

ЭЛЕКТРОННЫЙ КАТАЛОГ СТАТИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ КАК РАЗВИВАЮЩАЯСЯ ТЕХНОЛОГИЧЕСКАЯ КОМПОНЕНТА ИВС ГОСКОМСТАТА РОССИИ

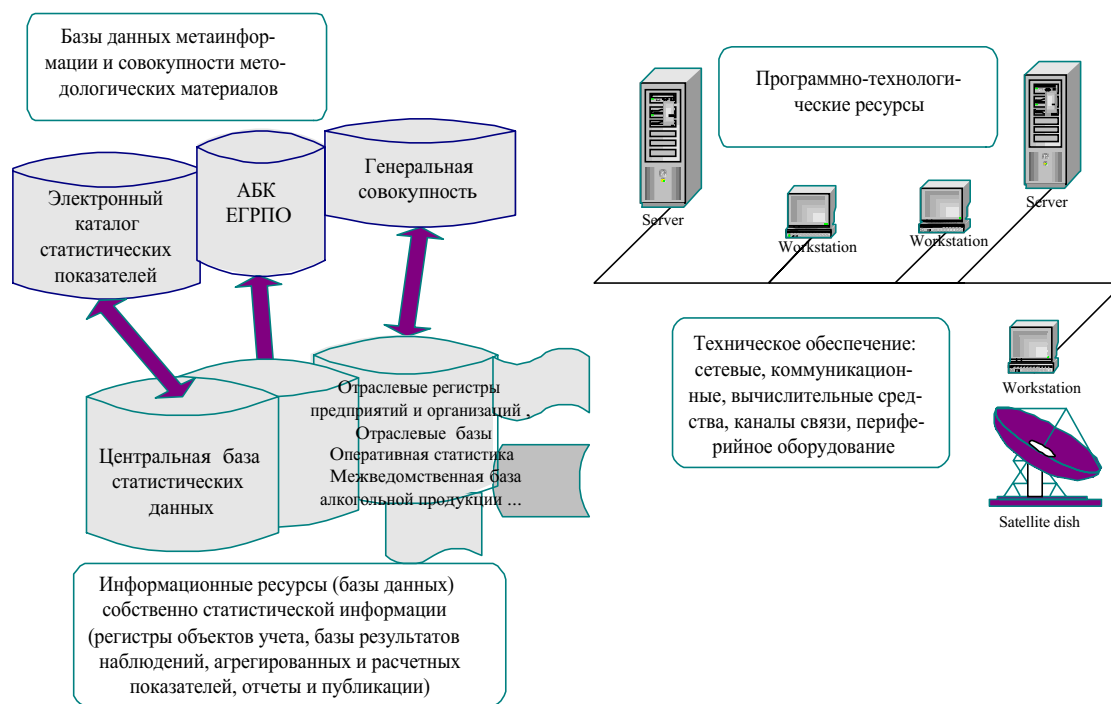
Н.А. Казакова,
НИПИСтатинформ Госкомстата России

Информационно-вычислительная система (ИВС) Госкомстата России состоит из целого комплекса подсистем, баз данных и нормативно-справочной информации, а также программно-технологических средств, обеспечиваю-

щих их функционирование и взаимодействие.

Одной из важных технологических компонент ИВС является Электронный каталог статистических показателей (см. рис. 1).

Рис. 1. Архитектура ИВС Госкомстата России



Электронный каталог статистических показателей (ЭКСП) - это информационно-справочная система метainформации, представляющая собой, с точки зрения экономического содержания, систематизированный перечень показателей, используемых в статистической практике, а с точки зрения организации данных - взаимосвязанную совокупность показателей, нормативно-справочной информации и таблиц метаданных, размещенную на сервере Госкомстата России и имеющую централизованный доступ пользователей посредством прикладного программного обеспечения с использованием операционной системы Windows и СУБД SQL Server.

Совокупность статистических показателей, классификаций, документации, методологических материалов и дру-

гой нормативно-справочной информации используется при проектировании и решении статистических задач, а также для информационно-справочного обслуживания пользователей статистической информации - органов управления, предприятий и организаций.

Проектирование Электронного каталога статистических показателей как базы данных и его первоначальное формирование осуществлялось НИПИСтатинформом Госкомстата России. В 1997 г. база была установлена на сервере Госкомстата России для ее дальнейшего администрирования в центральном аппарате. В то время создание базы ЭКСП позволило автоматизировать и соответственно ускорить процесс формирования Центральной базы данных Госкомстата России (ЦБСД).

Электронный каталог статистических показателей введен в действие с 1 марта 1997 г. и в настоящее время является важной технологической системной разработкой для создания единой метаинформационной системы Госкомстата России.

Статистические показатели, используемые органами государственной статистики, подлежат обязательной регистрации в ЭКСП. При регистрации конкретный показатель получает уникальный код, однозначно его идентифицирующий. Статистический показатель в ЭКСП характеризует качественную сторону социально-экономических явлений и процессов в обществе, отображающую сущность явления или процесса в конкретных условиях места и времени.

База ЭКСП содержит перечень статистических показателей, которые определены в формах государственного статистического наблюдения, аналитических расчетах и выходных статистических материалах, что в настоящее время составляет около 16 тыс. позиций, большой объем метаданных, описывающих разрезы разработки показателей, периодичность разработки, формы наблюдения, базы данных и т. д.

В основу действующего каталога положен рубрикатор, построенный в соответствии с наименованиями и иерархией существующей группировки показателей и отражающий основную деятельность Госкомстата России.

Объем действующей базы составляет 500 Мбайт на SQL-сервере Госкомстата России. К ЭКСП подключены рабочие станции всех структурных подразделений.

Для обеспечения функционирования Электронного каталога статистических показателей НИПИСтатинформом разработана технология и прикладная программная среда, включающая большой комплекс функциональных возможностей по формированию и актуализации базы и работы пользователей.

Среда разработки и функциональные возможности

Электронный каталог статистических показателей функционирует в локальной вычислительной сети (ЛВС) центрального аппарата Госкомстата России. Прикладное программное обеспечение разработано на языке Power Builder, который является мощным открытым инструментарием для создания сложных приложений, работающих не только в традиционной двухуровневой и трехуровневой архитектурах «клиент-сервер», но также и в многоуровневой архитектуре. Power Builder сочетает быстроту и удобство разработки с возможностью генерации компилируемых распределенных приложений.

Power Builder - программный продукт фирмы Sybase Inc, является лидером рынка, лучшим в своем классе инструментом разработки. В США не менее 50% всех приложений, работающих с СУБД, написано средствами Power Builder. В России многие известные фирмы, специализирующиеся на разработке прикладного программного обеспечения, для своих разработок используют Power Builder. НИПИСтатинформ Госкомстата России разрабатывает программы для работы с базами данных средствами Power

Builder с 1996 г. и на протяжении всех последних лет использует самые современные разработки этого программного продукта.

