

УДК 004.6

Л. С. Сорока, доктор технических наук, ректор
Академии таможенной службы Украины
В. И. Есин, кандидат технических наук, доцент
кафедры безопасности информационных систем
и технологий Харьковского национального
университета им. В. Н. Каразина

УПРАВЛЕНИЕ ИНФОРМАЦИЕЙ В ПРОСТРАНСТВЕ ДАННЫХ

Описано проблему управління інформацією в просторі даних, підходи до її розв'язання з їх перевагами й недоліками. Пропонується новий метод управління інформацією в просторі даних, що базується на використанні універсальної моделі даних і платформи для автоматизації спільної діяльності робочих груп.

Описываются проблема управления информацией в пространстве данных, существующие подходы к ее решению с их достоинствами и недостатками. Предлагается новый метод управления информацией в пространстве данных, основанный на использовании универсальной модели данных и платформы для автоматизации совместной деятельности рабочих групп.

The problem of management of information in the data space, the existing approaches to its solving with their advantages and disadvantages is described. The new method of management of information in the data space is offered. He is based on the use of universal data model and platform for automation of the joint activity of working groups (groupware).

Ключевые слова. Пространство данных, источник данных, универсальная модель данных, база данных с универсальной моделью данных.

Введение. Сегодня различные организации, учреждения, компании накопили значительный объем данных, располагающихся во всевозможных архивах систем управления оборудованием, в транзакционных СУБД, в офисных документах, в документальных системах и так далее, и обеспечивающих их деятельность. Причем число, объем, географическое расположение (согласно принципу глобализации [1]: “источники и потребители информации будут находиться в любой географической точке, где это окажется нужно”) и стоимость таких ресурсов продолжает стремительно расти. При этом данные, как правило, имея значительную практическую ценность, не согласованы. Более того, один и тот же объект в разных системах и приложениях может быть идентифицирован по-разному. Как связать такие данные? Как определить, какие из них относятся к одинаковым понятиям и объектам, а какие – к разным? Какие данные, когда, кем и для чего используются или будут использоваться и т. д.? На эти простые вопросы не просто ответить, имея разрозненные источники даже самых подробных данных. Таким образом, можно говорить о существовании проблемы управления данными в имеющихся информационных ресурсах.

Все эти и другие подобные причины явились предпосылками для широкого проведения в последние годы исследований и разработок в области интеграции неоднородных информационных ресурсов, к числу которых можно отнести [2, 3]: а) разработку архитектур системы интеграции данных; б) разработку интегрирующих моделей данных; в) разработку методов отображения моделей данных и построение отображений в интегрирующую модель для конкретных моделей, поддерживаемых отдельными источниками данных; г) интеграцию метаданных, используемых в системе источников данных; д) преодоление неоднородности источников данных (путем разработки адаптеров – компонентов, обеспечивающих интероперабельность интегрируемых неоднородных информационных ресурсов и создания посредников – компонентов, обеспечивающих семантическую интеграцию информационных ресурсов) и т. д.

© Л. С. Сорока, В. И. Есин, 2012

Так, в работе [4] проблема совместного, согласованного, а значит, непротиворечивого, использования данных “устаревших” информационных систем решалась путем проведения реинжиниринга последних и размещения данных этих систем в базу данных (БД) с универсальной моделью данных (УМД). Благодаря чему стало возможным использование неоднородных информационных ресурсов “устаревших” информационных систем, функционирующих на разных аппаратных платформах и под управлением различных операционных систем и иных программных сред, в системе, основанной на новой технологии.

В свое время в проекте “Синтез” [5] также приходилось решать подобную проблему путем создания интероперабельных сред информационных ресурсов. И тогда неоднородность информационных ресурсов компании как с точки зрения применяемых моделей данных, языков программирования, интерфейсов пакетов прикладных программ, так и с точки зрения операционного окружения предопределяла потребность использования специфических методов, обеспечивающих возможность их повторного использования в новых применениях.

Кроме того, данную проблему можно также решить путем интеграции данных в любое единое хранилище, обеспечивающее контролируемый доступ.

Однако такие подходы решения проблемы совместного согласованного использования имеющихся информационных ресурсов не всегда устраивают заказчика. Последний, например, хочет по различным

соображениям: от не имеющих свободных ресурсов (временных, человеческих, финансовых) до тривиального нежелания (по привычке) продолжать эксплуатацию всех своих имеющихся систем.

