

ПЕРЕПИСЬ НАСЕЛЕНИЯ: ПЕРВЫЕ РЕЗУЛЬТАТЫ И ПРОБЛЕМЫ ОБРАБОТКИ ДАННЫХ

О ПЕРВЫХ РЕЗУЛЬТАТАХ ВСЕРОССИЙСКОЙ ПЕРЕПИСИ НАСЕЛЕНИЯ 2010 ГОДА

И.А. Збарская, канд. социол. наук,
Федеральная служба государственной статистики

Всероссийская перепись населения 2010 г. в истории отечественных переписей была самой продолжительной по проведению - почти девять месяцев. Она началась 1 апреля переписью оленеводов в Ямало-Ненецком автономном округе и завершилась 20 декабря в труднодоступных населенных пунктах Свердловской и Тюменской областей, где доставка переписчиков и необходимой документации возможна по устоявшимся зимним дорогам. Перепись основной части населения страны прошла с 14 по 25 октября. По сравнению с предыдущей переписью продолжительность опроса была увеличена на четыре дня и 12-дневный период, по нашему мнению, является наиболее оптимальным с точки зрения возможности относительно равномерно распределения нагрузки переписного персонала.

Другой новацией 2010 г. было упрощение процедуры переписи. Традиционно в предыдущих переписях населения собственно опросу населения предшествовал предварительный обход, во время которого переписчик обходил каждое жилое помещение своего учетного участка для определения помещений с постоянно проживающим населением. Целью этого обхода был отбор жилых помещений для проведения переписи по программе выборочного наблюдения, так называемому «длинному» вопроснику. В связи с тем что при Всероссийской переписи населения 2010 г. все постоянное население страны опрашивалось по форме Л, необходимость в предварительном обходе отпала, и тем самым работа переписного персонала была существенно упрощена.

В первый день переписи - 14 октября - к обходу жилых помещений, опросу населения и заполнению переписных листов в субъектах Российской Федерации приступили более 425 тыс. переписчиков. Было открыто для населения почти 76 тыс. стационарных переписных участков (в 2002 г. работало 17 тыс. стационарных участков).

Во время переписи жители страны воспользовались предоставленным в соответствии с Федеральным законом «О Всероссийской переписи населения» выбором - ответить на вопросы программы переписи дома или дать о себе сведения на стационарном переписном участке. Преобладающая часть населения России ответила на вопросы переписчика по месту своего жительства. По

данным проводимого оперативного мониторинга, на стационарных участках было переписано более 8 млн. жителей страны (примерно 6% населения; в 2002 г. услугами стационарных участков воспользовались около 5% жителей), в Москве - почти каждый пятый житель (в 2002 г. - каждый третий), в Санкт-Петербурге - каждый 10-й житель (столько же, сколько и в 2002 г.). Такая форма организации переписи, как использование стационарных участков, впервые успешно примененная в 2002 г. и при настоящей переписи населения подтвердила свою эффективность.

Более миллиона жителей страны заявили переписчикам о своем отказе участвовать в переписи населения по личным (например, религиозным) соображениям; 2,6 млн. человек переписчики не застали дома, поскольку они отсутствовали на протяжении всего периода переписи.

Для обеспечения полноты сбора сведений о населении и получения при переписи общей численности населения, проживающего на территории Российской Федерации, в соответствии с процедурой, предусмотренной Федеральным законом «О Всероссийской переписи населения» (ст. 6, п. 3), из административных источников были получены и внесены в переписные листы данные о поле и возрасте на 3,6 млн. человек. Численность этой категории населения в 2002 г. составляла примерно 1,5 млн. человек.

Для обеспечения полноты и правильности учета населения, во избежание недоучета или двойного счета проведены контрольные мероприятия - инструкторами совместно с переписчиками был проведен выборочный (10%-ный) контрольный обход жилых помещений по учетным участкам (26-29 октября 2010 г.). В период переписи составлялась форма С «Список лиц, подлежащих Всероссийской переписи населения 2010 года», форма КС «Список лиц для контроля правильности заполнения переписных листов» и форма СПР «Справка о прохождении переписи» - для контроля за заполнением переписных листов.

Всего для работы на переписных, инструкторских и учетных участках было привлечено с учетом резерва 617 тыс. временных переписных работников. Основная часть временных переписных работников (36%) была сформирована из неработающего населения (до-

мохозяйки, пенсионеры); работники предприятий, социальных служб и общественных организаций составили 26%; студенты, учащиеся, преподаватели высших и средних учебных заведений - 25%.

С учетом опыта 2002 г. заблаговременно была начата работа с Министерством образования и науки Российской Федерации по участию студентов в переписи, поскольку, в первую очередь для крупных мегаполисов, решить проблему найма персонала очень затруднительно. Только в Москве надо было набрать свыше 33 тыс. переписчиков.

Совместно была разработана программа производственной практики - работа переписчиком после прохождения соответствующего обучения. Были направлены рекомендации в субъекты Российской Федерации, заключены соглашения между вузами и ТОГСаами, в которых определялись условия участия студентов и квоты по каждому государственному образовательному учреждению высшего профессионального образования. Такая процедура полностью соответствует ФЗ «О высшем и послевузовском профессиональном образовании» и Положению о порядке проведения практики студентов образовательных учреждений высшего профессионального образования (утв. Приказом Минобрнауки России от 25.03.2003 № 1154, п. 8).

С каждым студентом или учащимся заключался гражданско-правовой договор, в котором предусматривалась работа переписчиком, входящая в программу практики студента, его нагрузка, период работы, порядок сдачи-приемки выполненных работ и другие условия, установленные Гражданским кодексом Российской Федерации.

Важно отметить, что Всероссийская перепись населения 2010 г. прошла в спокойных с точки зрения общественного правопорядка условиях. Безопасность переписного персонала и правопорядок во время проведения переписи населения обеспечивались четкой и слаженной работой сотрудников органов внутренних дел. В целях обеспечения безопасности переписчиков и всего населения в период проведения переписи территориальными подразделениями МВД России осуществлялся круглосуточный мониторинг оперативной обстановки, охрана помещений переписных участков, в необходимых случаях сотрудники органов внутренних дел сопровождали переписчиков.

В обеспечении общественного порядка и безопасности ежедневно задействовалось в среднем около 140 тыс. сотрудников органов внутренних дел, в том числе непосредственно на обеспечение безопасности мероприятий по переписи населения - до 80 тыс. сотрудников. К проведению правоохранительных мероприятий ежедневно привлекались до 2 тыс. курсантов образовательных учреждений МВД России.

Подготовка к переписи населения в ряде субъектов совпала с проведением выборов различного уровня в

законодательно установленный единый день голосования (в 2010 г. - 10 октября). Росстат и ЦИК России осуществляли координацию действий и подготовили рекомендации по организации и выполнению мероприятий по подготовке и проведению Всероссийской переписи населения 2010 г. и проведению выборов. Распоряжением ЦИК России 19 января 2010 г. была образована Рабочая группа по взаимодействию избирательных комиссий и территориальных органов государственной статистики при подготовке к предстоящим выборам и переписи населения. Были разработаны совместные Рекомендации о взаимодействии избирательных комиссий и территориальных органов Росстата при подготовке и проведении выборов в единый день голосования 10 октября 2010 г. и Всероссийской переписи населения 2010 г., определены мероприятия, где возможно «пересечение интересов» и пути их решения: привлечение персонала; подбор помещений; проведение информационной работы.

Кроме того, рассматривались вопросы, связанные с совмещением доставки переписчиков, переписной документации, с доставкой работников избирательных комиссий, избирательных бюллетеней в отдаленные и труднодоступные территории. Так, например, в Магаданской, Амурской областях и Ямало-Ненецком автономном округе одними рейсами была осуществлена доставка всей документации и временных переписных работников в отдаленные и труднодоступные местности.

