

Комплексирование данных губернских документов и карт

© В.Г. Щекотилов

Тверское отделение РГО

globus-t@yandex.ru

© М.В. Шалаева

Тверь

Maria-geo@yandex.ru

© О.С. Лазарева

ТвГТУ

LazOSTvGU@mail.ru

Аннотация

Предложен подход для комплексирования информационных моделей табличной части поездных геокодируемых архивных документов XIX в. (списков населенных мест, статистических сведений, экономических примечаний) с комплексом обработки и представления архивных карт

Работа выполнена при поддержке РФФИ, грант 14-06-97507 p_центр_a.

1 Введение

С развитием техники в компьютерную обработку вовлекаются архивные текстовые, табличные и картографические документы [1]. В прикладных и научных исследованиях, образовании и просвещении начинают активно использоваться информационные ресурсы по архивным картам губерний и территорий России XIX века, а также сопутствующие табличные и текстовые архивные документы с элементами, которые могут быть геокодированы.

Среди архивных карт значительное место занимают крупномасштабные топографические межевые карты 8 губерний съемки А.И. Менде и военно-топографические карты [5].

Среди тестовых и табличных архивных документов можно отметить поездные списки населенных мест (СНМ), экономические примечания (ЭП) и статистические сведения (СС).

Для списков населенных мест известны практические реализации создания информационных ресурсов [3–5]. Для Карелии были созданы ГИС «Электронный каталог населенных мест Карелии XV–XX вв.» [3] и функционирующая «Картографическая справочно-информационная система «Генеральное межевание Олонецкой губернии» [5], в которой реализован комплекс по

данным генерального межевания XVIII в.

На сайте «Историко-генеалогической организации «Литера Ру» идет формирование компоненты «Списки населенных мест Российской Империи» [4]. Связь с картами на данном ресурсе отсутствует.

Для зарубежного ресурса с архивными картами «David Rumsey Map Collection» отсутствует геокодированная часть. (<http://www.davidrumsey.com/>) [9].

Созданный и практически апробированный авторами научный задел в области обработки и представления архивных карт в форме автоматизированных информационных ресурсов и пространственных данных XIX в. позволяет перейти к решению задачи разработки методов эргономичного комплексирования в информационных ресурсах картографических и пространственно атрибутированных архивных материалов (крупномасштабных топографических межевых карт, списков населенных мест, экономических примечаний, статистических сведений) на примере создания макета Интернет-ресурса Тверской губернии.

Данные исследования направлены на устранение несоответствия между существующими формами использования и комплексирования архивных документов XIX в. и современными возможностями вычислительной техники и коммуникаций.

2 Обработка и представление архивных карт и пространственных данных

На основе результатов исследований, проводимых с 2003 г., и развития разработанных методик обработки и представления архивных карт, а также списков населенных мест разрабатывается методическое, информационное и алгоритмическое обеспечение для формирования первичных информационных моделей (ИМ) архивных документов и создания общедоступного эргономичного функционального Интернет-ресурса для научного, образовательного и просветительского использования [7].

Для решения данной задачи предлагается формирование информационной модели для комплекса

Труды 16-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL-2014, Дубна, Россия, 13–16 октября 2014 г.

поездных документов XIX в.: списков населенных мест; экономических примечаний; статистических сведений, представляющих собой объемные статистические документы того времени. В настоящее время авторами созданы информационные ресурсы только в части СНМ (URL: <http://boxpis.ru/svg/?p=1590>).

Составной частью метода являются частные методики формирования данных для информационного ресурса на основе СНМ, ЭП, СС:

– методики создания функциональных автономных электронных версий СНМ (апробировано в форме автоматизированного электронного СНМ), СС, ЭП;

– методики комплексирования в единый функциональный информационный ресурс комплекса архивных документов: СНМ, ЭП, СС;

– методики комплексирования информационных ресурсов с ресурсом по крупномасштабным архивным картам;

– методики использования в прикладных исследованиях созданного Интернет-ресурса и частных электронных версий документов.