Острые проблемы управления информацией возникают сегодня не только в компаниях, использующих “устаревшие” ИС, но и в компаниях, имеющих в своем распоряжении большое число разнотипных источников данных, построенных на современных платформах и технологиях. И они не всегда имеют возможности управлять такими ресурсами удобным, интегрированным и обоснованным способом. Им приходится сталкиваться с набором слабо связанных источников данных (баз данных, web-сервисов, архивов систем управления оборудованием, систем электронного документооборота, офисных документов и т. д.), образующих так называемое пространство данных [6]. В этих компаниях вынуждены каждый раз решать повторяющиеся низкоуровневые задачи управления данными, например такие как:

- обеспечение возможностей поиска и запрашивания данных;
- соблюдение правил, ограничений целостности, соглашений об именовании; отслеживание происхождения данных;
- обеспечение доступности, восстановления и контроля доступа; управляемое развитие данных и метаданных и т. д. То есть подобные пространства данных требуют создания поверх существующих локальных систем (источников данных) дополнительного уровня программного обеспечения, предназначенного для предоставления необходимых функциональных возможностей.

В [6] авторы предлагают некоторый подход по проектированию и разработке платформ поддержки пространств данных (*Data Space Support Platforms – DSSP*), хотя не раскрывают его конкретной реализации.

Занимаясь проблемой Единого научного информационного пространства (ЕНИП) [7], российские ученые разработали и реализовали свой подход. Его основу составляет инструментальное ядро, которое обеспечивает обмен данными, интеграцию данных и поддержку распределенности. В первой версии реализация этого подхода называлась ИСИР (Интегрированная система информационных ресурсов). В основе инструментального ядра тогда лежала реляционная модель данных. В следующей версии системы в основе формальной модели данных ядра лежала модель данных RDFS (*Resource Definition Framework Schema*).

Авторы [7], говоря о своем подходе, не раскрывают его реализации, хотя выделяют и стараются решить проблемы общего вида, среди которых можно выделить следующие: техническую, синтаксическую, семантическую интероперабельность, использование и сбор метаданных, поддержку глобальной идентификации ресурсов, балансировку нагрузки, распределенную авторизацию доступа и принцип единой аутентификации.

Постановка задачи. Проведенный анализ рассмотренных выше подходов и требований к динамике, разнообразию типов информационных потоков, обрабатываемых в едином пространстве данных компании (ЕПДК), с учетом роста их объемов, а также требований к согласованности и разнообразию методов обработки привел к необходимости разработки собственного подхода (метода) построения и организации функционирования ЕПДК.

Результаты исследования. Предлагаемый метод, кроме разнообразных типов источников информации и приложений, работающих с ними, предполагает также использование:

- общих понятийных моделей (понятийного пространства предметной области – ПрО) в виде метаонтологий модели “объект-событие” [8], нормативно-справочной информации (НСИ) и системы классификации и кодирования (СКК);
- заранее определенных стандартных интерфейсов и механизмов доступа к различным информационным ресурсам;
- технологий хранилищ данных, интегрирующих исторические форматированные данные, архивные текстовые документы, звуковые и видеоархивы, а также картографические и прочие данные, с привлечением средств оперативной аналитической обработки;
- возможностей схемы базы данных с универсальной моделью и языка модели данных (ЯМД) [9];
- возможностей программного обеспечения совместной работы групп (*groupware*).

До сих пор часто встречается мнение, что СКК – это средство сокращенного представления информации в интегрированной БД. На самом деле отсутствие СКК или использование некорректно построенных СКК приводит к смысловой несовместимости информации, хранимой в различных БД или даже в одной БД [1]. То есть к СКК надо относиться как к началу и основе создания понятийного пространства ПрО. Использование общих понятийных моделей (понятийного пространства ПрО) обеспечивает семантическую интероперабельность пространства данных компании.

Одной из главных проблем эффективного использования, находящейся в ЕПДК информации, является правильно организованный обмен данными между различными его подсистемами. Нередко самая простая задача импорта/экспорта данных из одной системы в другую приводит к необходимости серьезных разработок модулей на стыке подсистем [10]. Задача существенно облегчается, если имеются заранее разработанные стандартные интерфейсы импорта/экспорта для различных типов источников данных. Так как именно такие стандартные интерфейсы и механизмы доступа обеспечивают свободный обмен между разнородными источниками информации ЕПДК, переносимость, масштабируемость и так далее, тем самым обеспечивая техническую интероперабельность пространства данных.

Технология хранилищ данных (ХД) позволяет решить проблемы, возникающие при интеграции данных распределенных и гетерогенных системы при внедрении методов OLAP [11]. Однако если использовать традиционную технологию построения ХД [12], то можно столкнуться с некоторыми трудностями. В этом случае требуется разработка специальных программ согласования и очистки данных различных объединяемых систем (выполняющих ETL-процедуру). Создание этих программ и их развитие, как правило, уникально для конкретной системы и требует больших затрат (людских, материальных, временных).

Поэтому предлагается воспользоваться технологией ХД, но при этом уже согласованные и очищенные данные вносить в хранилище из так называемой согласующей БД (СБД), построенной на основе универсальной модели данных [13].