Проанализировав итоги двух кампаний, можно сказать, что в целом удалось обеспечить их организованное проведение. Вместе с тем в ряде субъектов возникали достаточно серьезные проблемы. В первую очередь это сказалось на нехватке помещений (республики Дагестан, Тыва, Ростовская, Кемеровская, Новосибирская, Томская области) для размещения переписных и инструкторских участков, поскольку проведение выборов было более приоритетным мероприятием для местных властей.

Что же касается информационно-разъяснительной работы, то при медиапланировании размещения теле- и радиороликов на региональном телевидении и радио был учтен агитационный период избирательных кампаний и референдума, определенный законодательством Российской Федерации. Основная часть теле- и радиороликов, запланированных на октябрь, была размещена с 9 октября.

На отдельных территориях имела место договоренность с региональными избирательными комиссиями о совместном планировании сроков выпусков в СМИ публикаций, репортажей, проведения мероприятий агитационного и информационно-разъяснительного характера с учетом требований законодательства Российской Федерации при проведении избирательной и переписной кампаний. Осуществлялось вывешивание агитационных материалов на отдельных стендах.

В то же время совмещение мероприятий по информированию отразилось на проведении переписи населения:

- избирательная кампания отвлекла население от переписи. В больших количествах раздавались и распространялись газеты, листовки, плакаты и календари о выборах, в результате информация о переписи практически «тонула» в этом объеме. Две агитационные кампании, проводимые одновременно, вызывали негативное отношение и к переписи населения, и к выборам (Вологодская и Новосибирская области);
- решением органов местного самоуправления агитационные плакаты переписи населения и информационные материалы с указанием адресов и телефонов переписных участков размещались только 11 октября, что явилось причиной снижения уровня информированности населения и привело к обвальному количеству звонков населения с вопросами местонахождения переписных участков (Удмуртская Республика);
- по окончании избирательной кампании вместе с агитационными материалами выборов удалялись и плакаты по переписи населения (Московская и Нижегородская области).

Таким образом, совместные усилия ЦИК России и Росстата способствовали разрешению многих задач и вопросов в подготовке к Всероссийской переписи населения 2010 г. в условиях проведения избирательной кампании, но в последующем необходимо стараться максимально «разводить» во времени перепись населения и выборы.

Всероссийская перепись населения 2010 г. еще раз подтвердила важность информационно-разъяснительной работы и необходимость обратной связи с населением. Помимо мероприятий, предусмотренных общегосударственной Программой проведения информационно-разъяснительной работы, в субъектах Российской Федерации прошло много своих интересных и эффективных с точки зрения формирования позитивного отношения к переписи и обеспечения полноты охвата населения мероприятий.

Необходимо отметить информационную поддержку переписи со стороны федеральных органов исполнительной власти, общественных организаций. Так, Минкомсвязи России выпустило марку с эмблемой Всероссийской переписи населения 2010 г. и оказало содействие в организации рассылки информационного sms-сообщения 65 млн. абонентов сетей «БиЛайн», «МТС», «Мегафон» и «Теле2».

Центральный банк Российской Федерации выпустил в массовое обращение 10-рублевую монету и памятную серебряную монету достоинством 3 рубля с эмблемой Всероссийской переписи населения 2010 г.

При содействии Минспорттуризма России, Российского футбольного союза, Профессиональной футбольной лиги, отдельных футбольных клубов проведены

информационные мероприятия, связанные с переписью, на матчах 24-го тура чемпионата России и матчах Профессиональной футбольной лиги.

Активно тема переписи населения обсуждалась в средствах массовой информации. Всего за апрель-ноябрь 2010 г. по данным мониторинга было опубликовано в разных изданиях, в том числе и электронных, почти 38 тыс. публикаций. При этом за весь указанный период сохранялся позитивно-нейтральный тон этих публикаций; удельный вес негативных публикаций не превысил 3% общего числа.

В настоящее время подведены итоги конкурса, проводимого Медиасоюзом, на лучшее освещение подготовки и проведения переписи в региональных СМИ. Конкурс проводился по федеральным округам. Победителями по пяти номинациям стали журналисты республик Башкортостан, Бурятия, Коми, Карелия, Кабардино-Балкария, Калмыкия, Татарстан, Чувашия, Удмуртия; Алтайского, Краснодарского, Красноярского, Приморского, Ставропольского, Хабаровского краев; Архангельской, Белгородской, Брянской, Владимирской, Ивановской, Курганской, Липецкой, Ленинградской, Московской, Мурманской, Нижегородской, Новгородской, Новосибирской, Пензенской, Псковской, Ростовской, Сахалинской, Самарской, Свердловской, Смоленской, Тамбовской, Тюменской областей, г. Санкт-Петербурга, Ямало-Ненецкого автономного округа. Всего определено 98 победителей (как региональных СМИ, так и отдельных региональных журналистов, опубликовавших свои работы в течение периода январь-октябрь 2010 г.). Приятно отметить, что журналисты ряда приведенных территорий были победителями аналогичного конкурса и при переписи 2002 г.

В целях обратной связи с населением Росстатом на период с 5 сентября по 31 октября 2010 г. была организована «горячая линия» с федеральным номером 8-800-200-14-25. «Горячие линии» были созданы в каждом субъекте Российской Федерации.

С 2009 г. действует информационный сайт переписи www.perepis-2010.ru.

На «горячую линию» с федеральным номером 8-800-200-14-25 поступило почти 200 тыс. звонков из всех субъектов Российской Федерации. Среди звонивших жители Москвы составили 48% от общего количества обращений, Московской области - 17, Санкт-Петербурга - 12, Ленинградской области - 9,2%.

В период проведения Всероссийской переписи населения 2010 г. все дни, кроме 14 октября, когда был запущен фальшивый сайт с доменным именем www.peripis-2010.ru, работал официальный сайт www.perepis-2010.ru. С 1 сентября по 26 октября 2010 г. было более 900 тыс. посещений сайта, просмотрено почти 3,5 млн. страниц, в среднем за посещение просматривалось 3-4 страницы. Посетителями сайта являлись не только жители России (860 тыс.

посещений), но и Украины, Германии, США, Нидерландов, Великобритании, Казахстана, Франции и других стран.

Всего в Росстат поступило от населения свыше 2,5 тыс. обращений (письменные и электронные запросы).

Около 1,8 тыс. писем (70% общего числа обращений) было получено с отказом участвовать в переписи по религиозным соображениям. Адресатами этих писем были Московское представительство ООН, Росстат и территориальный орган Росстата по соответствующему субъекту Российской Федерации.

Оставшиеся обращения условно можно разделить на три группы:

1. *Запросы и жалобы населения* по адресам переписных участков, нарушениям процедуры переписи - задавались не все вопросы переписного листа, переписчик самостоятельно заполнял ответы на вопросы, не дожидаясь ответа опрашиваемого, отмечалось давление переписчика при ответе на вопрос о национальной принадлежности - 6,5%;

2. *Жалобы от переписчиков* по поводу принуждения участвовать в переписи (студенты), низкого уровня оплаты труда и большой нагрузки, неадекватной реакции населения на работу переписчиков, нарушения сроков выплаты вознаграждений - 21,5%;

3. *Предложения по совершенствованию* процедуры переписи, включению дополнительных тем в переписные листы и другие вопросы от населения - 2%.

Все поступившие запросы оперативно рассматривались специалистами центрального аппарата или, по нашему поручению, специалистами наших территориальных органов. Ответы направлялись заявителю.