Программно-технологические средства (ПТС) Электронного каталога статистических показателей подразделяются на два главных блока: администратора и пользователя, каждый из которых имеет несколько режимов работы.

Администрирование Электронного каталога статистических показателей осуществляется Управлением статистического планирования и организации статистического наблюдения Госкомстата России.

С помощью программных средств администратора обеспечена возможность ведения и актуализации ЭКСП, проектирования таблиц новых классификаторов и локальных справочников, формирования новых разрезов разработки, периодов отчетности и т. д. Для уменьшения трудоемкости работы администратора ЭКСП заполнение и актуализация справочников могут осуществляться в пакетном режиме. Для этого используется утвержденный обменный формат.

В составе ПТС ЭКСП имеется программный блок формирования системы методологических пояснений к показателям: каждый показатель имеет методологические пояснения, которые представляют собой научно обоснованные методологические подходы к формированию и определению конкретного статистического показателя.

Разработка методологических материалов в каждом управлении производится на основе анализа информационных потребностей органов управления и других пользователей статистической информации. Работа заключается в построении и совершенствовании системы статистических показателей и методов их исчисления, в подготовке бланков статистических форм и инструкций по их заполнению и представлению, а также в постановке статистических задач с целью их автоматизированного решения.

В статистической практике Электронный каталог статистических показателей представляет собой базу данных об исследованиях, проводимых Госкомстатом России. Каждое управление, используя каталог, имеет возможность получить:

- методологию расчета показателей;
- информацию о том, в каких формах статистической отчетности используется данный показатель;
- информацию о том, в каких базах данных содержатся данные о тех экономических явлениях, которые описывает данный показатель;
- по каким разрезам разрабатывается показатель.

Интерфейс пользователя обеспечивает получение информационно-справочных запросов к базе ЭКСП в виде паспорта показателя (рис. 2) и перечня показателей, возможность сквозного многокритериального поиска, поиска по контексту, формирования полной справки о показателях конкретной формы отчетности, печати и сохранения результатов поиска в стандартных средствах.

Паспорт показателя - это документ, сформированный с помощью программных средств Электронного каталога

статистических показателей и содержащий полную информацию о показателе: код показателя, наименование показателя, единица измерения, периодичность разработки,

разрезность разработки, признаки показателя, отдельные позиции признаков, формы государственного статистического наблюдения, базы данных.

Рис. 2. Пример паспорта показателя в интерфейсе пользователя

Паспорт показателя 1311000030001

Паспорт Перечень Окно Помощь

Основная информация

Код: 1311 0000 30001 Показатель: Действует Дата создания: 14/11/2001 Дата отмены:

Наименование: Инвестиции в основной капитал

Единица измерения: 372-Тысяча рублей

Ответственный: Бегунов И.В., Управление статистики основных фондов и инвестиций, 207-40-01

Объекты применения

Наименование ОП	Тип ОП
Оперативная статистика	БД
Сведения об инвестициях	форма
Центральная статистическая база данных (СБД)	БД

Периодичность

Периодичность	Характеристика	Тип
квартальная	нарастающая с начала года	время сверше
квартальная	не нарастающая	отсутствует

РАЗРЕЗЫ (таблицы 4 разрезы)

Наименование	Код
по территориям	S_ONATO
по отраслям экономики	S_ONOEN
по организационно-правовым формам	S_ONOFF
по государственному сектору	S_OS

ПРИЗНАКИ (таблицы 0 признаков)

Наименование

Перечень по организационно-правовым формам (таблицы 73 объекта)

Код	Наименование	Организационно-правовая форма
4	Крестьянские (фермерские) хозяйства и индивидуальные частные предприниматели (53+45+54)	
25	Другие предприятия и объединения (неофициальный код)	
35	Кредитные и кредитно-финансовые организации	
40	КОММЕРЧЕСКИЕ ОРГАНИЗАЦИИ	
41	ПРЕДПРИЯТИЯ	
42	ГОСУДАРСТВЕННЫЕ ПРЕДПРИЯТИЯ	
43	МУНИЦИПАЛЬНЫЕ ПРЕДПРИЯТИЯ	
44	ПРЕДПРИЯТИЯ ОБЩЕСТВЕННЫХ И РЕЛИГИОЗНЫХ ОРГАНИЗАЦИЙ	
45	ПРЕДПРИЯТИЯ ПОТРЕБИТЕЛЬСКОЙ КООПЕРАЦИИ	
46	ИНДИВИДУАЛЬНЫЕ (СЕМЕЙНЫЕ) ЧАСТНЫЕ ПРЕДПРИЯТИЯ (С ПРИВЛЕЧЕНИЕМ НАЕМНОГО ТРУДА)	
47	АКЦИОНЕРНЫЕ ОБЩЕСТВА ОТКРЫТОГО ТИПА, ОТКРЫТЫЕ	

Комплекс ПТС ЭКСП содержит автоматизированную систему взаимосвязи с базами данных Госкомстата России. В информации о каждом показателе в базе ЭКСП имеются сведения об использовании показателя в той или иной базе данных, заполняемые программно на основании справки, полученной из базы. По запросу пользователя Электронного каталога статистических показателей может быть получена справка о составе показателей в каждой базе данных с датой актуализации каждого показателя. Это позволяет своевременно вносить изменения в наименование и признаки показателей в базах данных.

Программные средства ЭКСП предусматривают механизмы защиты объектов и данных от несанкционированного доступа. Права доступа каждого пользователя определяются и назначаются администратором системы.

Роль ЭКСП в технологии электронной обработки информации и базах данных

При создании экономических описаний на разработку задач электронной обработки данных специалисты Госкомстата России пользуются программным интерфейсом Электронного каталога статистических показателей. Для этих целей имеются специальные программные средства, обеспечивающие взаимосвязь базы ЭКСП, базы норма-

тивно-справочной информации и Генеральной совокупности. Эти программные средства также разработаны НИПИстатинформом Госкомстата России.

Электронный каталог статистических показателей является источником метаданных для проектирования новых и актуализации и развития имеющихся информационных баз статистических данных.

Особенно возрастает роль ЭКСП с внедрением банковской технологии в электронную обработку информации.

Использование уникальных кодов позволяет вводить показатели из форм отчетности федерального государственного статистического наблюдения в микробазу, где осуществляются их контроль, агрегация, получение выходных таблиц и создание макробазы агрегированных и расчетных показателей.

В настоящее время с применением полной банковской технологии, когда все технологические этапы электронной обработки данных осуществляются в базе данных, разрабатываются такие задачи, как «Показатели производства спирта этилового, алкогольной продукции, спиртосодержащих растворов и розливе алкогольной продукции» и «Использование денежных средств предприятий»; с отчетности за I квартал 2002 г. переведены на полную банковскую

обработку показатели деятельности предприятий с участием иностранного капитала и выполнена новая разработка «Показатели деятельности естественных монополий».

