

О РАЗВИТИИ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ РОССТАТА

Р.И. Калина, канд. экон. наук,
Федеральная служба государственной статистики

Система государственной статистики в Российской Федерации всегда ориентировалась на максимально активное и эффективное использование информационных технологий на всех этапах производства и распространения статистической информации. Подтверждением этому факту является то, что каждый 12-й сотрудник системы государственной статистики работает в подразделениях, занимающихся информатизацией работы системы Росстата, а наиболее масштабные и ресурсоемкие проекты в системе статистики целевой функцией имеют повышение уровня информационной оснащенности нашего ведомства.

Наиболее значимым из таких проектов в начале 2000-х годов стал совместный с Всемирным банком проект развития системы государственной статистики (РСГС), заложивший основы новой структуры информационного пространства и технической оснащенности системы статистики. Современные компьютеры и серверы, вычислительные сети и программные продукты, новые информационные технологии массово пришли в нашу систему именно тогда.

Одним из ключевых результатов РСГС стало создание проекта новой информационно-вычислительной системы (ИВС) Росстата. Утвержденный еще в начале 2000-х годов, проект содержал в себе самые современные подходы к построению информационных систем, которые будут актуальны еще не один год.

Основное внимание в проекте было уделено созданию эффективной системы сбора и обработки статистической информации. Отметим цели, достичь которые представлялось особенно важным создателям проекта:

1. Централизация процессов сбора и обработки информации. Потребности системы статистики не позволяют ограничиться поднятием на федеральный уровень только агрегированных статистических данных по субъектам Российской Федерации. Система должна позволять работать с первичными статистическими данными на любом уровне;

2. Первичные и агрегированные статистические данные на федеральном и региональном уровнях должны размещаться в хранилищах данных (едином на федеральном уровне, и своем для каждого ТОГС). Данные в хранилище должны поступать автоматически из си-

стемы сбора статистической отчетности;

3. Для обеспечения возможности использования хранилищ данных применяется система нормативно-справочной информации (каталог статистических показателей, справочники, классификаторы и т. д.), через которые реализуется как размещение информации в хранилищах, так и последующий доступ к данным;

4. В проекте предусмотрена реализация функции онлайн-сбора (веб-сбора) статистических форм от предприятий [с использованием электронной цифровой подписи (ЭЦП) данных];

5. Система сбора и обработки статистических данных поддерживается в своей работе как функциональными (объединенная система регистров), так и обеспечивающими (универсальная транспортная подсистема, хранилища данных и т. д.) подсистемами ИВС Росстата.

Масштабность поставленных в проекте задач привела к тому, что было принято решение об их поэтапном внедрении. В настоящий момент у системы государственной статистики имеется достаточно ресурсов как для полного внедрения, так и для развития этих ключевых направлений.

Прежде всего это действующая с прошлого года Федеральная целевая программа «Развитие государственной статистики России в 2007-2011 годах». Только на поставку программно-технических средств для системы Росстата в ней предусмотрено финансирование на сумму 2,3 млрд. рублей, что позволит практически полностью переоснастить систему. Программой предусмотрены поставки 740 серверов, 15,5 тыс. персональных компьютеров с источниками бесперебойного питания, более 3 тыс. принтеров и терминалов сбора данных. Каждый ТОГС получит современное сетевое и стоечное оборудование, системы хранения данных на 8 и более терабайт, оборудование для создания учебных классов и оснащения залов коллегий, полиграфическое и многое другое оборудование.

Уже сегодня очевидно, что поставленные в Федеральной целевой программе (ФЦП) цели могут быть не только достигнуты, но и превышены. В результате проведения государственных закупок за первые два года действия Программы заключено государственных контрактов на поставку, к примеру, современ-

ных персональных компьютеров уже более чем на 13 тыс. единиц - и это при плане в 10 тыс. компьютеров! Таким образом, мы можем быть уверены в том, что современная техническая база для работы ИВС Росстата будет предоставлена уже в ближайшие год-два.

Найдется ли чем загрузить мощнейшую современную технику? Какие прикладные программные средства получит система статистики за этот же период?