В настоящее время началась автоматизированная обработка и подведение итогов переписи населения. В соответствии с постановлением Правительства Российской Федерации от 12 ноября 2010 г. № 896 итоги Всероссийской переписи населения 2010 г. подводятся поэтапно:

- предварительные итоги о численности переписанного населения (в том числе мужчин и женщин с разбивкой на городское и сельское население) подводятся на основе сводных ведомостей, составленных временными переписными работниками, до 30 апреля 2011 г.;

- окончательные итоги подводятся на основе автоматизированной обработки сведений о населении, содержащихся в переписных листах, согласно Программе итогов Всероссийской переписи населения 2010 г. до 31 декабря 2012 г.

В 2011-2012 гг. Росстатом будут оперативно размещаться на официальном сайте Росстата получаемые поэтапно данные об итогах Всероссийской переписи населения 2010 г. по всем темам, содержащимся в программе переписи. Доклады об итогах Всероссийской переписи населения 2010 г. будут опубликованы в Рос-

сийской газете. В течение 2012-2013 гг. планируется издать 11 томов официальной публикации итогов Всероссийской переписи населения 2010 г. по стране в целом и субъектам Российской Федерации. Каждым территориальным органом Росстата также будут опубликованы итоги переписи населения соответствующего субъекта Российской Федерации на более подробном территориальном уровне. При этом впервые в практике переписи итоги будут подведены как по административно-территориальным единицам субъектов Российской Федерации, так и по муниципальным образованиям. До настоящего времени статистика населения в разрезе муниципально-территориального деления не разрабатывалась.

Кроме того, планируется выпустить серию брошюр и предоставлять информацию не только на бумажной основе, но и на современных электронных носителях информации. Тома официальной публикации итогов Всероссийской переписи населения 2010 г. будут размещены на сайте Росстата в свободном доступе. По окончании издания тематических томов официальной публикации итогов переписи населения в 2013 г. Росстат обеспечит на своем официальном сайте онлайн доступ к микроданным Всероссийской переписи населения 2010 г. посредством формирования нерегламентированных таблиц, в которых будут применены средства защиты конфиденциальности первичной информации. Эта практика с недавнего времени применяется ведущими статслужбами мира и приобретает все большую популярность среди пользователей, так как дает большие возможности для анализа переписных данных.

* *
*

Проведение переписи населения осуществлялось во взаимодействии и под контролем Комиссии Правительства Российской Федерации по проведению Всероссийской переписи населения 2010 г., комиссиями по проведению Всероссийской переписи населения 2010 г. в федеральных округах, субъектах Российской Федерации, муниципальных образованиях. Большой комплекс работ по переписи был выполнен при слаженной работе и четком взаимодействии федеральных органов исполнительной власти, органов власти субъектов Российской Федерации и органов местного самоуправления.

Таковы самые первые результаты Всероссийской переписи населения 2010 г. Впереди глубокий анализ всех аспектов ее проведения: правового, методологического, организационного, финансового. Должно быть организовано широкое обсуждение не только итогов переписи, но и ее Программы, методов проведения; это необходимо потому, что к переписи населения 2020 г. нужно начинать готовиться уже сейчас.

АВТОМАТИЗИРОВАННАЯ ОБРАБОТКА ПЕРСОНАЛЬНЫХ ДАННЫХ ПЕРЕПИСИ НАСЕЛЕНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ 2009 ГОДА

Е.В. Ермолицкая,

Национальный статистический комитет Республики Беларусь

В период с 14 по 24 октября 2009 г. в Республике Беларусь проведена перепись населения, и уже в феврале 2010 г. опубликованы ее предварительные итоги - данные о численности населения в разрезе Республики Беларусь, областей, районов и городов. Однако об успешном завершении такого крупномасштабного статистического наблюдения, как перепись населения, можно говорить только тогда, когда закончена автоматизированная обработка персональных данных, содержащихся в переписных листах, сформированы итоговые данные, которые могут быть предоставлены Президенту и Правительству Республики Беларусь, органам государственного управления и другим пользователям.

Подготовка к автоматизированной обработке персональных данных началась с момента принятия в июле 2006 г. Закона Республики Беларусь «О переписи населения» [1], где оговорено, что персональные данные, содержащиеся в переписных листах, подвергаются обработке в соответствии с технологией обработки и программой формирования итоговых данных. В октябре 2008 г. была разработана и утверждена постановлением Министерства статистики и анализа Республики Беларусь Технология обработки персональных данных Республики Беларусь 2009 г. (далее - Технология обработки) [2], в которой определены стратегия обработки, ее основные этапы, а также инфраструктура аппаратно-программного комплекса, на базе которого организуется весь процесс обработки. В декабре 2008 г. была разработана и утверждена Программа формирования итоговых данных переписи населения Республики Беларусь (далее - Программа формирования итоговых данных), которая содержит перечень выходных регламентных таблиц по 10 тематическим блокам.

Успешная обработка персональных данных возможна только при обеспечении четкой взаимосвязи работ по трем основным направлениям. *Во-первых*, это методология обработки, которая определяет единицы наблюдения и признаки, по которым осуществляется сбор первичных статистических данных, перечень статистических показателей и группировочных признаков, по которым производится обобщение и представление итоговых данных. Методология обработки определяет методику редактирования персональных данных как при вводе информации в компьютер, так и на других этапах обработки, а также подходы в части формирования и распространения итоговых данных. *Во-вторых*, это выбор стратегии обработки, внедрение необходимых

технологических решений на всех этапах обработки персональных данных и распространения итоговых данных. *В-третьих*, это решение ряда организационных вопросов, касающихся закупки технических и программных средств, проведения работ по созданию и внедрению аппаратно-программного комплекса, найма и обучения персонала по обработке, организации всего процесса обработки.

Методологические аспекты автоматизированной обработки персональных данных

Основополагающим элементом, определяющим методологию обработки персональных данных, являются формы переписных листов [2], поскольку они содержат вопросы программы наблюдения и, следовательно, определяют специфику статистических данных, подлежащих обработке, а также состав информационного ресурса переписи населения.

При переписи населения Республики Беларусь 2009 г. было выделено три типа единиц наблюдения: респонденты (или лица), домашние хозяйства, помещения. В отношении лиц информация собиралась по следующим признакам: пол, возраст, состояние в браке, национальность и языки, миграционные признаки, страна рождения и гражданство, уровень образования, источники средств существования, экономическая активность, вид занятия респондента, а также вид экономической деятельности организации, в которой занят респондент, число рожденных детей у женщин.

С целью получения информации по домашним хозяйствам формы переписных листов содержали вопросы о родственных отношениях лиц в пределах данного домохозяйства, а также вопросы, касающиеся жилищных условий членов домохозяйства. По помещениям собиралась информация о типе, материале наружных стен и периоде постройки помещения, его видах благоустройства и размере общей площади. Для целей переписи населения были разработаны формы переписного листа, основными из которых являются:

форма 6 «Сопроводительный бланк» (далее - форма 6);

форма 1-помещение «Характеристика жилого помещения, список членов домохозяйства, постоянно проживающих в помещении» (далее - форма 1-помещение);

форма 2Н «Переписной лист» (далее - форма 2Н).

При разработке форм переписного листа необходимо было учитывать специфику выбранной технологии обработки, а также методологию кодирования отдельных вопросов переписного листа. Кроме того, формы переписного листа разных типов должны быть взаимосвязаны между собой, чтобы в результате формирования информационного ресурса переписи населения каждый респондент однозначно был отнесен к одному домашнему хозяйству, представляющему собой совокупность лиц, проживающих только в одном помещении. Программа наблюдения содержала также ряд географических признаков, позволяющих отнести каждую единицу наблюдения к конкретной административно-территориальной единице (далее АТЕ), а также адресу в пределах АТЕ.