Программно-технологический комплекс «Оперативная статистика», ЦБСД, отраслевые базы данных, базы «Структурное обследование предприятий», «Сплошное обследование малых предприятий» функционируют с использованием смешанной банковской технологии, когда некоторые этапы техпроцесса, такие, как ввод, контроль, агрегирование данных, выполняются в технологической базе отчетного периода, а затем данные в обменном формате загружаются в базу постоянного хранения.

С использованием Электронного каталога статистических показателей формируются интегрированные информационные ресурсы, объединяющие показатели из разных источников, например, «Межведомственная база данных по предприятиям и аккредитованным организациям, имеющим лицензии на осуществление деятельности в сфере производства спирта этилового, алкогольной продукции, спиртосодержащих растворов, розлива и оптовой реализации алкогольной продукции».

Унификация статистических показателей на основе ЭКСП обеспечила возможность применения банковской технологии обработки информации не только на федеральном, но и региональном уровнях.

Пользователи баз данных имеют возможность работать как с первичными показателями, полученными от предприятий, так и с результатами их обработки, производить свои аналитические расчеты, сохраняя результаты в стандартных средствах Word, Excel, SPSS.

Перспективы развития

В результате происходящих в экономике России реформ появляются принципиально новые экономические объекты, рождаются новые понятия, изменяется номенклатура предоставляемых предприятиями и учреждениями услуг. Радикальные изменения претерпевают информационные системы. Благодаря научно-техническому прогрессу появились новые технологии и программные средства, возникли новые подходы к их проектированию.

ЭКСП - открытая система для новых показателей, которая актуализируется с учетом особенностей развития экономики, социальной и демографической ситуации в обществе. Формирование рыночных отношений, системы национального счетоводства, учет требований международных стандартов в области статистики оказывают заметное влияние на состав и характер статистических показателей.

Одновременно с накоплением позитивного опыта действующего Электронного каталога статистических показателей были выявлены проблемы его функционирования, которые достаточно полно были отражены в статье специалистов Госкомстата России и НИИСтатистики «Проблемы совершенствования Каталога статистических показателей»¹.

Главной проблемой построения ЭКСП является разработка единой методологии определения основания показателя и его признаков, критериев отнесения к основанию показателя или его признаку, разработка методологических основ систематизации (группировки) показателей.

В существующей статистической практике нет однозначного определения и методологического обоснования критериев формирования основания показателя и его признаковой части.

Отсутствие четкого представления о содержании понятий «показатель» и «признаковая часть показателя» негативно влияет на состав показателей каталога, так как происходит подмена основания показателя на признаковую часть, и наоборот; при этом состав показателей в одних случаях обедняется, в других - неоправданно расширяется.

Основными методологическими проблемами являются:

- унификация статистических показателей на межотраслевой основе, которая означает приведение их к единообразию, когда однотипные по содержанию показатели не рассредоточиваются по блокам, отраслям, видам деятельности, а концентрируются в одном блоке, группе;
- методология разграничения основания показателя и его признаковой части;
- методология формирования признаковой части с учетом последовательности построения признаков, их многоуровневости;
- методология формирования межотраслевого рубрикатора ЭКСП, основанного на унификации показателей;
- методология формирования КСП на основе межотраслевого рубрикатора.

Наряду с этим, существует ряд частных методологических проблем:

- методология понимания показателей с одинаковым наименованием, имеющих разные единицы измерения, оценки (стоимостная, натуральная), методики исчисления и др.;
- показатели с общим содержанием нередко представлены одновременно в нескольких рубриках;
- разграничение и взаимосвязь понятий «показатель» и «графа», «строка», «графоклетка» статистической таблицы (формы).

Главными недостатками действующей системы ЭКСП являются:

- низкая производительность (частично от того, что в СУБД не развит механизм обработки транзакций);
- несовершенство инструмента кодирования классификатора (рубрикатора) и самих показателей;
- отсутствие поиска по объектам разработки (например, не существует возможности найти показатели, разрабатываемые по конкретной форме собственности ОКФС);
- несовершенство пользовательского интерфейса;
- низкая степень защищенности системы и данных от несанкционированного доступа;

¹ См.: Вопросы статистики. № 2 за 2001. С. 11-17.

- потеря данных.

В настоящее время НИПИСтатинформ Госкомстата России ведет разработку новых программно-технологических средств формирования и ведения Электронного каталога статистических показателей федерального уровня на новой методологической основе, утвержденной Постановлением Госкомстата России № 118 от 30 ноября 2000 г.

Разрабатываемые ПТС призваны обеспечить совершенствование ЭКСП, его построение как сквозной системы, в которой каждый статистический показатель уникален, имеет свой код и используется для всего комплекса задач статистического наблюдения, а не только в привязке к конкретной форме федерального государственного статистического наблюдения.

С точки зрения достижения косвенного эффекта, целью новой разработки является возможность добавления и корректировки показателей в ЭКСП, добавление рубрик, изменение рубрик, формирование призначной части показателя, формирование перечней объектов использования показателей; получение сведений о статистических показателях и их признаках, распределение статистических показателей по объектам их использования, возможность проследить историю изменения статистических показателей; создание новых форм статистической отчетности, получение сведений о показателях, используемых в формах.

Программные средства ЭКСП должны обеспечивать упорядоченность статистических показателей с учетом системы научных взглядов на проблемы развития экономики в их единстве: определение факторов (ресурсов и условий) и результатов социально-экономического развития экономики. Рубрикатор нового ЭКСП будет иметь три главные рубрики:

- факторы развития экономики;
- результаты развития экономики;
- статистическая оценка степени реализации целей развития экономики.

С точки зрения получения прямого эффекта, целью разработки являются:

- сокращение времени решения задачи, а следовательно, увеличение оперативности появления результатов;
- увеличение достоверности и точности результатов;
- сокращение стоимостных и трудовых затрат на обработку входных документов;
- увеличение аналитической эффективности результативной информации;
- повышение своевременности получения информации для предоставления.

НИПИСтатинформ в своих разработках применяет принципиально новые системы программирования, представляющие собой инструментальные средства на основе информационных технологий, предназначенных для проектирования функционального программного обеспечения.

Для разработки программного обеспечения выбраны современные программные продукты, такие, как case-средство Power-Designer, СУБД MS SQL 2000, средство быстрой разработки приложений (RAD) Power Builder 7.0. Сред-

дой для функционирования прикладного программного обеспечения является Windows 2000, платформой для СУБД - Windows NT. В работах принимают участие специалисты с высоким уровнем подготовки и опытом работы, что гарантирует высокое качество разрабатываемых прикладных программ.

В настоящее время выполнены работы по моделированию и проектированию базы ЭКСП с учетом новой методологии его формирования, ведения и применения. Разработаны программные средства, реализующие систему иерархии статистических показателей в соответствии с утвержденными методологическими положениями о КСП.

Поскольку Электронный каталог статистических показателей - это развивающаяся информационная система, в которой информация постоянно изменяется (актуализируется), в составе ПТС ЭКСП разрабатываются программные средства формирования архивов и возможности работы с ними: информационно-поисковая система и возможность восстановления отмененных показателей.