В ФЦП на цели внедрения и развития проектных решений ИВС Росстата выделено 432 млн. рублей на период до 2011 г. Пока окончательно не утверждены параметры проекта РСГС-2, предусматривающего сотрудничество с Всемирным банком на 2008-2012 гг., но, скорее всего, это также будут серьезные суммы, пусть и не превышающие финансирование по ФЦП, но вполне сравнимые с ней в этой части.

Управление информационных ресурсов и технологий, воссозданное в структуре Росстата в 2007 г., видит свою основную задачу в полной реализации целей разработанного проекта ИВС Росстата с учетом состояния рынка информационных технологий, принятых нормативных правовых актов и существующего уровня развития программно-технических средств.

Поясним эту мысль на следующем примере.

В 2000 г., когда закладывались концептуальные основы действующего проекта ИВС Росстата, еще не было принято, к примеру, закона «Об электронной цифровой подписи», а доступ в Интернет на рядовом предприятии был редкостью. Сегодня закон об ЭЦП не только принят, но и уверенно действует - в стране развилась инфраструктура специализированных операторов связи, которые снабдили цифровыми подписями и через Интернет принимают налоговую отчетность уже у более миллиона (!) предприятий и организаций Российской Федерации. Уже в 2006 г. свыше 93% предприятий использовали в своей работе персональные компьютеры, более 62% предприятий имели подключение к глобальным информационным сетям. Значения этих показателей растут год от года.

Система статистики не должна игнорировать новые обстоятельства. Принятое законодательство уже не позволяет нам использовать несертифицированные средства ЭЦП, как это было предусмотрено в проекте. Да и сами предприятия стремятся к тому, чтобы максимально унифицировать свое общение с государственными органами, и как следствие, выражают желание пользоваться уже приобретенным сертификатом ЭЦП и привычными программными продуктами для сбора данных.

Простые подсчеты показывают, что для выхода на такие же масштабы взаимодействия с предприятиями по предлагаемой дистанционной системе сдачи отчетности (веб-сбор) системе государственной статистики пришлось бы отвлечь до 10% своей штатной численности только на организацию технического взаимодей-

ствия с сотнями тысяч предприятий. Это и невозможно, и нецелесообразно.

Понятно, что в сложившихся условиях мы должны использовать инфраструктуру специализированных операторов для электронного сбора статистических данных от предприятий. Отметим, что задача активной организации сбора отчетности в электронном виде соответствует как целям, установленным в ФЦП (до 30% предприятий в результате выполнения программы должны перейти на сбор статистики в электронном виде), так и другим задачам Правительства Российской Федерации, которое планирует уже в 2010 г. организовать для любого желающего предприятия возможность сдать отчетность электронным образом, дистанционно.

Такие работы нами уже ведутся как в технической, так и в нормативной части. В частности, в июле 2008 г. Росстатом впервые в истории ведомства утвержден Временный порядок сбора отчетности в электронном виде, который предусматривает возможность для предприятий не дублировать электронный отчет в бумажном виде при его сдаче с ЭЦП через каналы специализированных операторов. Не секрет, что многие предприятия резко отрицательно реагировали на необходимость дублирования информации.

С нашей точки зрения, поступивший в электронном виде от предприятия отчет не может и не должен обрабатываться исключительно на региональном уровне. Правильная маршрутизация полученных отчетов на компьютеры в районном отделе статистики, по месту нахождения предприятия, позволит эффективно проверить его качество, и уже только потом проверенный отчет будет размещен в хранилище данных.

Взаимодействие со специализированными операторами предъявляет более жесткие требования к степени интеграции различных компонент ИВС Росстата. В новой ситуации уже недопустимо неавтоматическое, полуручное взаимодействие многих подсистем, как это, к сожалению, часто встречается до сих пор.

К примеру, для встраивания всех форм отчетности в свои программные средства операторы рассчитывают централизованно и оперативно получать от нашей системы все необходимое: формы статистической отчетности, включая формально-логические контроли, статистические показатели из единого каталога, иную нормативно-справочную информацию, электронные форматы обмена данными.

Между появлением этой информации (назовем ее для простоты «метаданными» системы статистики, хотя, разумеется, в метаданные входит не только то, что нужно для сдачи отчетности) у специализированного оператора и ее использованием для сдачи конкретного отчета предприятием может пройти час, а то и меньше времени. Если аналогичная метainформация еще не будет находиться в базах данных ТОГС, то может воз-

никнуть ситуация, когда корректный отчет предприятия будет отвергнут системой.