Согласно традиционным подходам организации и проведения переписей населения, в период подготовки к переписи в целях обеспечения полноты охвата территории переписью, а также расчета потребности во временном персонале по проведению опроса населения был проведен ряд работ по разбиению территории государства на переписные, инструкторские и счетные участки - переписное районирование. В результате по каждой АТЕ (в данном случае это административный район, город областного подчинения, район в городе) был составлен организационный план проведения переписи населения (далее - оргплан). Переписное районирование производится при помощи переписных карт, которые в период подготовительных мероприятий по проведению переписи населения были изготовлены подрядными организациями Государственного комитета по имуществу Республики Беларусь. К сожалению, при подготовке к переписи населения Республики Беларусь 2009 г. не использовались возможности географических информационных систем (далее - ГИС), и переписные карты были предоставлены на бумажном носителе.

Вместе с тем было принято решение автоматизировать процесс составления оргпланов. Для ввода результатов переписного районирования в компьютер был разработан программный комплекс «Оргплан». В результате была организована база данных, содержащая перечень переписных, инструкторских, счетных участков по сельской местности в разрезе населенных пунктов. По городским поселениям и крупным сельским населенным пунктам (с численностью населения более 3000 человек) база данных включала информацию как о перечне участков, так и о жилых и нежилых строениях, входящих в конкретный счетный участок. Каждое жилое или нежилое строение в базе данных было отнесено к АТЕ, на территории которого оно находится, к кварталу в пределах АТЕ, а также имело формализованную адресную часть - код элемента дорожно-транспортной сети (улица, переулок и т. д.), номер дома, номер корпуса.

Результаты переписного районирования в электронном формате были также использованы при выполнении работ по подготовке портфелей с заполненными формами переписного листа к автоматизированной обработке. Эти работы проводились по завершении переписи населения специалистами главных статистических управлений областей и г. Минска (далее - областной уровень) согласно Указаниям по подготовке портфелей с формами переписного листа к автоматизированной обработке (далее - Указания по подготовке портфелей). По мере поступления портфелей с полевых работ формировались массивы портфелей. Состав каждого массива, а именно перечень АТЕ, определялся Указаниями по подготовке портфелей; при этом размер массива в среднем составил около 700 портфелей. Подготовка портфелей в пределах данного массива включала следующие этапы:

- укладка форм переписного листа в таком порядке: первой в портфеле располагалась форма 6, затем укладывалась первая по счетному участку форма 1-помещение в порядке возрастания номеров домашних хозяйств, потом формы 2Н по данному помещению и т. д. в порядке следования форм 1-помещение. Если портфель содержал переписные листы нескольких населенных пунктов, то перед переписными листами каждого из них располагалась форма 6;

- подсчет по данному портфелю (счетному участку) в разрезе населенных пунктов количества форм переписного листа каждого типа и внесение этой информации в контрольную ведомость. Контрольная ведомость - документ в электронном формате, сопровождающий массив на автоматизированную обработку и содержащий перечень портфелей (счетных участков) в разрезе населенных пунктов, а также номер портфеля в пределах массива. Контрольная ведомость формируется в период подготовительных работ на основании данных переписного районирования;

- расположение наклейки со специальным штрих-кодом на боковой стороне портфеля, а также на лицевой стороне формы 6. Штрих-код содержит следующую информацию: код населенного пункта по Системе обозначений административно-территориальных образований Республики Беларусь (далее - СОАТО), номера переписного, инструкторского и счетного участков. Если в состав счетного участка входило несколько населенных пунктов, то необходимо было расположить несколько наклеек. Наклейки были изготовлены в период подготовительных работ, штрих-коды разработаны также на основании данных переписного районирования;

- нанесение красным карандашом на ярлыке портфеля номера портфеля в пределах массива.

По мере готовности портфелей по данному массиву контрольная ведомость по линиям связи передавалась в Белстат, где производилась ее загрузка в базу дан-

ных складского учета, а массив портфелей отправлялся на обработку. Необходимо подчеркнуть, что поступление массивов из областного уровня в Белстат было строго регламентировано и производилось в соответствии со специальным графиком.

Наличие результатов переписного районирования в электронном формате, организация базы данных складского учета на основании контрольных ведомостей, использование наклеек со штрих-кодами позволили:

- полностью автоматизировать процесс формирования и заполнения сводных таблиц оргплана, облегчить работы по его корректировке;
- исключить ошибки при подготовке портфелей к автоматизированной обработке и формировании массивов;
- автоматизировать работы по складскому учету при поступлении портфелей с формами переписных листов на автоматизированную обработку;
- обеспечить контроль за полнотой охвата обработкой всех АТЕ, счетных участков, форм переписного листа по каждому счетному участку в разрезе населенных пунктов;
- использовать информацию базы данных оргплана о жилых и нежилых строениях (в частности, адресную часть) при выполнении работ по редактированию таких вопросов переписного листа, как наименование улицы, номер дома, номер корпуса, а также максимально автоматизировать эти работы;
- подключить информацию базы данных оргплана о кварталах к базе данных переписи населения с целью использования ее при распространении итогов переписи населения при помощи ГИС.

К методологическим аспектам переписи населения относятся также методологические решения в части кодирования и редактирования персональных данных, выбор которых во многом определяется технологическими решениями. Согласно Технологии обработки, для ввода персональных данных использовалась технология поточного сканирования и интеллектуального распознавания символов и меток, содержащихся в формах переписного листа. Специфика этой технологии заключена в том, что процесс ввода практически полностью автоматизирован, ответы на вопросы переписного листа, заполненные в виде меток, автоматически кодировались. Если ответ был записан в специально отведенных для этого полях в виде наименований, то такой ответ подвергался процедуре распознавания и автоматического кодирования. Для целей автоматического кодирования были разработаны кодификаторы национальностей, языков, стран мира, территорий Республики Беларусь, наименований улиц.

Редактирование персональных данных было разбито на два этапа. *Первый этап редактирования* производился при вводе персональных данных; при этом нужно было обеспечить:

- необходимую точность ввода информации, то есть введенная информация должна максимально соответствовать данным, содержащимся в формах переписного листа;

- контроль полноты введенной информации, а именно выполнения следующих требований: АТЕ, а также переписной, инструкторский и счетный участки должны соответствовать данным переписного районирования; все формы переписного листа по данному счетному участку (как правило, портфелю) в разрезе населенных пунктов должны быть введены в компьютер; все формы 1-помещение должны принадлежать данному счетному участку; по форме 1-помещение должно быть переписано хотя бы одно домашнее хозяйство; все формы 2Н должны быть отнесены к конкретному домашнему хозяйству (далее - редактирование структуры);

- правильность заполнения ответов на вопросы формы 2Н о возрасте, поле, родственных отношениях, состоянии в браке (далее - редактирование семейных связей), что обеспечило правильность формирования домашнего хозяйства уже на этапе ввода и корректную выгрузку введенной информации в базу данных переписи населения;

- правильность кодирования ответов на вопросы форм переписного листа (далее - редактирование другое).

Для обеспечения вышеуказанных требований была разработана система логических контролей и редактирования, которая включала более 400 проверок. Решения по исправлению ошибок, принимаемые на данном этапе редактирования, были максимально упрощены в целях наиболее быстрого и эффективного ввода информации.

Второй этап редактирования производился уже при формировании базы данных переписи населения и обеспечивал логическую взаимосвязку ответов респондентов на вопросы форм переписного листа. Для реализации этого этапа редактирования была разработана система логических контролей, содержащая около 600 видов контролей, а для исправления некоторых логических ошибок - алгоритмы автоматического редактирования.