Известны попытки создания ограниченных по функциональности информационных ресурсов по СНМ: «Родная Вятка» (URL: <http://rodnaya-vyatka.ru/>); «Литера Ру. Историко-генеалогическая организация» (URL: <http://www.litera-ru.ru/>).

Создание подходов комплексного автоматизированного использования архивных документов различной структуры (картографических, табличных, описательных) является нетривиальной задачей, требует применения технологий баз данных, географических информационных систем, интернет-технологий [7]. По этой причине для уникальных региональных (губернских) комплексов информационно насыщенных документов архивные крупномасштабные карты – списки населенных мест – экономические примечания – статистические сведения (АК-СНМ-ЭП-СС; общая аббревиатура – КНЭС) автоматизированные информационные ресурсы отсутствуют.

В мировой практике известны реализации отдельных ресурсов по архивным картам: коллекция D. Rumsey (<http://rumsey.com/>) объединения листов отдельных карт Франции и Германии. Комплексирование ресурса с геокодируемыми документами отсутствует.

Известен отечественный региональный ресурс с использованием данных XVIII в. (планы генерального межевания и экономические примечания) – «Картографическая справочно-информационная система «Генеральное межевание Олонецкой губернии» (URL: <http://maps.karelia.ru/mez/>).



Рис. 1. Схема формирования и использования информационных ресурсов (ИР)

3 Комплексование картографического информационного ресурса с архивными документами

Развитием картографического информационного ресурса является его комплексование с архивными документами, имеющими повторяющиеся геокодируемые элементы. Схема формирования и использования информационного ресурса по архивным картам, текстовым и табличным документам XIX в. представлена на рис. 1.

На обобщенной схеме отражены три функциональных группы:

1) наполнение информационного ресурса ИМ архивных карт, текстовых и табличных документов;

2) выполнение геокодирования пространственно-атрибутируемых объектов информационного ресурса (установление связей между картами и текстово-табличными документами);

3) использование ресурса в научных, образовательных, справочных и просветительских целях по направлениям: истории, географии, краеведения, топонимики, генеалогии, туризма и др.

Примеры пространственно-атрибутируемых объектов: дачи, владельцы, мельницы, заводы, постоянные дворы, почтовые станции.

4 Геокодируемые элементы поузедных архивных документов

В таблице 1 представлены повторяющиеся геокодируемые элементы из СНМ, ЭП, СС в части Калязинского уезда.

Можно отметить, что количество населенных мест по СНМ составляет 819, селений по СС 707, дач по ЭП 1078.

Таблица 1. Повторяющиеся геокодируемые объекты Калязинского уезда

| № | Геокодируемый элемент | СНМ | ЭП | СС | Кп ¹ | ИМ ³ |
|---|--------------------------------|-----|------|-----|-----------------|-----------------|
| 1 | Населенные места | 819 | | | 9 | 7,4 |
| 2 | Алфавитный указатель селений | | | 707 | 7 | 4,9 |
| 3 | Таб. 1. Население | | | 707 | 62 | 43,8 |
| 4 | Таб. 2. Землевладение и аренды | | | 707 | 67 | 47,4 |
| 5 | Таб. 3. Земледелие и т.д. | | | 707 | 66 | 46,7 |
| 6 | Примечания по селениям | | | 707 | 102 | 7,1 |
| 7 | Дачи, межевая ведомость | | 1078 | | 274 | 29,1 |
| 8 | Алфавит дачам | | 1078 | | 4 | 4,3 |
| 9 | Владельцы, 26 стр. | | 3685 | | 4 | |

В таблице 1 обозначено:

¹ – Кп – количество параметров для объекта;

² – параметры (характеристики) представлены в сплошном текстовом виде;

³ – характеристика количества данных – произведение количества записей на число параметров, деленное на 1000;

⁴ – для большей части дач несколько записей для разных деревень (селений, частей);

⁵ – для многих указано несколько совладельцев; для многих приводится список дач во владении; не указано 341 не частных владельцев (государственных имуществ и крестьян, ведомств и учреждений).

В таблице 1 сделана попытка сравнить количественный объем структурированных данных СНМ, СС, ЭП. Можно заключить, что количество данных из СС и ЭП на порядок превышает данные СНМ. Это обстоятельство подчеркивает актуальность перевода их в электронную форму и геокодирование.