В 2002-2003 гг. предполагается осуществить поэтапный переход к формированию ЭКСП на новой методологической основе. При этом в течение переходного периода сохранится действующая база ЭКСП.

Особенно возрастает значение Электронного каталога статистических показателей в связи с отменой Общероссийского классификатора технико-экономических и социальных показателей с 1 января 2003 г.

В декабре 2001 г. осуществлен перевод действующей базы из среды СУБД SQL Server 6.5 в среду СУБД SQL Server 2000, которая должна улучшить параметры работы администратора и пользователей с базой ЭКСП.

Следует отметить перспективное направление развития ПТС ЭКСП для целей региональной статистики как новой отрасли социально-экономической статистики.

Превращение регионов Российской Федерации в экономически самоуправляемые территории привело к существенному возрастанию значения субъекта Федерации как объекта статистического анализа. На протяжении ряда лет стала формироваться региональная статистика, центральным звеном которой является создание системы статистических показателей, разрабатываемых с учетом особенностей развития отдельных регионов.

В основу разработки ЭКСП регионального уровня должны быть положены такие принципы, как обеспечение единства методологии и критериев построения статистических показателей, их сопоставимости с аналогичными показателями в системе государственной централизованной статистики, а также удовлетворение потребностей местных органов самоуправления в качественной оперативной информации, отражающей основные социально-экономические явления и процессы, происходящие на территории.

С одной стороны, программно-технологические средства формирования и администрирования ЭКСП должны обеспечивать работу с базой как информационно-справочной системой метаинформации. С другой стороны, база ЭКСП должна формироваться как территориально-распределенная база статистических показателей, ко-

торая бы включала показатели для различных уровней власти: федеральный уровень, уровень федерального округа, уровень региона, районный (муниципальный) уровень. В зависимости от конкретного региона перечни показателей одного уровня не должны существенно различаться. Отличаться может только степень наполнения этих показателей, зависящая от специфики региона. Тогда ЭКСП будет служить основой для формирования единого информационно-статистического пространства и использования информационных ресурсов системы Госкомстата России.

Многие территориальные комитеты государственной статистики уже используют действующую федеральную базу «Электронный каталог статистических показателей» для проектирования своих информационных ресурсов (микро- и макробаз данных).

Разработка ПТС ЭКСП регионального уровня позволит проектировать базы данных показателей отдельных регионов и осуществлять обработку региональных пока-

зателей (агрегацию, расчеты, контроль, формирование выходных таблиц) и получать информационно-справочные оперативные запросы с использованием унифицированных программно-технологических средств, разрабатываемых НИПИстатинформом и внедряемых в систему Госкомстата России.

Создание информационных ресурсов на базе единого Электронного каталога статистических показателей позволит наиболее эффективно решать основные задачи управления и прогнозирования во всех сферах, включая анализ состояния и динамики развития национальной экономики, финансовой системы, демографической и социальной ситуации.

В частности, создание развитой информационной базы в регионах и на муниципальном уровне позволит осуществлять комплексный анализ совокупности происходящих в этих территориальных образованиях социальных, экономических и технологических процессов, необходимых для своевременного принятия решений.

ПРОЕКТИРОВАНИЕ ИНФОРМАЦИОННЫХ ХРАНИЛИЩ ДЛЯ СТАТИСТИЧЕСКИХ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ СИСТЕМ

Ю.Ф. Тельнов, канд. экон. наук,

Московский государственный университет экономики, статистики и информатики (МЭСИ)

Одним из эффективных инструментов поддержки принятия управленческих решений являются информационно-аналитические системы (ИАС), которые аккумулируют и эффективно обрабатывают большие объемы статистической информации. При этом используются методы и средства современных информационных хранилищ (Data Warehouse), а именно программные средства: оперативного анализа данных (OLAP), интеллектуального анализа данных (Data Mining), методов статистического и математического моделирования.

Достоинства применения ИАС заключаются в следующем:

- достигается интеграция данных в масштабе всех основных функций управления;
- обеспечиваются сложные аналитические запросы в различных разрезах: видов продукции и услуг, отраслей, регионов, технологий и т. д.;
- реализуется оперативный анализ эффективности финансово-хозяйственной деятельности предприятий с учетом отраслевых тенденций;
- устанавливаются корреляции, причинно-следственные и динамические связи экономических показателей;
- прогнозируется развитие предприятий и отраслей на длительную перспективу.

Вместе с тем большие объемы разнообразной аналитической информации, интегрируемой в информацион-

ных хранилищах, предъявляют серьезные требования к эффективной структурной организации данных статистических информационно-аналитических систем. В связи с этим возрастает значение разработки технологии проектирования информационных хранилищ.

Особенности архитектуры информационных хранилищ

Информационное хранилище (ИХ) представляет собой базу обобщенной информации, формируемую из множества внешних и внутренних источников, на основе которой выполняются статистические группировки и интеллектуальный анализ данных [5]. По сравнению с базами данных для оперативной обработки транзакций ИХ обеспечивают более гибкое и простое формирование произвольных справочно-аналитических запросов, а также применение специализированных методов статистического и интеллектуального анализа данных.

В основе информационного хранилища лежит понятие многомерного информационного пространства, или гиперкуба, в ячейках которого хранятся анализируемые числовые показатели (например, объемы оборота, издержек, инвестиций и т. д.). Измерениями (осями) гиперкуба являются признаки анализа, например, время, группа продукции, регион, тип процесса, тип клиента и др. При хранении признаки анализа отделяются от фактических данных, об-

разу так называемую инвертированную организацию хранения данных, или структуру данных типа «звезда».

К особенностям организации хранимой информации в ИХ относятся [1-4]:

- интеграция, или обобщение, данных в ИХ из транзакционных баз данных по всем основным процессам предприятий в виде единого многомерного информационного пространства. Например, ИХ аккумулирует показатели объемов производства, сбыта, сервиса в продуктовом, территориальном, отраслевом, временном и других разрезах;
- произвольность агрегации данных на основе отделения от фактических данных независимых и равноправных измерений информационного пространства (признаков анализа информации, разрезов) в виде иерархий агрегации. Например, региональный признак представляется в виде иерархии агрегации «область - район - город - село», а временной признак - «год - квартал - месяц - день»;