Форма 2Н содержала вопросы о занятии респондента, наименовании организации, в которой респондент занят, и о виде ее экономической деятельности. Ответы на эти вопросы заполнялись в произвольной форме, поэтому не могли быть закодированы автоматически, и их кодирование производилось в полуавтоматическом режиме. Для этих целей были разработаны методика кодирования, кодификаторы занятий и видов экономической деятельности. При кодировании использовался Регистр предприятий и организаций Республики Беларусь, что позволило автоматизировать поиск в нем организации, в которой занят респондент, и облег-

читать процесс принятия решения по определению ее вида экономической деятельности.

Основная цель любого государственного статистического наблюдения - получение и предоставление пользователям итоговых данных, полученных в результате обработки и группировки первичной статистической информации. В период подготовки переписи населения были выработаны методологические подходы формирования информационного ресурса переписи населения, а также решения в части доступа пользователей к итоговым данным. Согласно Программе формирования итоговых данных, сводная информация должна быть представлена в виде выходных регламентных таблиц по следующим АТЕ: республика, области (г. Минск), административные районы в разрезе городских поселений, города областного подчинения в разрезе внутригородских районов. Таблицы должны быть получены как на бумажных носителях, так и в электронном виде для распространения на электронных носителях информации и размещения в сети Интернет.

Альтернативным способом получения итоговых данных является доступ к сводной информации по нерегламентным запросам. Для реализации такой возможности была разработана система статистических показателей и группировочных признаков, которая легла в основу методологических решений по формированию базы агрегированных данных. Такой подход позволил:

- формировать итоговые данные в любых тематических разрезах, как предусмотренных Программой формирования, так и выходящих за ее рамки;
- получать информацию на уровне любой АТЕ, вплоть до каждого населенного пункта;
- разработать и внедрить интеллектуальную систему поддержки принятия решений на базе OLAP-технологий, обеспечив при этом доступ к аналитической информации пользователя-статистика, не являющегося специалистом в области информационных технологий.

Стратегия и технология автоматизированной обработки персональных данных

Ввод персональных данных осуществлялся централизованно - на территории Национального статистического комитета Республики Беларусь (далее - Белстат), с использованием технологии поточного сканирования и интеллектуального распознавания символов. Централизация процесса обработки позволила минимизировать бюджетные затраты на закупку сканерного и серверного оборудования, лицензионного программного обеспечения, а также обеспечить конфиденциальность персональных данных, содержащихся в переписных листах на всех этапах обработки, и в максимальной степени применить политику информационной безопасности. Вместе с тем при этом обеспечивалось применение единых методологических подходов в части редактирования персональных данных.

Согласно выбранной стратегии обработки, на территории Белстата был создан центр по обработке, где на базе локальной вычислительной сети разработана и введена в эксплуатацию автоматизированная система переписи населения. Создание системы включало следующие этапы:

- закупка сканерного и компьютерного оборудования, лицензионного программного обеспечения, организация локальной вычислительной сети (состав технических и программных средств определен в Технологии обработки);
- разработка и внедрение Системы ввода и обработки персональных данных переписи населения Республики Беларусь 2009 г.;
- разработка и внедрение Системы переписи населения Республики Беларусь 2009 г. (СПН).

Система ввода и обработки персональных данных переписи населения Республики Беларусь 2009 г.

Разработчиком Системы ввода и обработки персональных данных переписи населения Республики Беларусь 2009 г. (далее - Система ввода) была выбрана израильская компания *Top Image Systems inc.* Решение было принято в результате оценки конкурсных предложений в рамках международного конкурса. Предложение данной компании оказалось наиболее приемлемым как по функциональным характеристикам системы, так и по цене; при этом был принят во внимание обширный опыт компании в области обработки данных переписей во многих странах мира. Разработка, тестирование и отладка Системы ввода производились на территории Республики Беларусь на технической базе Белстата, что позволило существенно оптимизировать и ускорить процесс внедрения системы в эксплуатацию, а также максимально обеспечить информационную безопасность.

Система ввода включала три функциональных блока: модуль автоматизации складского учета, модуль обработки персональных данных (на базе платформы *eFlow*), модуль кодирования занятий и видов экономической деятельности.

Модуль автоматизации складского учета предназначен для регистрации поступления портфелей с формами переписного листа с областного уровня в Белстат на обработку, выдачи портфелей со склада на участок сканирования, возврата портфелей на склад после сканирования, отправки портфелей, прошедших обработку, обратно на областной уровень. Портфель в базе данных мог иметь следующие статусы: «неизвестно», «поступил», «выдан на сканирование», «возвращен после сканирования», «отправлен в область».

Процесс регистрации заключался в считывании при помощи ручного сканера штрих-кода, расположенного на боковой стороне портфеля, поиске этой информации в базе данных складского учета и установке порт-

фелю одного из статусов; при этом оператору выдавалось соответствующее сообщение. Как было указано ранее, база данных складского учета заполнялась на основании данных контрольных ведомостей, поступающих с областного уровня. При помощи модуля автоматизации складского учета можно было получить информацию о любом портфеле и его статусе на складе.

Модуль обработки персональных данных - основ-

ной элемент Системы ввода, предназначенный для сканирования форм переписного листа, распознавания символов и меток, содержащихся в них, логического контроля, автоматического кодирования и редактирования персональных данных, выгрузки информации в базу данных переписи населения. На рис. 1 представлена обобщенная схема процесса ввода персональных данных.

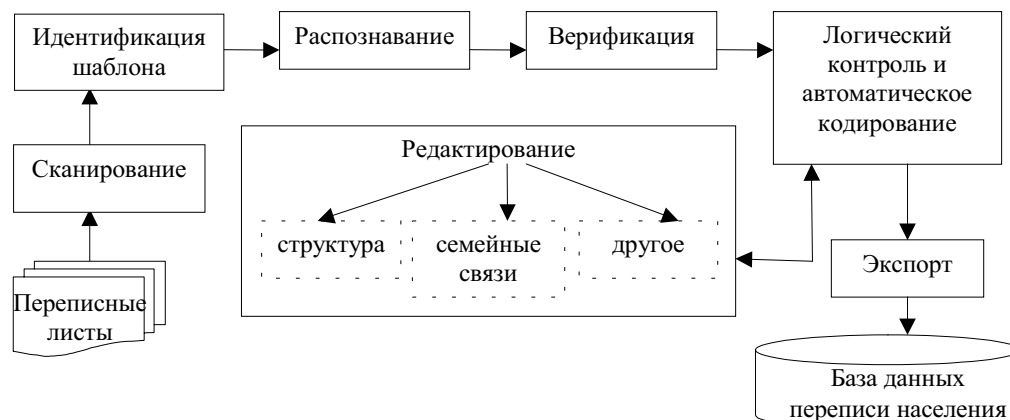


Рис. 1. Обобщенная схема процесса ввода

На этапе сканирования производилось считывание форм переписного листа с целью получения их изображений в электронном формате, а именно образов форм переписного листа. На этом же этапе по технологии цветоистирания (color drop out) из образов форм переписного листа происходило автоматическое удаление «лишней информации» зеленого цвета - шаблона. Шаблон - это набор полей данного типа формы переписного листа, предназначенных для заполнения респондентом и соответственно подвергаемых обработке. После сканирования и преобразования содержимого данного портфеля в электронный формат формировался так называемый *пакет*, который содержал информацию по данному счетному участку и имел в системе уникальный идентификатор.

Следующим этапом обработки пакета являлась автоматическая идентификация шаблонов образов форм переписного листа. Это необходимо для того, чтобы определить, к какому типу формы переписного листа отнести данный образ. Для этого использовалась информация, находящаяся в нижней части формы в виде штрих-кода, нанесенного типографским способом. Если шаблон не мог быть определен автоматически, то он определялся оператором в ручном режиме.