Согласующая БД имеет много общего с ХД. СБД также строится как “переводчик” терминов и обеспечивает согласованность данных разных систем. Отличие состоит в динамике формирования СБД. Технология СБД не фиксирует статическое состояние существующих систем, а поддерживает постоянную согласованность с изменяющимися данными, которые появляются в процессе функционирования ИС. Расширения и дополнения в СБД могут проводиться обычным техническим персоналом в рамках стандартных работ по сопровождению и не требуют участия разработчиков программного обеспечения, как это оказывается необходимым при использовании ХД.

Но все же чтобы предоставить пользователям возможность простого и быстрого доступа к разнородным источникам данных ЕПДК в соответствии с их правами необходима какая-то согласующая (интегрирующая) система. В рамках предлагаемого подхода как раз и была разработана такая системы, которая получила название системы согласования данных (ССД).

Решая свою основную задачу – согласование данных, ССД, во-первых, объединяет все справочники и классификаторы, используемые для работы функционирующих приложений, формируя таким образом единую НСИ. Но это не единый стандартизированный справочник для всего информационного пространства компании, полученный путем какого-либо интегрирования (преобразования) имеющихся справочников и классификаторов (внедрение единого нормативного справочника, как правило, препятствует нормальной работе функционирующих приложений). Это некая система соответствия между различными справочниками и классификаторами существующих (и будущих) подсистем подразделений компании. То есть ССД, включая в свой состав все справочники и классификаторы (в ней организовано их хранение), устанавливает соответствие между ее элементами. Эта работа, естественно, должна проводиться экспертами предметных областей до использования элементов того или иного справочника, классификатора, но не обязательно в самом начале построения ЕПДК. Эксперт может выполнять эти действия с помощью специальных приложений по мере поступления задач на своем рабочем месте, которое территориально может быть расположено в любом месте, где есть доступ к ССД.

При таком подходе один элемент данных может включаться в разные классификаторы. Поэтому чтобы избежать конфликтов, когда одно и то же имя используется для различных объектов (омонимы) или различные имена используются для одного и того же объекта (синонимы), ССД должна содержать связи между одинаковыми элементами данных, имеющими разные имена или идентификаторы в разных системах, или, наоборот, указывать на имена (идентификаторы) других объектов при совпадении имени, но не семантики.

Такой подход к построению ССД: 1) исключает риски по нарушению нормального функционирования подсистем подразделений компании, источники данных которых входят в уже созданное ЕПДК; 2) позволяет не прекращать ведение работ по согласованию данных; 3) дает возможность пользователям не менять своих стереотипов при работе со своими приложениями, так как ССД не требует внесения изменений в различные эксплуатируемые приложения, данные которых в определенный момент должны быть согласованы между собой.

Во-вторых, системой согласования данных решаются задачи собственно организации обмена данными между разнотипными источниками, входящими в ЕПДК, и прослеживания связей между ними. Для их решения в рамках создания ССД были предложены: а) форматы документов обмена и спецификации на создание программных средств обмена между различными источниками данных, входящих в ЕПДК; б) стандартные программные интерфейсы обмена между различными источниками данных, входящих в ЕПДК; в) сценарии информационного обмена; г) программы внесения согласующей информации в базу данных ССД (программы формирования соответствия); д) программы обмена.

Основой для построения ССД стала платформа для автоматизации совместной деятельности рабочих групп (*groupware*), которую предлагается реализовать с помощью программного продукта фирмы IBM – *Lotus Notes/Domino*. Это связано с тем, что использование для своих решений платформы *Lotus Notes/Domino* позволяет создавать территориально-распределенные системы, построенные на различных аппаратных и программных платформах (Microsoft Windows (32 и 64 бит), Linux (Red Hat, SuSE), Sun Solaris, IBM i5/OS (OS/400), IBM AIX, IBM z/OS (OS/390), Mac OS X, HP UX), интегрирующие данные различных СУБД и иных источников информации (использующие технологии CORBA, COM, XML, ODBC, OLE и т. д.), с хорошей масштабируемостью, наличием средств криптозащиты, с возможностью организации надежного механизма репликаций (путем обеспечения гарантированной доставки данных потребителю независимо от качества коммуникационных каналов) и т. д.

Структурная схема единого пространства данных компании и приложений, работающих с его данными, в соответствии с предлагаемым методом, изображена на рис. 1.

ЕПДК включает:

- единую БД системы согласования данных;
- множество баз данных подразделений компании, построенных на основе универсальной модели данных;
- множество источников данных различного типа (базы данных, web-сервисы, архивы систем управления оборудованием, системы электронного документооборота, офисные документы и т. д.);
- интерфейсы обмена между БД с УМД и иными источниками информации;
- интерфейс обмена между БД с УМД и БД ССД;
- интерфейсы обмена между БД ССД и иными источниками информации;
- хранилище и витрины данных.