К счастью, у системы государственной статистики есть необходимые цифровые каналы связи для того, чтобы обеспечить получение метаинформации в ТОГС. Однако если обработка такой информации будет производиться вручную, то зависимость от человеческого фактора окажется серьезным риском для функционирования всей системы. Таким образом, мы приходим к выводу о том, что распространение метаинформации в системе государственной статистики, распространение версий программных средств (претерпевающих нередко изменения) должно быть должным образом автоматизировано. Одним из первых шагов к реализации этого подхода является начало использования Универсальной транспортной подсистемы (УТП) для передачи статистических данных по самым оперативным (цифровым) каналам связи в нашей системе. С мая 2008 г. прочие каналы связи являются для системы статистики лишь резервными. С помощью УТП и необходимых средств версионного контроля, подсистемы автоматического распространения и обновления метаданных на всех уровнях системы вполне реально решить и эту проблему.

Главной трудностью на этом пути является необходимость целостной реализации внедрения всех компонентов проекта. Лоскутный, позадачный подход к информатизации такого крупного ведомства, как Федеральная служба государственной статистики, с неизбежностью привел бы к невозможности сопряжения между собой разработанных программных средств, технического обеспечения и, как следствие, к невозможности комплексно подойти к решению всех задач системы статистики в информационно-вычислительной системе.

С этой целью Управлением информационных ресурсов и технологий подготовлено техническое задание на полное внедрение и развитие проектных решений ИВС Росстата с соблюдением всех целей, поставленных в утвержденном проекте.

Для этого в техническом задании предусмотрены следующие шаги по отдельным компонентам ИВС Росстата:

- формирование интегрированного информационного ресурса государственной статистики на базе внедрения единого хранилища данных;
- развитие направления анализа и распространения данных и метаданных системы государственной статистики, в том числе на базе внедрения OLAP-технологии для анализа статистических данных;
- создание системы распространения данных государственной статистики на основе разработанной подсистемы хранения и подсистемы метаданных;
- использование каталога статистических показателей (КСП) как средства навигации (поиска) системы

распространения данных и метаданных государственной статистики;

- обеспечение доступа сторонних пользователей к метаданным подсистемы метаданных;
- распространение первичных данных с применением средств, предотвращающих раскрытие данных;
- разработка новой компоненты - формализованной библиотеки экономических описаний (БЭО), развитие компонент КСП и Системы стандартных экономико-статистических классификаций (ССЭСК) для обеспечения взаимодействия с БЭО, создание подсистемы метаданных, интегрированной с процессом производства и распространения статистических данных и метаданных.

Прочие функциональные компоненты должны будут модернизироваться при внедрении для обеспечения решения поставленных задач. К примеру, хранение первичных и агрегированных данных должно будет производиться с обеспечением возможности хранения нескольких версий одного массива данных; должно обеспечиваться ведение архива электронных версий форм с ЭЦП, полученных от предприятий.

Обработка первичных данных должна вестись в хранилище данных (ХД) по соответствующим алгоритмам, получаемым из БЭО, размещение агрегированных данных также должно производиться в ХД.

Уже начат проект по построению интранет-сайта Росстата, который в будущем должен интегрировать внутренний портал ведомства с внутренними же информационными ресурсами и внутренними корпоративными системами Росстата, обеспечивая доступ к нему всех сотрудников системы государственной статистики, подключенных к цифровой сети (в том числе в ТОГС).

Одновременно доработке или разработке подвергнутся и обеспечивающие подсистемы, такие, как подсистема обеспечения безопасности информации, которая должна более тесно и прозрачно интегрироваться в автоматизированное рабочее место сотрудника системы государственной статистики, или Универсальная транспортная подсистема, которая вкуче со средствами версионного контроля должна автоматизировать, в частности, вопросы распространения обновлений к прикладным программным продуктам и распространения метаинформации внутри нашей системы, что должно существенно снизить влияние человеческого фактора на работу ИВС Росстата.