На этапе распознавания производилось автоматическое распознавание меток и символов, содержащихся в образах форм переписного листа. Ответы респондентов, содержащие нераспознанные символы, поступали на этап верификации, на котором оператору предоставлялась возможность внести соответствующие исправ-

ления. На экране компьютера отображалась часть образа переписного листа, содержащая ответ респондента, а также поле для корректировки, в котором нераспознанный символ был специальным образом выделен. Оператору были доступны кодификаторы национальностей, языков, стран мира, территорий Республики Беларусь, улиц, что помогало принять решение и ускорило процесс корректировки.

С целью повышения точности распознавания проводилась повторная верификация - контроль качества. Количество пакетов, поступающих на повторную верификацию, можно было менять в настройках программного обеспечения в зависимости от загрузки системы. Результаты первого и второго этапов верификации сравнивались, и если они не совпадали, то окончательное решение принимал эксперт - оператор высшего уровня.

После верификации содержимое пакета подвергалось логическому контролю и автоматическому кодированию. Если обнаруживались ошибки или ответ респондента не был закодирован автоматически, то такой пакет отправлялся на этап редактирования. Первыми исправлялись ошибки структуры, затем ошибки в семейных связях, потом прочие ошибки. Соответственно было выделено три модуля редактирования (структура, семейные связи и др.), которые были аналогичны по своей функциональности, но отличались типами ошибок, для исправления которых были предназначены. Пакет не мог быть отправлен из модуля редактирования, пока все ошибки не были исправлены.

При этом редактор имел возможность:

- видеть на экране дерево образов форм переписного листа, принадлежащих данному пакету, и при необходимости просматривать любую из них;
- в окне ошибок видеть все ошибки данного типа в пределах пакета;
- при выборе конкретной ошибки просмотреть сообщение об ошибке, а также получить справку о действиях редактора по ее исправлению;
- просматривать образ формы переписного листа, где поле, содержащее ошибку, выделялось соответствующим образом; при этом в качестве подсказки ему отображалась информация, логически связанная с данным полем (например, если респондент указывал номер матери (отца), под которым в списке обозначено лицо в возрасте моложе 15 лет, то в качестве подсказки отображался образ формы 2Н, заполненной на это лицо);
- вносить исправления в специально предназначенные для этого поля;
- пользоваться кодификаторами и справочными пособиями в электронном формате;
- «отключить» ошибку, то есть не исправлять ее (например, если указана национальность, не найденная в кодификаторе);
- отправить ошибку эксперту - редактору высшего уровня, который примет окончательное решение (например, число лет противоречит дате рождения, а по другой информации невозможно установить правильный ответ).

После исправления ошибок пакет снова отправлялся на этап логического контроля и автоматического кодирования. Если все ошибки были исправлены, то пакет подвергался процедуре экспорта с целью загрузки персональных данных в базу данных переписи населения. При экспорте формировался XML-файл, содержащий информацию по данному пакету; при этом выгрузке подлежали также образы форм переписного листа. На данном этапе подготавливалась информация для кодирования занятий и видов экономической деятельности, а именно по конкретной форме 2Н «вырезалась» часть образа, содержащая ответы респондента на соответствующие вопросы, а в формат XML выгружалась вспомогательная информация (возраст, образование, место работы и т. д.). В результате по данному пакету формировалась совокупность файлов-образов и XML-файлов, которая в виде папки с файлами передавалась на областной уровень для кодирования.

Отдельно необходимо оговорить специфику интеграции двух систем - Системы ввода и СПН. На этапе разработки обеих систем был утвержден протокол обмена информацией, а также внедрена система специальных контролей, что позволило исключить поступление в базу данных переписи населения некорректной информации (повторная загрузка пакета, попытка загрузить данные по несуществующему счетному участ-

ку и т. д.). В случае возникновения ошибки пакет возвращался обратно в Систему ввода с пометкой «не прошел экспорт»; при этом оператору выдавалось сообщение об ошибке, а также реквизиты формы переписного листа, в которой обнаружена ошибка.

Для реализации этапов обработки, обозначенных на рис. 1, были разработаны соответствующие программные модули - станции, как автоматические (идентификация шаблона, распознавание, логический контроль и автоматическое кодирование, экспорт), так и предназначенные для работы операторов (сканирование, ручная идентификация шаблона, верификация, редактирование). Оператор мог активировать любую из станций (автоматические станции не были доступны операторам) из специального приложения - *активатора модуля*.

Необходимо отметить высокие возможности модуля обработки персональных данных в части администрирования всего процесса обработки. Разработчиком был предложен специальный программный модуль - *контроллер*, использование которого позволило администратору:

- видеть на экране компьютера в виде специальных ярлычков все типы станций, при нажатии кнопкой мыши на соответствующий ярлык войти в станцию;
- увидеть перечень пользователей, работающих с данной станцией, а также перечень пакетов, находящихся на ней в обработке, просмотреть образы форм переписного листа конкретного пакета, проанализировать динамику его обработки;
- менять приоритеты пакетов, удерживать и удалять пакеты из обработки;
- найти пакет в системе, а именно определить, на каком этапе обработки он находится;
- менять количество автоматических станций каждого типа и тем самым регулировать информационные потоки с целью исключения «узких» мест.

Модуль кодирования занятий и видов экономической деятельности был разработан как отдельное приложение, которое устанавливалось на областном уровне, куда информация, предназначенная для кодирования, поступала из Белстата согласно специально разработанному графику. Результаты кодирования передавались обратно в Белстат, где производилась загрузка их в базу данных переписи населения.

Система переписи населения Республики Беларусь 2009 г. Разработчиком Системы переписи населения Республики Беларусь 2009 г. является белорусская компания *CompIT Developer Systems*. Выбор был определен обширным опытом компании в разработке информационных систем, предназначенных для эффективной обработки больших объемов информации, в том числе на базе программных продуктов *Oracle*. Кроме того, предложенный компанией вариант реализации

системы оказался наиболее приемлемым как по технологическим, так и функциональным характеристикам.

СПН предназначена для импорта информации из Системы ввода, формирования базы данных переписи населения и электронного архива образов форм

переписного листа, логического контроля и редактирования персональных данных, формирования базы агрегированных данных, получения итоговых данных. Функциональная структура СПН представлена на рис. 2.