Среди геокодируемых объектов из таблицы 1 по структуре данных и взаимным связям можно выделить следующие группы:

1) первичные геокодируемые: населенные места из СНМ и дачи из ЭП;

2) повторяющиеся с небольшим отличием иной: селения из СС (требуется сопоставление со СНМ);

3) ссылающиеся на геокодированные:
– на селения (из СС таб.1, таб.2, таб.3, примечания по селениям);
– на дачи: из ЭП владельцы (множественное); алфавит дачам.

Для первой группы объектов следует выполнить геокодирование (координатное атрибутирование записей).

Для второй группы необходимо выполнить:

– идентификацию списка селений со СНМ (атрибутирование записей номером по СНМ);

– для неидентифицированных по СНМ селений необходимо выполнить геокодирование.

Для третьей группы объектов возможна алгоритмическая (программная) реализация запросов на формирование координатной информации в табличной и картографической формах представления.

Следует отметить, что в дополнение к объектам, перечисленным в табл. 1, могут формироваться наборы иных геокодируемых объектов. Примерами могут являться:

– храмовые сооружения («Храмы России», URL: <http://temples.ru/>), список которых формируется в основном по текстовым архивным документам;

– ветряные и водяные мельницы, перечень которых в основном формируется по карте.

5 Информационная модель геокодируемых элементов

На рис. 2. представлена структура информационных блоков непосредственно или опосредованно имеющих атрибутирование на местности.

На рис. 2 в изображениях блоков приняты следующие обозначения: ГК – записи подлежат геокодированию; Кл. – уникальный ключ для записей.

Информационная модель списков, представленных на рис. 2, может быть представлена в виде:

$$M^{SNM} = \{N^{SNM}_i; P^{SNM}_i; G^{SNM}_i; i=1...K^{SNM}\} \quad (1)$$

$$M^{SS} = \{U^{SS}_j; V^{SS}_j; S^{SS}_j; N^{SNM}_j; G^{SS}_j; j=1...K^{SS}\} \quad (2)$$

$$T_1^{SS} = \{P1^{SS}_j; j=1...K^{SS}\} \quad (3)$$

$$T_2^{SS} = \{P2^{SS}_j; j=1...K^{SS}\} \quad (4)$$

$$T_3^{SS} = \{P3^{SS}_j; j=1...K^{SS}\} \quad (5)$$

$$Pr^{SS} = \{P1^{SS}_j; j=1...K^{SS}\} \quad (6)$$

$$Md^{EP} = \{N^D_i; G^D_i; i=1...K^D\} \quad (7)$$

$$Mv^{EP} = \{Fv^{EP}_p; Pv^{EP}_p; Sd^{EP}_p; p=1...K^V\}, \quad (8)$$

где

M^{SNM} – ИМ списка населенных мест; N^{SNM}_i – номер НМ; P^{SNM}_i – атрибуты НМ по СНМ; G^{SNM}_i – координаты НМ; K^{SNM} – количество НМ по СНМ;

M^{SS} – ИМ списка селений по СС; U^{SS}_j – номер уезда; V^{SS}_j – номер волости; S^{SS}_j – номер селения; N^{SNM}_j – номер соответствующего НМ по СНМ; G^{SS}_j – координаты селения (если оно отсутствует в СНМ); K^{SS} – количество селений;

T_1^{SS} – ИМ характеристик таблицы 1 СС; $P1^{SS}_j$ – атрибуты селения;

T_2^{SS} – ИМ характеристик таблицы 2 СС; $P2^{SS}_j$ – атрибуты селения;

T_3^{SS} – ИМ характеристик таблицы 3 СС; $P3^{SS}_j$ – атрибуты селения;

Pr^{SS} – ИМ текстовых характеристик селения по примечанию СС; Pr^{SS} – характеристика селения;

Md^{EP} ИМ дач по ЭП; N^D_i – номер дачи; G^D_i – координаты дачи; K^D – количество дач;

Mv^{EP} ИМ перечня владельцев дач по ЭП; Fv^{EP}_p – ФИО владельца; Pv^{EP}_p – атрибуты владельца; Sd^{EP}_p – список дач владельца; K^V – количество владельцев.