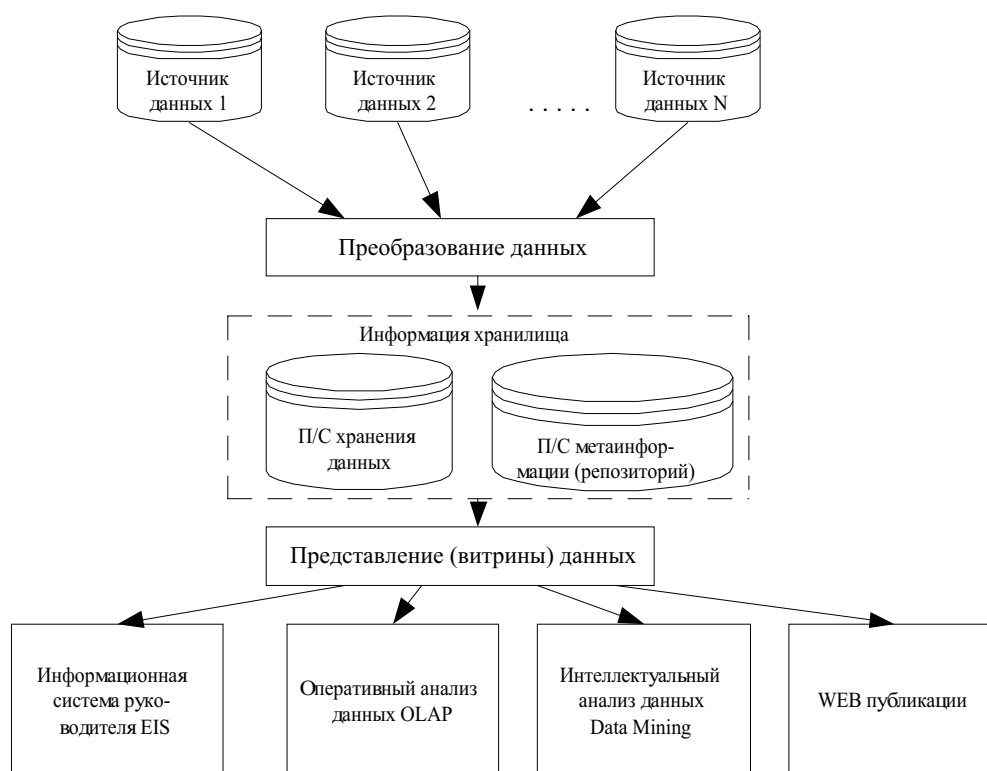
- обязательное хранение временного признака в данных, дающего возможность отслеживать динамику изменения показателей в течение длительного периода времени;

- непротиворечивость данных во всех используемых источниках в течение определенного периода времени, например дня, которая позволяет обеспечить единую точку зрения всех пользователей на экономическую систему;

- обеспечение множества представлений структуры информационного хранилища для различных категорий пользователей: руководителей, аналитиков, менеджеров направлений деятельности. Отбор набора показателей и признаков анализа определяет предметную ориентированность информационного хранилища или организацию витрин данных.

Архитектура системы оперативного анализа данных представлена на рис. 1.

Рис. 1. Архитектура информационного хранилища



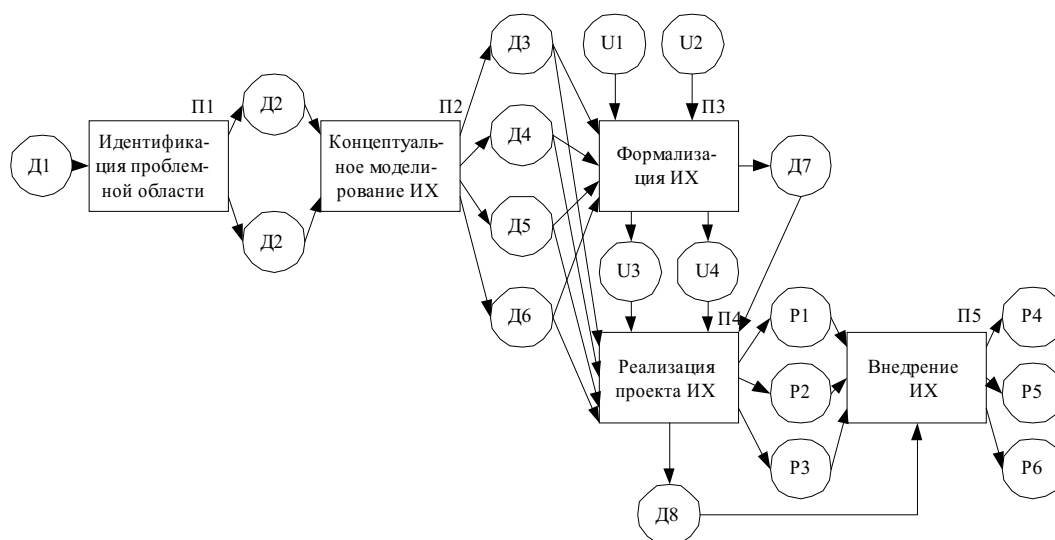
Технология проектирования информационного хранилища

Интеграция множества источников данных в рамках единого информационного хранилища представляет собой трудоемкую и дорогостоящую проектную задачу. Поэтому к процессу проектирования ИХ в наибольшей степени относятся требования: очередности внедрения компонентов ИХ, обеспечивающей быструю отдачу от внедрения, и адаптивности логической и физической структуры ИХ к изменяющимся в ходе проектирования и эксплуатации информационным потребностям. Рассмотрим технологию

ческую сеть проектирования информационного хранилища (см. рис. 2).

П1. Идентификация проблемной области. На основе материалов предпроектного обследования (Д1) осуществляется параметризация проекта создания информационного хранилища и выделяются все необходимые материальные, финансовые, людские и временные ресурсы на выполнение проектных работ, то есть составляется техническое задание (Д2) и технико-экономическое обоснование проекта (Д3). В частности, в рамках технического задания в разрезе конкретных видов деятельности формулируются цели и задачи, области применения и пользователи

Рис. 2. Технологическая сеть проектирования информационного хранилища



ИХ, устанавливаются источники исходных данных, определяются информационные потребности пользователей.

Цели и задачи:

Цели построения информационного хранилища во многом определяют характер используемых источников данных, направлений и методов анализа извлекаемой информации. В качестве цели создания ИХ обычно выступает повышение качества и оперативности принятия управленческих решений на стратегическом, тактическом и оперативном уровнях.

К важнейшим задачам, которые решаются с помощью ИХ, относятся:

- обоснование принятия стратегических решений;
- контроллинг - анализ финансово-хозяйственной деятельности и выявление резервов совершенствования процессов финансово-хозяйственной деятельности предприятий;
- оперативный мониторинг и сравнительный анализ (benchmarking) важнейших показателей деятельности предприятий в отраслевом и межотраслевом аспектах.

Круг пользователей: руководители; референты руководителей, подготавливающие информацию для принятия решений; менеджеры; аналитики.

Области применения: анализ и прогнозирование экономических показателей финансово-хозяйственной деятельности предприятий и организаций в разрезах типов потребителей, продуктов, используемых технологий, каналов распределения, направлений функциональной деятельности и др.

Перечень источников данных:

- внутренние источники: базы оперативных данных об объемах продаж, производства, закупок, издержках по центрам затрат, состоянии материальных, финансовых, людских ресурсов;
- внешние источники: официальные статистические данные о деятельности отрасли, смежных отраслей, о состоянии финансов; нормативная государственная информация; маркетинговая информация о зондировании рын-

ка, состоянии конкурентов; коммерческие базы данных специализированных компаний в области информационного бизнеса.