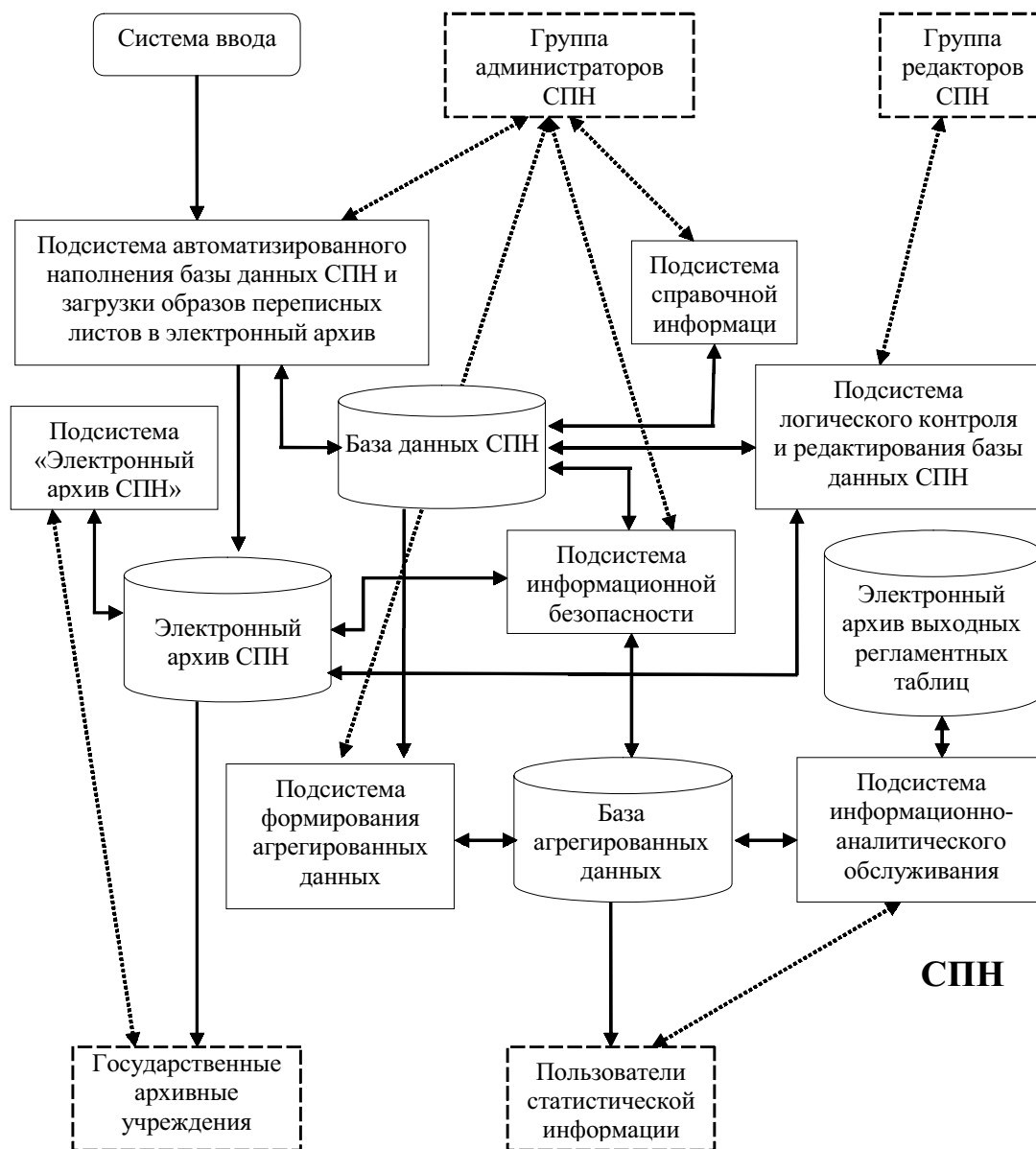


Рис. 2. Функциональная структура СПН

Подсистема автоматизированного наполнения базы данных СПН и загрузки образов переписных листов в электронный архив (далее - Подсистема загрузки) осуществляла автоматическую загрузку персональных данных, поступающих из Системы ввода, в базу данных СПН. Единицей загрузки являлся пакет, данные передавались в XML-формате. Одновременно производилась загрузка образов форм переписного листа в электронный архив; при этом обеспечивалась увязка

информации конкретной формы переписного листа и ее образа. В рамках взаимодействия Подсистемы загрузки и Системы ввода разработчику Системы ввода был предоставлен интерфейс, который соединен с сервисом на сервере СПН. Сервис проводил проверку данных на основе XSD схемы и в случае отсутствия ошибок загружал их в базу данных переписи населения (далее - база данных СПН). В случае неуспешного выполнения загрузки (например, неверный формат дан-

ных), код ошибки, ее текстовое описание и местоположение в пакете передавались в Систему ввода для исправления.

Пользователем подсистемы загрузки мог быть любой сотрудник управления переписи населения, наделенный соответствующими правами доступа; при этом он имел возможность по выбранной АТЕ:

- просмотреть, сколько пакетов должно быть загружено и сколько загружено фактически (в разрезе форм переписного листа каждого типа);
- просмотреть протокол загрузки, а именно перечень пакетов с их статусами (загружен или ошибка) и датой загрузки;
- по выбранному пакету просмотреть ошибки загрузки.

Функцией Подсистемы загрузки являлся также импорт результатов кодирования занятий и видов экономической деятельности, которые поступали из областного уровня в Белстат. Данная функция запускается автоматически по мере поступления информации с результатами кодирования, организованной в XML-формате.

База данных СПН содержит обезличенные персональные данные [1] и является основным информационным элементом СПН. Логическая структура базы данных СПН соответствует реляционной модели данных, поскольку база организована в среде Oracle Database, являющейся RDBMS. База данных СПН относится к типу VLDB (very large database) и требует больших объемов дискового пространства. С целью упрощения администрирования и оптимизации операций ввода-вывода табличные пространства были разбиты на функциональные группы.

Для повышения управляемости, производительности и надежности приложений СПН применена опция Oracle Partitioning, которая позволяет разбивать таблицы на части, называемые секциями. Оптимизация работы в секционированной таблице достигается за счет:

- отсечения секций (Partition pruning) - при выполнении запроса просматривается только необходимая часть таблицы (выбранная секция), что повышает производительность в несколько раз;
- соединение по секциям (Partition-wise-joins) - разбивает большие соединения на маленькие соединения между каждой из пар секций, что позволяет осуществить полное соединение за более короткое время.

Механизм секционирования применим и к индексам. В базе данных СПН секционированию подлежат индексы секционированных таблиц по типу локальных секционированных индексов.

Подсистема справочной информации предназначена для информационного обеспечения входящих в состав СПН подсистем и обеспечивает:

- загрузку в базу данных СПН справочников и кодификаторов, используемых для кодирования ответов

на вопросы форм переписного листа. Справочники и кодификаторы были представлены разработчику специалистами Белстата в электронном формате;

- просмотр пользователем справочников и кодификаторов, поиск нужной информации, при необходимости - внесение корректировок;
- формирование и ведение служебных кодификаторов, используемых при функционировании СПН.

Подсистема информационной безопасности реализует функции идентификации, аутентификации и авторизации, функции управления доступом и аудита.

Подсистема «Электронный архив СПН» предназначена для обеспечения выгрузки образов форм переписного листа из электронного архива СПН и данных переписного районирования для передачи на постоянное хранение в государственные архивные учреждения, протоколирования событий и объектов выгрузки. Наименьшей единицей выгрузки является пакет. При работе с данной подсистемой пользователю предоставлены следующие возможности:

- просмотреть в разрезе АТЕ: сколько всего пакетов, сколько выгружено пакетов и форм переписного листа каждого типа, выбрать объект выгрузки, выгрузить информацию и просмотреть протокол выгрузки;
- в разрезе счетных участков увидеть статус пакета (выгружен/не выгружен), выбрать объект выгрузки, выгрузить информацию и просмотреть протокол выгрузки;
- получить отчет в формате Excel - ведомость объектов выгрузки, которая содержит перечень выгруженных образов форм переписного листа и соответствующие им имена GIF-файлов;
- выгрузить данные переписного районирования (наименование АТЕ и ее код по СОАТО, номера переписных, инструкторских и счетных участков) в формате Excel и просмотреть протокол выгрузки.

Подсистема логического контроля и редактирования базы данных СПН (далее - Подсистема логического контроля и редактирования) предназначена для выявления и устранения логических несоответствий, содержащихся в персональных данных с целью обеспечения их логической целостности и непротиворечивости.

Пользователь-редактор имел в своем распоряжении удобный инструмент, позволяющий в пределах выбранной АТЕ:

- видеть количество ошибок в разрезе типов форм переписного листа;
- по данной ошибке просмотреть количество записей, обратиться к любой из них и проанализировать данные, просмотреть образ формы переписного листа по данной записи, внести, если это необходимо, соответствующие исправления. Редактору предоставлялась возможность обращаться к персональным данным, образам других форм, заполненных в пределах

данного помещения, и вносить необходимые исправления;

- по данной ошибке запускать процедуру автоматического редактирования; при этом автоматически редактировались все записи с данной ошибкой в пределах выбранной АТЕ;

- производить поиск записей по заданным критериям; при этом выбранные записи могли быть отредактированы в интерактивном режиме.