С разработанными Управлением информационных ресурсов и технологий материалами можно будет ознакомиться на сайте Росстата в ближайшее время. Первые результаты от внедрения проектных решений ИВС Росстата будут заметны уже в 2008 г., а выход на промышленное применение спроектированной ИВС Росстата в полном объеме ожидается в начале 2010 г.

ОПТИМАЛЬНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ ВЫПОЛНЕНИЯ ПЕРВИЧНОЙ ОБРАБОТКИ МАТЕРИАЛОВ СТАТИСТИЧЕСКИХ ПЕРЕПИСЕЙ

В.П. Божко, д-р экон. наук,

Московский государственный университет экономики, статистики и информатики (МЭСИ),

А.В. Лури, канд. техн. наук,

Е.Б. Сычев, канд. экон. наук,

ГМЦ Росстата

Важной задачей эффективного проведения различных статистических переписей (всероссийские переписи населения, сельского хозяйства, экономическая перепись и др.) является определение последовательности первичной обработки их материалов в каждом региональном центре, создаваемом в органах государственной статистики. Первичная обработка материалов переписей, независимо от предметной области, как правило, осуществляется по единой технологии. Она заключается в автоматизированной обработке массивов машиночитаемых документов (МЧД) и состоит в последовательном выполнении двух технологических операций:

- сканирование МЧД и ввод данных;
- формальный и логический контроль введенных данных.

Первая технологическая операция выполняется оператором ввода информации на автоматизированном рабочем месте (АРМ), состоящем из персонального компьютера и подключенного к нему сканирующего устройства. При сканировании МЧД данные автоматически распознаются средствами прикладного программного обеспечения АРМ и записываются в базу данных (БД). Если распознавание данных программным способом затруднено или даже невозможно (например, ввиду некачественного ручного заполнения переписчиком текста на бланке), оператор самостоятельно на клавиатуре АРМ присваивает нераспознанным данным действительные значения.

В результате выполнения этой операции формируется массив данных, представляющий собой электронную копию машиночитаемых документов с внесенными при распознавании необходимыми изменениями.

Вторая технологическая операция состоит в формальном и логическом контроле введенных данных, который выполняет экономист на соответствующем АРМ. Формальный контроль данных осуществляется автоматически программным средством и состоит в проверке наличия обязательных показателей и соответствия присутствующих данных требуемому диапазону значений. Логический контроль данных состоит в проверке непротиворечивости данных, семантически связанных между собой в одном или нескольких документах одного массива. При обнаружении ошибок формального и логического контроля экономист выполня-

ет их исправление по заранее определенным правилам.

Первичная обработка материалов переписей завершается формированием электронного файла, формат которого доступен для последующей обработки.

Таким образом, первичная обработка одного массива МЧД осуществляется в два этапа, причем второй этап может быть выполнен только после полного завершения первого этапа (сразу или по истечении некоторого времени); при этом один центр может обрабатывать МЧД нескольких субъектов Федерации.

Результаты исследования показали, что очередность обработки машиночитаемых документов влияет на такой важный показатель, как общее время выполнения работы в центре. Кроме того, для центров, обрабатывающих информацию нескольких субъектов Федерации, от очередности обработки МЧД зависят также сроки завершения работ по каждому субъекту.

Организовав оптимальную последовательность выполнения работ в каждом центре, можно, стремясь достичь минимизации общего времени, получить также возможность предоставления информационных услуг с результатами переписей по некоторым субъектам до завершения всех работ (под работой будем понимать обработку одного массива МЧД).

Сформулируем постановку задачи в следующем виде.

Дано:

M - общее число работ, которое будет выполнено в центре;

N - количество субъектов, МЧД которых обрабатываются в центре;

t_{m1} и t_{m2} - продолжительность соответственно первого и второго этапов работ, $m = \overline{1, M}$.

Обозначим через s_{m1} и s_{m2} соответственно сроки начала 1-го и 2-го этапов m -й работы, $m = \overline{1, M}$.

Требуется:

определить такое упорядоченное множество работ $P = \{p_1, p_2, \dots, p_m, \dots, p_M\}$, для которого

$$s_{p_1 1} < s_{p_2 1} < \dots < s_{p_m 1} < \dots < s_{p_M 1},$$

где p_m - номер работы;

$p_m \in [1, M]$, $m = \overline{1, M}$, определяющее последовательность выполнения работ с заданной продолжитель-