На данном этапе исследований связи между таблицами предлагается реализовать на программном уровне в запросах.

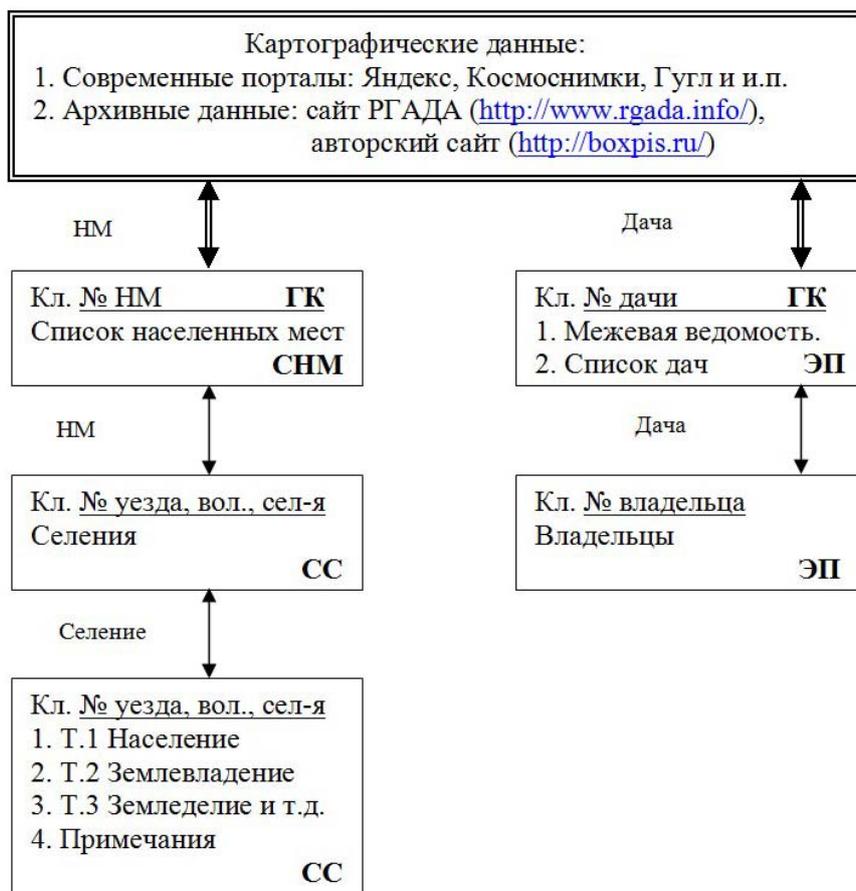


Рис. 2. Структура геокодируемых информационных блоков

На рис. 2 отражены следующие связи между таблицами:

– таблицы статистических сведений («Т.1 Население», «Т.2 Землевладение», «Т.3 Земледелие», «Примечания») связаны с таблицей «Селения» через поле «Селение»;

– таблица «Селения» статистических сведений связана со «Списком населенных мест» через поле «Селение/Населенное место»;

– таблица «Владельцы» экономического примечания связана через поле «Дача» с таблицами «Межевая ведомость», «Список дач».

На рис. 2 отражены следующие объекты геокодирования:

– по положению населенного места для таблицы «Список населенных мест»;

– по положению внутренней точки дачи для таблицы «Список дач» экономического примечания.

Для выбранной записи одного из исходных документов или сформированной выборки возможно позиционирование на карте.

На представленном на рис. 2. множестве списков могут быть реализованы следующие запросы на формирование выборок:

- дачи около выбранного населенного места;
- населенные места около выбранной дачи;
- дачи выбранного владельца;
- владельцы выбранной дачи;
- дачи владельца выбранной дачи.

В части архивных карт подход также апробирован для доступных листов других крупномасштабных карт (топографических межевых Владимирской, Нижегородской, Симбирской, Ярославской губерний; карты Кавказа; военно-топографической карты Европейской России). Комплекс обработанных крупномасштабных архивных карт охватывает территорию 20 субъектов РФ, 7 государств.