Необходимо отметить высокие возможности Подсистемы логического контроля и редактирования в части настроек правил логического контроля и алгоритмов редактирования, что обеспечило максимальную эффективность при работе с большими объемами информации. Разработчиком СПН был предложен удобный инструмент, который позволял пользователю, наделенному определенными правами доступа:

- выбирать АТЕ, в пределах которой необходимо проверить данные, а также включать/отключать отдельные правила логического контроля и тем самым регулировать временные издержки на процедуры логического контроля;

- без помощи специалистов разработчика менять часть программного кода процедуры логического контроля;

- включать/отключать возможность автоматического редактирования по данному правилу логического контроля;

- блокировать возможность редактирования персональных данных по выбранным переменным, что обеспечило логическую последовательность всего процесса редактирования (например, изначально исправлялись ошибки возраста респондента, затем переменные «возраст» и «число исполнившихся лет» блокировались, а другие переменные исправлялись уже по отношению к ним и т. д.).

База агрегированных данных - база данных, структура которой оптимизирована для выполнения задач данной предметной области. При ее разработке использовалась пространственная модель (dimensional model), которая описывает данные с помощью измерений и фактов. База агрегированных данных состоит из набора звездообразных схем, каждая из которых включает в себя центральную таблицу фактов, окруженную таблицами измерений. Проектирование базы агрегированных данных осуществлялось посредством использования Oracle Warehouse Builder. При описании предметной области за основу был взят принцип структуризации статистического показателя на основании и признаковую часть [3]. Перечень показателей и признаков был разработан специалистами Белстата и предоставлен разработчику.

Подсистема формирования агрегированных данных предназначена для извлечения данных из базы данных СПН, группировки и приведения их к формату, под-

держиваемому моделью базы агрегированных данных, расчета статистических показателей и загрузку их в базу агрегированных данных. Реализация функций данной подсистемы основана на использовании компонентов программного продукта Oracle Business Intelligence. Формирование агрегированных данных выполнялось пользователями, которым предоставлены соответствующие полномочия.

Подсистема информационно-аналитического обслуживания предназначена для формирования выходных регламентных таблиц, организации электронного архива выходных регламентных таблиц и получения информации по нерегламентным запросам пользователей. Источником информации для формирования как выходных регламентных таблиц, так и итоговых данных нерегламентированного характера является база агрегированных данных, что существенно снижает время выполнения запроса.

При работе с подсистемой информационно-аналитического обслуживания пользователь имеет возможность:

- в пределах данной АТЕ выбрать из списка нужную таблицу, в необходимых тематических разрезах сформировать ее и экспортировать в форматы MS Excel, MS Word; при этом происходит автоматическое сохранение сформированного отчета в электронном архиве;

- быстрого поиска и извлечения выходной таблицы из электронного архива, ее просмотра и выдачи на печать;

- упаковки сформированного отчета (отчетов) в архив (например, WinRAR) с целью передачи информации сторонним пользователям.

Реализация функций получения информации по нерегламентным запросам основана на использовании Oracle BI Answers. При помощи удобного интерфейса пользователь имеет возможность в рамках тематического блока (например, демографические характеристики населения) выбрать статистический показатель (например, численность населения), определить разрезы, в которых необходимо данный показатель отразить (например, мужчины в возрасте до 24 лет, никогда не состоявшие в браке, по сельским населенным пунктам Витебской области) и получить таблицу (плоскую или многомерную), в которой будет отображена информация по запросу. Результат запроса может быть выдан на печать, экспортирован в формат MS Excel.

Организация базы агрегированных данных и использование для работы с ней компонентов программного продукта Oracle Business Intelligence открывает неограниченные возможности в плане доступа к информационному ресурсу переписи населения, среди которых в первую очередь необходимо отметить:

- оперативное формирование итоговых данных переписи населения в различных тематических разрезах,

Таблица

Численность персонала по обработке
(человек)

	Численность персонала по обработке
Всего	91
в том числе:	
начальник смены	2
редактор-эксперт	2
редактор-верификатор	72
оператор сканирования	8
кладовщик	2
грузчик	4
техник по обслуживанию средств вычислительной техники	1

построения и оформления статистических таблиц, что позволяет в максимально сжатый срок подготовить и опубликовать статистические бюллетени и сборники по итогам переписи населения;

- легкий доступ к базе агрегированных данных аналитика, не являющегося специалистом в области информационных технологий, что существенно сокращает время на подготовку информации по запросам сторонних пользователей (органов государственного управления, организаций, частных лиц);

- возможность получения итоговых данных в любом территориальном разрезе (вплоть до населенного пункта) с целью интеграции данных переписи населения с *географической информационной системой* для их распространения в виде тематических карт.

Организационные аспекты автоматизированной обработки персональных данных

К организационным аспектам автоматизированной обработки относятся:

- закупка технических и программных средств, организация и проведение конкурсов на выбор разработчика прикладного программного обеспечения, оценка конкурсных предложений, заключение договоров на выполнение работ и услуг;

- организация взаимодействия с разработчиками прикладного программного обеспечения в части тестирования информационной системы, приемки ее в опытную и промышленную эксплуатацию, определение условий технического сопровождения;

- создание вычислительного центра по обработке, внедрение в эксплуатацию аппаратно-программного комплекса на базе локальной вычислительной сети;

- разработка стратегии взаимодействия Белстата с главными статистическими управлениями областей и г. Минска в плане передачи портфелей на автоматизированную обработку, а также с точки зрения организации процесса кодирования видов экономической деятельности и занятий и многое другое.

Однако наиболее сложным и трудоемким организационным моментом является определение количества персонала по обработке, его наем, обучение и организация труда. Работы по обработке персональных данных включают два этапа.

Первый этап связан с вводом персональных данных в компьютер и реализован в рамках функционирования Системы ввода. Согласно Технологии обработки, в течение шести месяцев необходимо было обеспечить ввод информации с порядка 14 млн. форм переписного листа. Исходя из этого, а также из производительности Системы ввода, была рассчитана численность персонала по обработке, которая в разрезе функциональных обязанностей приведена в таблице.

Отбор персонала проводился на конкурсной основе. Согласно условиям заключаемых гражданско-правовых договоров, претендент должен был пройти курс обучения и выдержать специальный тест. С целью обучения и тестирования персонала специалистами Белстата были разработаны специальные программы обучения и программы тестирования. Обучение было организовано по группам, каждая из которых обучалась не более двух дней. С претендентами, успешно прошедшими тесты, заключались договоры.

Ввод персональных данных начался 5 января 2010 г. и был завершен к концу июня 2010 г., работа персонала по обработке была организована в две рабочие смены по 6 часов каждая.

Второй этап обработки персональных данных связан с редактированием базы данных СПН, получением выходных регламентных таблиц. С целью реализации этого этапа работ с 1 июля 2010 г. на условиях гражданско-правовых договоров привлечено 15 редакторов.

В августе 2010 г. основные итоги переписи населения Республики Беларусь 2009 г., а именно данные о численности населения страны, его характеристиках по полу и возрасту, уровню образования, национальности и языках, были предоставлены Президенту, а 1 сентября 2010 г. опубликованы в виде статистического бюллетеня.

Литература

1. О переписи населения: Закон Республики Беларусь, 13 июля 2006 г., № 144-З // Нац. реестр правовых актов Респ. Беларусь. 2006. № 122, 2/1241.
2. Национальный статистический комитет Республики Беларусь [Электронный ресурс]. 2010. Режим доступа: www.belstat.gov.by/hoper/ru/perepic/main_new.php.
3. Методологические положения по формированию и ведению каталога статистических показателей (КСП). - М.: Госкомстат России, 2000. - 75 с.