. Фиксировались и категории датировки: «первое упоминание» и «другая», если таковая присутствовала и сделана по другому источнику. Таблица содержит 725 записей.

Таблица «Опубликованные источники о НП» включает данные, также собранные по материалам картотек Ю.С. Булыгина и публикациям. В этой таблице 380 записей. В основном это сведения о дореволюционных источниках [12–17]. Сведения этих таблиц позволяют уточнить основной круг источников по истории населенных пунктов Алтая. Хранение таких данных в справочно-информационных целях более чем оправдано. В целом можно положительно оценить такую «открытость» базы данных для изменений и включения дополнительных наборов данных. Это дает возможность расширить информационную базу исследований, например, путем добавления сведений о численности населения. Так, в процессе работы над источниковой базой был выявлен источник [18], сведения которого не нашли отражения в исходном списке Ю.С. Булыгина. В связи с тем, что в 1920-е гг. новые населенные пункты образовывались достаточно интенсивно, информация из этого «Списка...» для части территории Алтайского края позволяет уточнить имеющиеся датировки, собственно списочный состав селений и в определенной мере их месторасположение. Указанный список был переведен в формат электронных таблиц MS Excel.

Заметим, что хотя в большинстве случаев для создания электронных источников применяется технология баз данных, в некоторых случаях именно комплексное применение различных средств дает в оптимальные сроки эффективно перевести в машиночитаемый вид данные исторических источников.

Так, подготовку исходных данных формулярного характера, различного рода таблицы со статистическими показателями удобно производить в электронных таблицах, например в MS Excel, и далее импортировать их базы данных. Но особенно это важно при формировании нового информационного массива на основе другого машиночитаемого источника, например файла в текстовом формате, отсканированного и сохраненного в графическом формате изображения и т.п. Это актуально в настоящее время, когда исследователь может получить доступ к таким источникам, выставленным в электронных библиотеках в сети Интернет [19].

Одним из способов, позволяющим сопоставить сведения разных источников, является организация дополнительной базы данных с последующей ее связью с другой БД на основе общих ключевых полей, т.е. для этого нужно создать и заполнить в соответствующей таблице присоединяемой базы столбец, содержащий в нашем случае идентификатор населенного пункта. Далее, в MS Access выполняется команда на организацию связи основной БД с таблицами подключаемой базы, после чего можно выполнять все необходимые запросы.

Таким образом, можно автономно разрабатывать БД окружения с собственными структурами. А собственно БД «Населенные пункты Алтайского края» в таком случае становится своеобразным «каркасом», на который «нанизываются» другие информационные массивы. По сути дела это реальный путь к организации банка данных по населенным пунктам Алтая, содержащий самую разнообразную информацию.

Приведенная выше методика формирования «окружения» для основной базы данных была опробована нами при создании еще одного машиночитаемого массива, который получил условное название «СНМСК-Алтай». В основу его положен известный «Список населенных мест Сибирского края». Данные этого весьма представительного по объему списка могут оказаться полезными для локализации исчезнувших к настоящему времени селений, для пространственного анализа размещения их на территории региона, анализа размещения населения и т.д.

Необходимо отметить, что территория, к которой относились Барнаульский, Бийский, Каменский, Славгородский и Рубцовский округа Сибирского края, в совокупности больше территории нынешнего Алтайского края, и значительное количество населенных пунктов располагалось на территории сопредельных регионов (Новосибирской и Кемеровской областей, Республики Алтай). Учитывая то, что определенный задел в виде базы данных «Населенные пункты Алтайского края» уже имеется, решено было заносить сведения списка по принципу «как есть», с тем, чтобы в дальнейшем локализовать населенные места, зарегистрированные в 1926 г., в рамках существующего

на настоящий момент административно-территориального деления.

В связи с большим объемом данных списка данные вводились «послойно», т.е. не все характеристики населенного пункта сразу, а только некоторые из них, остальные же решено было «подключать» позднее. В первую очередь таковыми стали название населенного пункта, его статус, год возникновения, число хозяйств, учтенных при проведении переписи 1926 г., численность мужского и женского населения, данные по преобладающей национальности. Для удобства дальнейшей работы с источником вносились порядковые номера селений по округу, по району и номер страницы в издании. Такие «метки» позволяют, если понадобится, быстро и точно сослаться на данные оригинального источника.

Занесенный с использованием MS Excel список населенных пунктов, представленный в отдельных выпусках (всего 5408 строк), был подвергнут сплошной проверке. Далее полученный материал был импортирован в созданную в MS Access базу данных.

Самым трудоемким стало сопоставление записей списка с записями базы данных «Населенные пункты Алтайского края» для их однозначной идентификации и локализации в рамках существующего административно-территориального деления. Во-первых, далеко не всегда легко сразу сопоставить весьма распространенные названия населенных пунктов (Малиновка, Покровка и т.п.), особенно находящихся в соседних районах или в даже пределах одного района. Это потребовало проверки и обращения к другим, как к более ранним, так и более поздним по происхождению, источникам, сверки с географическими картами. Во-вторых, необходимо было проверять все, включая и уже зафиксированные в базе названия населенных пунктов. В-третьих, некоторые названия населенных пунктов в обеих базах так сильно различались по написанию (особенно это касалось фамилий владельцев хуторов и заимок), что для их соотнесения приходилось очень тщательно сверять списки названий селений в пределах административных районов.

