

О ЦЕНТРАЛИЗАЦИИ ОБРАБОТКИ СТАТИСТИЧЕСКИХ ДАННЫХ

М.В. Бурдаков

Росстат, с точки зрения информационных технологий (ИТ), существенным образом отличается от других федеральных органов исполнительной власти, поскольку представляет собой практически фабрику по сбору, обработке и представлению огромных массивов информации. Информационно-вычислительная система (ИВС) Росстата является одной из крупнейших в стране и в рейтинге C-News занимает пятое место.

Предметная область, в которой работает Росстат, весьма сложная, поэтому, естественно, требует очень проработанной и качественной поддержки технологических процессов со стороны ИТ. Более того, разнообразие постоянно возникающих производственных задач требует еще и высокой гибкости.

Чтобы стал понятен размах идущих преобразований в ИТ, сначала совсем коротко обратимся к предыдущему состоянию ИВС, затем перейдем к ее текущему положению и уже далее - к ближайшим перспективам.

На момент подготовки самой первой концепции развития ИВС Росстата в 2002 г. мы имели, с точки зрения ИТ, децентрализованную территориально распределенную информационно-вычислительную структуру.

Главными минусами такого построения системы ИТ было то, что каждое из обособленных территориальных подразделений (включая, естественно, федеральный уровень) имело собственные устоявшиеся технологические цепочки и соответствующие им инструменты информатизации. Каждый был сам по себе и самостоятельно выполнял производственные задачи, в одиночку сражаясь с постоянно возникающими организационными и технологическими трудностями.

Кроме того, имела место так называемая «лоскутная» автоматизация, при которой автоматизировались не процессы целиком, а некоторые этапы процессов и отдельные задачи. И каждое подразделение дописывало

для своего удобства так называемые «бантики», что еще больше делало процесс несходящимся. Далее при таком построении системы ИТ и речи быть не могло ни о каком действенном контроле за правильностью, качеством и ходом технологических процессов на местах.

Фактически прежняя концепция развития принципиально поддерживала такой подход и подразумевала модернизацию существовавшей децентрализованной ИВС, исходя из текущих задач и потребностей подразделений Росстата.

Такое положение в ИТ было не уникально, историческое наследие автоматизации конца 1990-х - начала 2000-х годов во всех крупных компаниях и холдингах страны в подавляющем большинстве случаев было не менее плачевно: отдельные системы и «системки», работающие на разных программных платформах и использующие разные локальные справочники, никак не связанные между собой, однозначно препятствовали информационному обмену, консолидации данных и эффективности работы в целом.

Между тем, жизнь все время ставит новые вызовы, которые необходимо учитывать:

- прежде всего, это постоянное увеличение потребности в различной статистической информации со стороны общества, с одновременным требованием повышения ее оперативности и точности;

- стандартизация форматов статистической информации, таких, как Open Data, SDMX, DDI, а также появление и возможность обработки новых источников данных, например BIG DATA;

- кадровые проблемы, связанные с невысоким уровнем заработной платы ИТ-специалистов на местах, а в ряде случаев вообще с отсутствием квалифицированных ИТ-специалистов;

- общая тенденция к сокращению государственных служащих;

Бурдаков Михаил Владиславович (stat@gks.ru) - начальник управления информационных ресурсов и технологий Росстата (г. Москва, Россия).

- моральное устаревание технологических и программных решений, внедренных в Росстате в начале-середине 2000-х годов;

- существенное сокращение финансирования Росстата. На сегодняшний день, например, недофинансирование ИТ Росстата составляет порядка 2 млрд рублей в год;

- общее повышение стоимости услуг по ремонту, замене всего эксплуатируемого оборудования и т. д., которое делает финансово невыгодным содержание серверных помещений в ТОГСах.

Подобные трудности предприятия, а особенно территориально распределенные, во всем мире решают одним способом - внедрением у себя так называемых систем управления ресурсами организации. В случае Росстата - это переход от децентрализованной ИТ-структуры к максимально централизованной.