Для каждого источника данных определяются параметры: территориальное расположение, административное подчинение, периодичность обновления, конфиденциальность и достоверность хранимой информации, форматы данных и характеристики программно-технической среды, объемы данных.

Информационные потребности пользователей. Для обоснования информационных потребностей выполняется анализ конкретных функций управления, например планирования, бюджетирования, маркетинга и т. д. В результате выявляется перечень регламентированных информационно-справочных документов и предполагаемых направлений формирования произвольных запросов.

П2. Разработка концептуальной модели ИХ. Этап разработки концептуальной модели ИХ соответствует этапу логического проектирования, который выполняется на основе технического задания (Д2) и технико-экономического обоснования (Д3). На выходе этого этапа получается логическая структура данных ИХ (Д4), схема преобразования данных (Д5), логическая структура данных витрин (Д6) и схема представления данных (Д7).

Проектирование логической структуры ИХ осуществляется на основе анализа статистики использования конкретных информационно-справочных документов в процессе решения основных задач принятия решений. В результате выполнения операции производятся:

- отбор признаков анализа;
- построение схем агрегации показателей;
- построение схем обобщения признаков;
- определение временного горизонта хранения показателей;
- отбор первичных и производных показателей для хранения;
- выбор типа логической структуры ИХ;
- распределение показателей по типам логической

структуры.

Основными методами выполнения операции отбора и структуризации показателей и признаков являются матричные, графо-аналитические и тезаурусные методы. В частности, большое значение имеет формирование объемно-частотных характеристик использования типов показателей и признаков их группировки в различных типах информационно-справочных запросов. На этой операции происходит также обобщение непосредственно сформулированных пользователями типов запросов к ИХ.

Сложность структуры данных показателей предопределяет выбор ее типа: «звезды» с однородной структурой признаков для всех показателей или «расширенной снежинки» с применением нескольких типов хранилищ показателей. В последнем случае осуществляется распределение показателей по типам хранилищ.

Проектирование процессов извлечения и схемы преобразования данных производится путем анализа выявленных на этапе идентификации проблемной области источников данных. На выходе операции формируется уточненный состав источников данных с определенными схемами фильтрации и агрегации данных для помещения в ИХ.

В частности, на этом этапе осуществляется анализ альтернативных источников данных, например, выбор из числа коммерческих баз данных, а также устанавливаются схемы преобразований исходных данных в хранимые структуры ИХ. Сложность схем отображения источников данных в структуру хранилища предопределяет выбор типа структуры ИХ: MOLAP (многомерная), ROLAP (реляционная), HOLAP (гибридная).

Проектирование логической структуры витрин и схемы представления данных предполагает распределение показателей вместе с измерениями по витринам данных на основе выявленных информационных потребностей пользователей. Для витрин данных точно также, как и для информационных хранилищ, проектируется структура данных и устанавливается схема отображения структуры ИХ на структуры витрин.

Данная операция может предшествовать разработке структуры информационного хранилища, когда сначала создаются структуры данных витрин данных, например, по основным видам деятельности, а затем эти структуры данных интегрируются в общую структуру ИХ.

В рамках логически спроектированных витрин данных осуществляется выбор методов анализа данных для конкретных категорий пользователей. В частности, выявляется потребность в применении определенных видов статистического и интеллектуального анализа данных.

П3. Формализация ИХ. Этап формализации завершает техническое проектирование информационного хранилища. На основе спроектированной на предшествующей операции архитектуры ИХ (Д3-Д6) и универсумов программно-технических средств (U1-U2) осуществляется выбор схемы размещения ИХ в сетевой вычислительной среде (Д7) и программно-технических средств реализации ИХ (U3-U4).

Выбор схемы размещения ИХ в сетевой вычислительной среде осуществляется в зависимости от выбранного

типа организации и предполагает определение числа уровней хранения:

- структура данных реализована централизованно на одном MOLAP-сервере;
- структура данных распределена на нескольких серверах в соответствии с ROLAP-организацией;
- наиболее оперативные и агрегированные данные хранятся на быстродействующем MOLAP-сервере, а детальные данные - в ROLAP-хранилище на менее производительных серверах.

Определение требований к конфигурации и числа клиентских мест выполняется на основе структуры витрин данных, выявленных категорий пользователей и используемых методов интеллектуального анализа, которые в совокупности определяют требования подключения к OLAP-серверу. Для каждого пользователя устанавливаются права доступа к ИХ.

Выбор программно-технических средств ИХ (серверов, клиентских мест, телекоммуникационного оборудования, инструментальных программных средств) выполняется на основе требований к физической конфигурации системы в части объемов памяти, быстродействия, надежности и выбранной клиент-серверной архитектуры ИХ.

Расчет объемов ИХ осуществляется путем суммирования объемов хранимых данных на всех MOLAP-серверах с учетом необходимого индексирования (специальных индексирующих таблиц для доступа к основным данным), а также объемов метаданных репозитория для MOLAP и ROLAP-организации. Объемы ИХ рассчитываются на текущий момент времени и на перспективу с учетом внедрения всех компонентов системы.

П4. Реализация проекта ИХ. Этап реализации проекта ИХ выполняется на основе выбранных программных (U3) и технических средств (U4), а также построенных на этапе концептуального моделирования компонентов ИХ (Д3-Д6) и схемы размещения ИХ (Д7) путем наполнения репозитория (Р1), настройки или программирования других инструментальных средств (Р2), наполнения информационного хранилища для MOLAP-структуры (Р3), создания проектной документации (Д8).

Наполнение репозитория ИХ осуществляется путем ввода определений:

- структуры ИХ, источников и витрин данных;
- правил ввода данных в ИХ из одного источника, из нескольких источников, при отсутствии данных;
- правил преобразования форматов при поступлении данных из источника и при выводе данных в представление пользователя;
- параметров использования методов интеллектуального анализа данных.

Разработка и отладка программных компонентов производятся в основном путем параметрической настройки ППП. В случае функциональной неполноты выбранного инструментального программного средства в части процедур начальной и периодической загрузки данных, а также процедур анализа данных выполняется программирование отдельных программных модулей.

Наполнение ИХ предполагает автоматическую загрузку

ку информации из источников данных в ИХ с MOLAP - организацией, которая повторяется с заданной в репозитории периодичностью. Эта операция в последующем предполагает очистку ИХ от ненужных и устаревших данных; управление данными на различных уровнях хранения; автоматическое обновление агрегированных данных.

П5. Внедрение и опытная эксплуатация. Заключительный этап создания ИХ предполагает комплексное тестирование всех компонентов ИХ (Р1-Р3) с исправлением всех возникающих ошибок (Р4-Р6), последующим обучением пользователей и постоянным администрированием в соответствии с установленными правилами и документацией проекта (Д8).