Формируется практика размещения частей информационного ресурса (например, информационная модель карты губернии) на внешних исследовательских и образовательных Интернет-ресурсах и локальных вычислительных сетях.

6 Заключение

Практика разработки и сопровождения интернет-ресурса по крупномасштабным архивным картам XIX в. [7] показала повышенную востребованность сегмента с комплексированием посредством геокодирования списка населенных мест и крупномасштабной карты. Создаваемые комплексные ресурсы по архивным картам и документам России XIX в. могут стать составной частью Интернет-ресурсов библиотек и архивов (например, www.rgada.info/) [2].

На примере Тверской губернии предложено формирование комплекса из списка населенных

мест, таблиц, атрибутирующих селения из статистических сведений, а также списки дач и владельцев из экономических примечаний. Для данного комплекса таблиц предлагается выполнить с использованием интернет-ресурса с топографической межевой картой съемки А.И. Менде геокодирование населенных мест и дач. Набор формируемых запросов позволит связать таблицы между собой, а также с архивной и современной картами.

Предложенный подход позволяет создавать ресурсы, расширяющие возможности по доступу и автоматизированному использованию архивных документов с исторической, географической и краеведческой информацией в научных, образовательных (вузы, школы), просветительских (общества «Знание», РГО, историческое, краеведов, генеалогии), управленческих (кадастр, градостроительство, туризм) и справочных целях.

Исследования выполнены в рамках гранта РФФИ и Правительства Тверской области «Разработка метода комплексирования в информационных ресурсах архивных материалов XIX века (карт, списков населенных мест, экономических примечаний, статистических сведений) на примере создания макета интернет-ресурса Тверской губернии и апробации его для задач изучения и оценки особенностей природопользования для XIX и начала XXI веков» №14-06-97507 р_центр_а.

Литература

- [1] Белова Е.Б., Бородкин Л.И., Гарскова И.М. и др. Историческая информатика / под ред. Л.И. Бородкина, И.М. Гарсковой. – М., 1996.
- [2] Голубинский А.А., Якушин Г.Б. Российский государственный архив древних актов. – <http://www.rgada.info/>
- [3] Жуков Ю.А., Ляля Е.В. Создание геоинформационной системы (ГИС) «Электронный каталог населенных мест Карелии XV–XX вв.», «Рябининские чтения – 2007», Музей-заповедник «Кижы». – Петрозаводск, 2007. – 497 с.
- [4] Списки населенных мест Российской Империи. Историко-генеалогическая организация «Литера Ру». – <http://maps.litera-ru.ru/spiski.php>
- [5] Чернякова И.А. Картографическая справочно-информационная система «Генеральное межевание Олонецкой губернии». – <http://maps.karelia.ru/mez/>
- [6] Щекотилов В.Г. Разработка информационной системы «Историческая, географическая и социально-экономическая параллель XIX–XXI веков по архивной информации съемки А.И. Менде» на основе интернет-технологий // 4-я Междунар. конф. по исторической географии «Глобальные и региональные проблемы исторической географии», РГО, Санкт-Петербург, 2011 г. – С. 244–247.

- [7] Щекотилов В.Г. Обработка и представление архивных карт [Электронный ресурс]. – URL: <http://www.boxpis.ru>
- [8] Щекотилов В.Г., Бугрова Н.И., Щекотилова М.В. Комплексирование описательных документов объектов историко-культурного наследия с картографическими ресурсами и системами // Записки Филиала РГГУ в г. Великий Новгород: науч. сб. – 2011. – Вып. 9. – С. 57–62.
- [9] David Rumsey, Edith M. Punt. Cartographica Extraordinaire: The Historical Map Transformed. 2004. – <http://www.davidrumsey.com/>

Aggregation of Data from Provincial Documents and Maps

Vladimir G. Shekotilov, Maria V. Shalaeva,
Oksana S. Lazareva

The proposed approach for the integration of information models of the structured part of the geocoding archival documents of the XIX century (including lists of settlements, statistics, economic notes) with the facilities for processing and presentation of archival maps is presented.