Этот подход и был зафиксирован руководителем Росстата в 2011 г. в новой Концепции развития ИВС Росстата на период до 2017 г. (далее - Концепция). Помимо реализации централизованной системы сбора, обработки и представления статистической информации, в ней заложены такие требования к гибкости и унификации технологических и организационных схем, а также к модульности всех создаваемых решений, которые позволят проводить наращивание или изменение функционала в создаваемых ИТ-системах без существенных доработок и программирования с максимальным использованием механизмов и настроек внутри самой системы. Это значительно сократит количество масштабных работ и связанных с ними затрат.

Естественно, в Концепцию закладывалась возможность предоставления ТОГСам инструмента, позволяющего учесть региональные особенности и удовлетворить их региональные потребности, однако оставаясь в рамках единых унифицированных технологических подходов и организационных схем.

Реализация Концепции обеспечит:

- работу всех подразделений по единым правилам, включающим в себя детерминированные каталоги респондентов, установленные экономическими описаниями, общие для всех справочники, классификаторы, то есть всю систему нормативно-справоч-

ной информации, единые алгоритмы обработки информации, а также единую для всех систему электронного сбора. В Росстате будет создан непрерывный производственный конвейер по сбору, обработке и представлению официальной статистической информации;

- перенос всех основных задач по сопровождению и развитию ИТ-инфраструктуры, а также совершенствованию программного обеспечения на федеральный центр, создавая условия, при которых в ТОГСах достаточно будет иметь специалистов среднего уровня, занимающихся обслуживанием и подключением офисного оборудования, что обеспечит оптимизацию ресурсов;

- специальными средствами ИТ тотальный контроль за качеством и правильностью производственных процессов и работой пользователей по всей системе Росстата;

- возможность оперативно вносить изменения в работу подразделений, особенно в условиях реорганизации системы Росстата, путем перераспределения ролей исполнителей в системе, что повысит эффективность работы.

Кроме того, будут обеспечены:

- технологическая возможность для гибкого представления статистических данных во всех актуальных форматах для всех уровней пользователей;

- возможность оперативного межведомственного информационного взаимодействия с внешними системами;

- создание единого межведомственного Реестра отчетности и форм наблюдения, что приведет к упорядочению сбора статистической информации, исключению дублирования показателей, повышению открытости данных и снижению отчетной нагрузки на бизнес.

Необходимо отметить, что все, о чем пойдет речь дальше, относится, прежде всего, к унифицированным формам. На базе федерального центра в ГМЦ и в центральном аппарате Росстата создается мощная информационная система, включающая в себя такие основные взаимосвязанные блоки, как:

- Единая система нормативно справочной информации - НСИ;

- Система подготовки электронных экономических описаний - СПЭЭО;

- Централизованный статистический регистр - АС ГС ОФСН;
- Единая система сбора и обработки отчетности - ЕССО;
- Централизованная система обработки данных - ЦСОД;
- Единое хранилище данных - ЕХД;
- системы представления и распространения статистической информации;
- множество иных подсистем обеспечивающих совместную согласованную работу как информационных систем, так и пользователей.

Архитектура, заложенная в этих системах и подсистемах, позволяет обеспечить доступ для работы в них с любого уровня (федерального, регионального и городского) всем сотрудникам Росстата в соответствии с теми ролями и правами, которые будут установлены необходимостью выполнения пользователями производственных задач.

По сути, перед нами стоит задача создания и внедрения принципиально новых ИТ-систем с учетом непрерывности исполнения Производственного плана Росстата.

В 2011-2013 гг. была заложена основа ИВС Росстата нового поколения.

В части обеспечивающих систем были созданы Единая ведомственная мультисервисная сеть (МСС), Единая служба технической поддержки, внедрена система электронного документооборота (СЭД), созданы Интернет-портал и Интранет-портал Росстата.