Сопоставление сведений двух баз данных позволило определить населенные пункты, как расположенные вне современной территории края, так и находящиеся в ее пределах, но, вероятно, пропущенные при формировании списка Ю.С. Булыгина. Также были обнаружены дублирующиеся записи о селениях. В результате проделанной работы имеются два взаимосвязанных набора данных, которые можно под-

вергнуть дальнейшей обработке и анализу. Объединяющим эти данные «стержнем» служит уникальный идентификатор населенного пункта, взятый из одной базы данных и хранящийся в другой.

На следующем этапе был произведен отбор в базе данных «СНМСК-Алтай» записей о тех селениях, которые находятся на территории края, вне зависимости от того, удалось ли их точно сопоставить с записями базы данных «Населенные пункты Алтайского края» или нет. Таких осталось 5012, из них пока не удалось соотнести 29.

После сопоставления и идентификации населенных пунктов в базу данных «СНМСК-Алтай» были добавлены данные о наличии объектов производственной и социальной сфер, а также об их количестве. Общее число записей в таблице «Объекты в населенных пунктах» составило 3696.

Полученный массив данных можно анализировать по различным параметрам, например, оценить качество датировки возникновения населенных пунктов, представленной в списке, основываясь на сопоставлении баз данных. Так, удалось выяснить, что расхождение по датировке, например в пределах ± 5 лет в XVIII в. отмечено у 29 селений, в XIX в. – у 67, в XX столетии – у 1544. Были уточнены некоторые характеристики системы расселения середины 1920-х гг., в частности, распределение населенных пунктов по статусу с учетом их локализации в рамках территории современного Алтайского края [20, с. 37–46].

Подобным образом можно организовать информационные массивы на основе списков населенных мест Томской губернии 1893 и 1911 гг. и ряда других источников. Это позволит обеспечить не только их более быстрое формирование, но и перевести в машиночитаемый вид значительный объем данных, который, например, при выборочном извлечении определенных информационных пластов был бы совершенно не затронут.

Таким образом, информационно-справочный потенциал рассмотренных баз данных весьма значителен. Их анализ, несомненно, позволит расширить и углубить наши представления о процессах заселения и освоения территории Алтая в XVIII–XX столетиях. Представленная в настоящей статье технология их формирования может быть применена при создании новых таких источников и формирования полноценного банка данных по истории населенных пунктов Алтайского края.

Библиографический список

1. Гарскова И.М. Базы и банки данных в исторических исследованиях. – М. ; Геттинген, 1994.
2. Владимиров В.Н., Колдаков Д.В. Образование населенных пунктов Алтайского края: история во времени

и пространстве // История. Карта. Компьютер : сб. науч. статей. – Барнаул, 1998.

3. Владимиров В.Н., Колдаков Д.В. К истории образования населенных пунктов Алтайского края // Инфор-

мационные технологии в гуманитарных науках. – Казань, 1999.

4. Мазур Л.Н., Бродская Л.И. Эволюция сельских поселений Среднего Урала в XX веке: опыт динамического анализа. – Екатеринбург, 2006.

5. Справочник административно-территориальных изменений в Алтайском крае. 1981–2001. – Барнаул, 2002.

6. История населенных пунктов Алтая [Электронный ресурс]. – URL: <http://new.hist.asu.ru/base/rindex.html>.

7. Список населенных мест Сибирского края. Т. 1: Округа юго-западной Сибири. – Новосибирск, 1928.

8. Алтайский край. Административно-территориальное деление на 1 января 1973 г. – Барнаул, 1973.

9. Алтайский край. Административно-территориальное деление. – Барнаул, 1978.

10. Алтайский край. Административно-территориальное деление. – Барнаул, 1992.

11. Административно-территориальные изменения Алтайского края за 1939–1991 гг. – Барнаул, 1992.

12. Хозяйственно-статистическое описание крестьянских волостей Алтайского округа. – Барнаул, 1886.

13. Описание заселенных и незаселенных переселенческих участков, образованных до 1896 года. – Барнаул, 1896.

14. Материалы по исследованию мест водворения переселенцев в Алтайском округе. – Барнаул, 1899.

15. Список населенных мест Томской губернии на 1911 год. – Томск, 1911.

16. Книга образования переселенческих участков. 1885–1912 гг. – Томск, 1913.

17. Цифровые данные по учету сельского населения при поземельном устройстве Алтайского округа. – Барнаул, 1916.

18. Списки населенных мест 41-й районной волости Алтайской губернии. На 10-е апреля 1924 г. – Барнаул, 1924.

19. Томский государственный университет. Электронная библиотека [Электронный ресурс]. – URL: http://www.lib.tsu.ru/index_main.php?id=511

20. Колдаков Д.В. Список населенных мест Сибирского края» как источник по истории населенных пунктов Алтая // Пятые научные чтения памяти Ю.С. Булыгина : сборник научных трудов / под ред. Ю.М. Гончарова, В.А. Скубневского. – Барнаул, 2009.