Заключение. Представленная технология проектирования информационных хранилищ позволяет сократить затраты на создание и поддержание статистической информационно-аналитической системы, что обеспечивается применением системного подхода, который исключает несогласованность формулируемых требований на концеп-

туальной стадии разработки и при их реализации на последующих стадиях. Кроме того, обеспечивается повышение качества принятия управленческих решений на стадии эксплуатации за счет постоянной готовности статистической информационно-аналитической системы адекватно отражать состояние объектов управления и своевременно выявлять тенденции в их развитии.

Литература

1. **Львов В.** Создание систем поддержки принятия решений на основе хранилищ данных. // Системы управления базами данных. 1997. № 3.
2. **Спирли Эрик.** Корпоративные хранилища данных. Планирование, разработка, реализация. Том. 1. Пер. с англ. - М.: Издательский дом «Вильямс», 2001.
3. **Тельнов Ю.Ф.** Интеллектуальные информационные системы в экономике. - М.: СИНТЕГ, 2002.
4. **Codd E.F., Codd S.B., Salley C.T.** Providing OLAP to User-Analysts: An IT Mandate, Arbore Software Corp. Papers, 1996.
5. **Inmon W.** Building the Data Warehouse. - New York: John Wiley & Sons, 1992.

ПЕРСПЕКТИВНЫЕ РЕШЕНИЯ В ОБЛАСТИ СЕТЕВЫХ ТЕХНОЛОГИЙ

Ю.К. Голованов,

Г.Ф. Супрун,

Госкомстат России,

В.А. Безруков,

Ростовский областной комитет государственной статистики

Локальные вычислительные сети уже давно стали основой для автоматизации обработки статистической информации на федеральном и региональном уровнях системы государственной статистики, внутри них созданы и функционируют разнообразные информационные ресурсы, но все это доступно для пользователей данной ЛВС и обмен информацией между такими сетями осуществляется с помощью электронной почты с использованием низкоскоростных аналоговых каналов связи. Это резко ограничивает возможности пользователей по доступу к информационным ресурсам, применению современных информационных технологий обработки и распространения статистической информации. Качественно новый уровень работы предоставляют глобальные, корпоративные вычислительные сети, и вопрос организации корпоративной сети Госкомстата России, объединяющей все вычислительные и информационные ресурсы федерального уровня и территориальных органов государственной статистики, встал в числе первоочередных задач, что и отмечено в Концепции развития ИВС Госкомстата России.

Развитие корпоративной сети Госкомстата России требует решения многих сложных задач, опыта решения которых в органах статистики еще нет. Одной из таких задач является выбор каналов связи, с помощью которых можно было бы объединить локальные вычислительные сети

Госкомстата России, ГМЦ Госкомстата России и территориальных органов государственной статистики в единую сеть. Анализ зарубежного опыта создания и эксплуатации глобальных сетей показывает, что они предоставляют возможности не только для различных технологий сбора, обработки информации, но и позволяют решить на качественно новом уровне вопросы голосовой и видеосвязи и, таким образом, более эффективно использовать каналы связи, на которых строится информационно-вычислительная сеть.

С целью решения поставленных задач в 2002 г. в рамках программы ТАСИС был реализован пилотный проект по созданию фрагмента защищенной корпоративной сети на базе цифровых каналов связи сети передачи данных «Атлас». В проекте принимали участие Госкомстат России, Ростовский облкомстат и исполнитель работ ЗАО «Комфакс», выбранный на конкурсной основе.

Пилотным проектом предусматривалось решение следующих задач:

- разработка и реализация типовых схем телекоммуникационных узлов ИВС Госкомстата России федерального и регионального уровней;

- отработка технологий телекоммуникационного взаимодействия информационных систем Госкомстата России различных уровней и первичных источников статистиче-

ской информации на фрагменте телекоммуникационной сети;

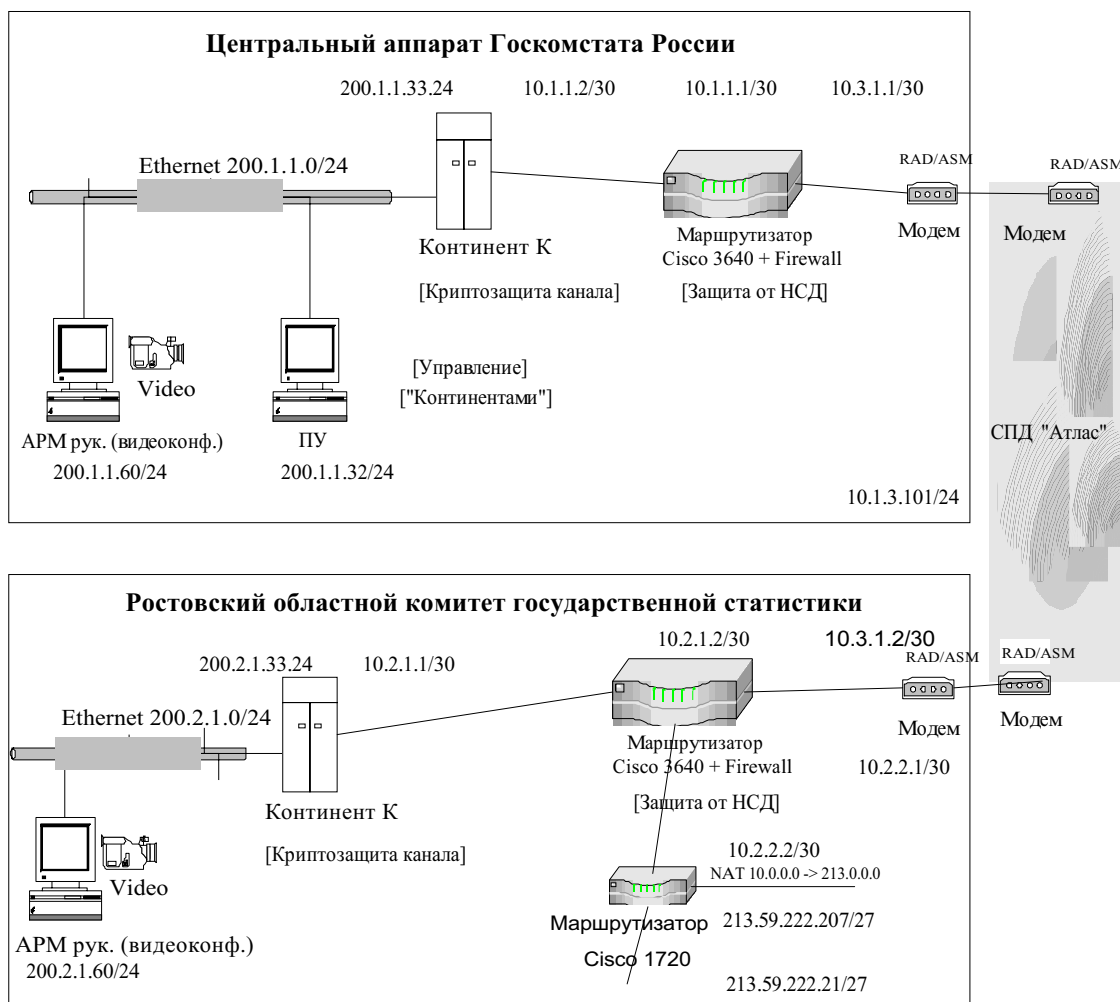
- создание модели Интранет-сети ограниченного доступа с использованием цифровых каналов связи;
- отработка технологии удаленного доступа к базам данных;
- решение вопросов защиты информации при приеме-передаче данных по цифровым каналам связи;
- внедрение системы видеоконференцсвязи.