Организовано Единое хранилище данных и система многомерного анализа данных с возможностями формирования произвольных запросов к ЕХД и подключения к ЕХД аналитических систем (IBM SPSS, IBM Cognos BI, Super STAR). В рамках работ по загрузке показателей в ЕХД проделана колоссальная работа по созданию базы единого, непротиворечивого источника данных НСИ, без чего было бы просто невозможно предоставлять данные показателей Росстата в ЕМИСС.

Создана и внедрена базовая компонента системы веб-сбора.

В 2014 г. введена в промышленную эксплуатацию Система подготовки электронных экономических описаний, которая позволяет формировать XML-шаблоны для их размещения на официальном сайте Росстата уже на этапе подготовки экономических описаний,

а также предоставляет возможности проектирования форм статистического наблюдения в непосредственной увязке с метаданными НСИ.

В 2015 г. были выполнены работы по модернизации Единой системы сбора и обработки, направленные на централизацию обработки первичных статистических данных.

Указанная работа тесно связана с проведенной в том же году работой по совершенствованию технологии ведения и использования Статистического регистра хозяйствующих субъектов на основе единого централизованного ресурса - АС ГС ОФСН, выполнение которой позволило централизованно формировать согласованные каталоги хозяйствующих субъектов, представляющих первичные статистические данные по унифицированным формам федерального статистического наблюдения. Также на базе ЕХД и НСИ в составе СМАД создан функционал формирования витрин данных.

На этапе планирования обследования подготавливается формализованное экономическое описание, то есть тот документ, который описывает методологию (правила, порядок, алгоритмы, формы предоставления результатов) разработки показателей деятельности предприятий на основе форм ФСН, а также содержит требования к информационному и техническому обеспечению.

Формализация ЭО выполняется с помощью Системы подготовки электронных экономических описаний. Критерии отбора респондентов СПЭЭО получает из централизованной АС ГС ОФСН. Источником информации о справочниках и показателях для СПЭЭО и АС ГС ОФСН является единая НСИ на базе СМАД.

На выходе СПЭЭО получается формализованное электронное экономическое описание в тех форматах, которые принимаются смежными системами, включая:

- технологические карты;
- сводные показатели;
- алгоритмы расчета;
- формальные и логические контроли;
- импортируемые показатели;
- выходные таблицы;
- выгрузки во внешние системы.

Такое решение уже сейчас позволяет использовать подготовленные подобным обра-

зом данные как готовые настройки для централизованной системы обработки данных.

Таким образом, за формализацию ЭО будут отвечать специалисты центрального аппарата Росстата, а специалисты ТОГСов - за ведение региональной составляющей НСИ.

Сбор информации от респондентов осуществляется в единую систему сбора ЕССО. Одновременно централизованная система обработки данных:

- производит загрузку электронного ЭО, загрузку каталогов респондентов;
- осуществляет передачу информации в ЕССО по респондентам и формам ФСН, в рамках которых следует организовать сбор данных;
- контролирует сбор обработанных отчетов респондентов в ЕССО.

В процессе обработки информации ЦСОД:

- осуществляет контроль и формирование пообъектной БД, содержащей данные по респондентам и данные, собранные в рамках ФСН;
- рассчитывает сводные показатели;
- производит выгрузку показателей во внешние системы;
- формирует выходные и публикационные таблицы. В состав ЭЭО включаются технологические карты для каждой формы ФСН, которые представляют собой совокупность процессов, например таких, как расчет сводных показателей на РУ, ФУ и т. д. Таким образом, пользователям в соответствии с их уровнем доступа (пользователи ТОГСов, ГМЦ, ЦА) будут доступны на выполнение только «свои» процессы.

Для открытой и закрытой частей разработаны разные технологические карты, учитывающие особенности обработки данных в каждой части.

Невозможно не сказать несколько слов о ЦСОД. Сложность задачи по его реализации повлекла за собой множество неординарных технических решений, и на сегодняшний день ЦСОД Росстата по вычислительной мощности сравним с вычислительными фермами таких гигантов, как Google и Яндекс.