Фрагмент корпоративной сети (см. рисунок) был организован из телекоммуникационного узла центрального аппарата Госкомстата России на базе маршрутизатора Cisco 3640, телекоммуникационного узла Ростовского облкомстата на базе маршрутизатора Cisco 3640. К маршрутизаторам были подключены модемы RAD ASMi-50, к ко-

торым подключена выделенная линия, соединяющая их с СПД «Атлас». Данное оборудование обеспечивало пропускную способность канала до 1,5 Мбит/сек., но реальное подключение было на скорости до 512 Кбит/сек.

Для защиты передаваемой информации по цифровому каналу связи в сегменте локальной вычислительной сети использовались программно-аппаратные комплексы криптографической защиты информации «Континет-К», которые обеспечивали шифрование всей поступающей из ЛВС в цифровой канал информации и ее дешифрацию при передаче из канала в ЛВС. Таким образом, был решен вопрос защиты любой передаваемой по каналу связи информации от несанкционированного доступа и обеспечения конфиденциальности переговоров во время проведения сеансов видеосвязи.

Рисунок. Схема изолированного сегмента макета защищенной телекоммуникационной сети Госкомстата России, моделирующей системы видеоконференцсвязи, передачи файлов и удаленного доступа к региональным информационным ресурсам



Моделирование работы электронной почты в корпоративной сети проводилось на примере создания одноранговой сети, состоящей из двух рабочих станций: одной в Госкомстате России, второй - в Ростовском облкомстате. На рабочей станции в Госкомстате России был развернут

почтовый узел электронной почты, и таким образом она моделировала почтовый сервер, а на рабочей станции в Ростовском облкомстате установлен локальный абонент электронной почты. Это позволило осуществить работу электронной почты с использованием цифрового канала

связи. В процессе моделирования производилась передача информации в обоих направлениях штатными средствами электронной почты, которая прошла успешно и показала высокие скоростные характеристики, как минимум на порядок превышающие скоростные характеристики при работе в режиме удаленного абонента по существующим каналам связи.

Для организации видеоконференцсвязи использовалась система VCON. Данная система обладает удобным интерфейсом и средствами управления сеансом видеосвязи, включая и возможность управления видеокамерой удаленного абонента, что очень удобно при проведении различного рода совещаний, презентаций с участием нескольких лиц и демонстрацией каких-либо макетов, стендов, документов и т. д. Подключение к рабочей станции мультимедиа проектора во время сеанса видеосвязи позволило всем участникам видеоконференции наблюдать за происходящими событиями.

Была проведена серия опытных сеансов видеосвязи, во время которых оценивались качество и надежность работы системы. В июле 2002 г. был организован сеанс видеоконференцсвязи председателя Госкомстата России с руководством Ростовского облкомстата о ходе работ по под-

готовке к проведению ВПН-2002. Во время всех сеансов канал связи работал устойчиво и его пропускная способность обеспечивала синхронную передачу видео- и голосовой информации. Использование системы видеоконференцсвязи показало, что она позволяет более оперативно решать вопросы в условиях территориальной разобщенности подразделений, задействованных в решении общей задачи.

Полученные в ходе пилотного проекта результаты выявили существенные преимущества по сравнению с действующей телекоммуникационной системой Госкомстата России. Используемые технологии передачи интегрального трафика в случае их внедрения в ИВС Госкомстата России обеспечат пользователям различные информационные сервисы и позволят существенно сократить время на сбор и передачу данных между федеральным и региональным уровнями, организовать удаленный доступ к информационным ресурсам, систему ведомственной IP-телефонии, в том числе в защищенном режиме, а также организовать видеоконференцсвязь между Госкомстатом России и территориальными органами государственной статистики и дистанционное обучение сотрудников системы Госкомстата России.

ВНИМАНИЮ АВТОРОВ, ГОТОВЯЩИХ СТАТЬИ К ПУБЛИКАЦИИ!

Редакция принимает к рассмотрению материалы общим объемом до 0,5 авторского листа на дискете (Word любой версии) с приложением распечатанного текста в одном экземпляре.

В тексте не должно быть двойных пробелов и пустых строк. Ни в коем случае не делать красную строку у абзаца несколькими пробелами - только абзачными отступами. Можно делать необходимые смысловые выделения **полужирным** и **курсивным** начертаниями, которые при верстке, как правило, сохраняются. **Не рекомендуется выделять текст подчеркиванием**, поскольку подчеркнутый текст зрительно засоряет страницу и мешает нормальному восприятию материала.

Оформляя **таблицы**, нужно стремиться к тому, чтобы в одной ячейке не было разных по смыслу строк, то есть каждому «пункту» таблицы должна соответствовать своя строка ячеек. Несколько строк таблицы лучше объединять, делая невидимыми границы ячеек. Все видимые границы в таблицах лучше делать сплошной черной линией толщиной 0,5 пункта. Выравнивание текста и цифр внутри ячеек необходимо выполнять только стандартными способами, а не с помощью пробелов или дополнительных пустых строк.

Все фреймы (рамки с текстом) необходимо убрать, так как после удаления фреймов (рамок) находящийся в них текст может попасть в другое место.

Если графические иллюстрации делаются в Word, **все элементы рисунка должны быть сгруппированы** внутри единого объекта, иначе при малейшем сдвиге границ страницы, неизбежном при работе с документом, элементы могут сместиться.

Настоятельно рекомендуется авторам ничего не сканировать самим: в 90% материал требует переработки. **Для графиков и диаграмм обязательно предоставлять таблицы цифровых данных** (часто диаграммы приходится перестраивать заново).

При небольшом количестве иллюстративного материала проще применять обычную сквозную нумерацию иллюстраций (Таблица 1, Таблица 2 или Рис. 1, Рис. 2, Рис. 3 и т. п.).

Электронную версию фотографий следует передавать в формате TIFF максимально возможного разрешения и глубины цветопередачи (300 dpi, CMYK, без альфа-каналов и компрессии). Не нужно пытаться сжимать файл в формат JPEG для уменьшения объема. Потеря качества при этом может оказаться для полиграфии совершенно невосполнимой.

Грамотно подготовленный к верстке текст ускоряет его публикацию.

К статье должны быть приложены краткая справка об авторах с указанием места работы, должности, ученого звания и степени, адреса постоянного места жительства и контактного телефона.

Редакция журнала рукописи не рецензирует и не возвращает.

Справки по телефону: (095) 207-48-82

Тел./факс: (095) 207-48-90

E-mail: voprstat@mtu-net.ru