Использование в ЦСОД для хранения информации нереляционной базы данных Cassandra обеспечило нам бесконечную масштабируемость системы и повышенную ус-

тойчивость к сбоям, с одновременной минимизацией затрат на это.

Кластер Кассандры позволяет выполнять 10000 операций записи в секунду и 90000 операций чтения в секунду на каждом из 33 входящих в него серверов. Это очень важно, так как в рамках одного шага одного процесса вычисления происходит запись в среднем 1-2 млн строк.

В качестве базы данных для проведения расчетов и ФЛК был выбран Oracle. Кластер Oracle состоит из 4 серверов, по 1 ТБ оперативной памяти на каждом сервере. Это позволяет осуществлять расчет со скоростью 50000 записи строк в секунду. При проведении среднего расчета происходит запись 3 млн строк.

Как пример успешной реализации централизованного подхода и демонстрации его жизнеспособности можно привести организацию вполне успешного перевода в прошлом году аппаратно-программного комплекса по сбору цен на новую технологическую платформу. В результате это позволило получить следующие преимущества:

- полное снятие с системных администраторов ТОГСов работы по поддержке и обновлению ПО АПК РЦ. Нет необходимости в поддержке серверов приложений и серверов БД, их настройке, администрировании, обновлению;
- отсутствие временных задержек в доступности данных. Как только котировка вносится в режиме реального времени в базу данных ФУ, она тут же становится доступной на всех уровнях с учетом прав доступа. Больше не нужно ждать, пока запустится синхронизация локальных копий БД с ФУ. В любой момент времени все участники обследования видят единую картину данных. Вероятность выполнения расчета РУ на неактуальных данных сводится к нулю;
- возможность доступа к приложению АКП РЦ из любой точки, имеющей доступ к сети Интернет, с любого устройства, имеющего браузер, поддерживающий html 5.

Консолидированная платформа открывает технические возможности для развития методологии обследования с целью повышения скорости и качества сбора данных.

Особо отметим трудности, которые приходится преодолевать при реализации Кон-

цепции. Это весьма существенным образом сказывается на скорости разработки и внедрения новой системы. В основном все сложности методологического характера унаследованы от старой системы. Сейчас все наши управления стали это понимать и включились в решение этих задач. Более того, без их помощи сделать бы не удалось вообще ничего.

В рамках реализации Концепции развития ИВС Росстата остается еще решение многих задач, как технологических, так и организационных, и методологических.

С точки зрения организации и методологии, ключевыми являются:

- планомерный перевод бумажных ЭО в электронный вид в СПЭЭО силами специалистов центрального аппарата Росстата (решая обозначенные выше проблемы в части ЭО и НСИ);

- разработка и утверждение регламентов работы со статистической информацией и метаданными, а также контроль за их соблюдением в централизованной технологии обработки данных.

С точки зрения технологии, должны быть решены следующие ключевые задачи:

- расширение перечня унифицированных форм ФСН, обработка данных по которым должна выполняться в централизованной технологии;

- вывод из эксплуатации всех разрозненных децентрализованных систем ведения НСИ с заменой их функционалом централизованной НСИ на базе СМАД;

- автоматизация процесса ведения региональной составляющей как в НСИ, так и в экономических описаниях (региональные сборки и выборки, региональная часть справочников);

- вывод из эксплуатации разрозненных отраслевых баз данных с переносом их информационных фондов в ЕХД на базе СМАД и созданием для каждой отдельной базы витрины с возможностью размещения на Инtranет/Интернет портале Росстата.

Проект РСГС Международного банка реконструкции и развития, в рамках которого стала возможной такая основательная модернизация ИВС Росстата, находится на своей завершающей стадии: остался этап 2017-2019 гг. По его завершению масштабных, крупных проектов в рамках развития ИВС Росстата уже не будет. Доработки ранее созданных систем будут осуществляться исключительно в рамках работ по сопровождению.

Наша основная задача до завершения этого проекта - построить основательный, надежный фундамент, конвейер из гибко интегрированных между собой систем, который в дальнейшем может быть модернизирован в рамках сопровождения.