Для работы с данными ЕПДК предполагается использование следующих типов приложений: приложений проектировщиков и пользователей различных БД, в том числе и БД с УМД; приложений пользователей, работающих с витринами и хранилищем данных; приложений для работы с различного рода источниками информации.

В единой базе данных ССД, реализованной на платформе *Lotus Notes/Domino*, хранятся все используемые в компании справочники, словари, классификаторы, включая международные, государственные, отраслевые и так далее, с иерархической вложенностью до 32-го уровня.

Пример представления одного из возможных справочников-классификаторов приведен на рис. 2.

При этом по каждому из объектов такого справочника в БД ССД хранится так называемый документ обмена. Один из вариантов формата документа обмена, хранящегося в БД ССД, приведен на рис. 3.

Также в состав структуры БД ССД входит таблица (возможен вариант с несколькими таблицами – каждая для различного типа источников данных) согласования, в которой (которых) хранится информация о соответствии между элементами данных согласуемых ис точников. А именно в ней (них) хранятся данные о соответствующих источниках, подлежащих согласованию, времени формирования данных от этих источников, датах согласования, ссылках на соответствующие документы обмена, некоторых признаках, позволяющих, например, активизировать требуемые программы-агенты и т. п.

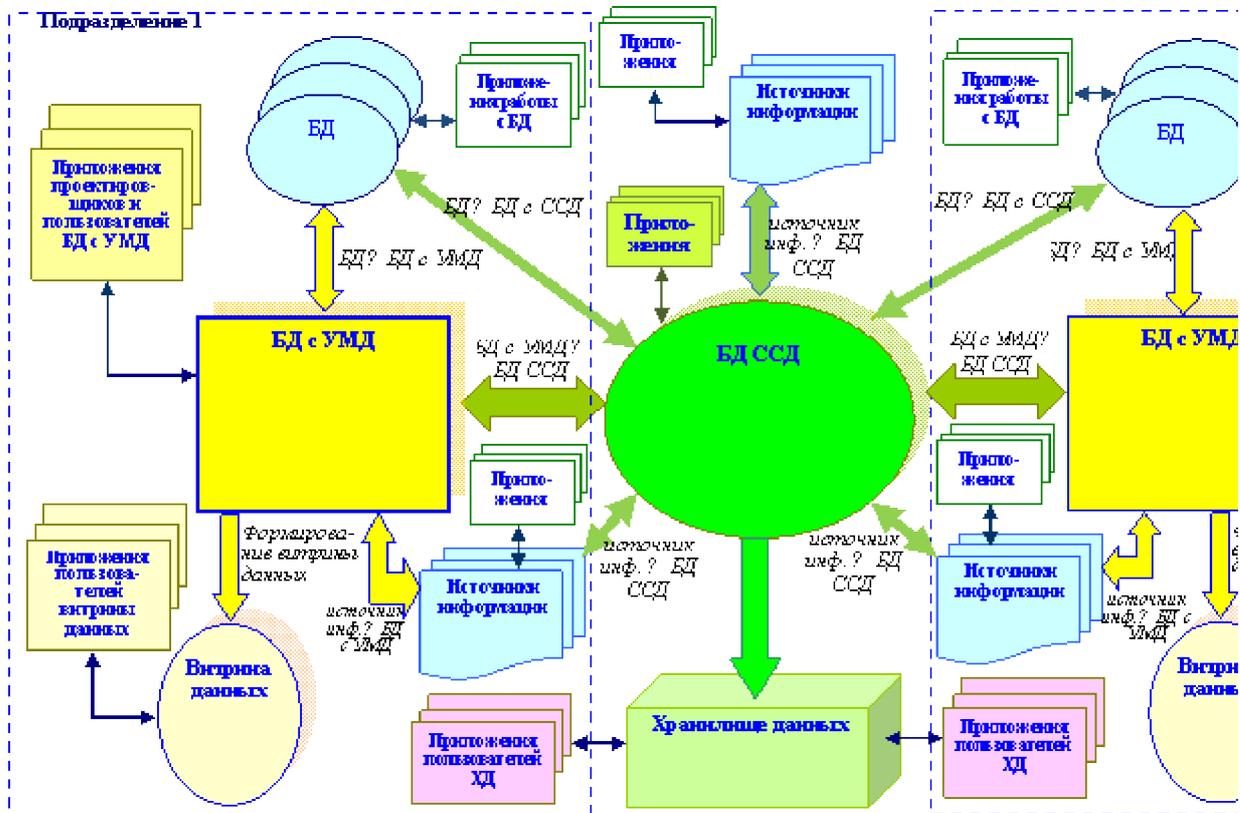


Рис. 1. Структурная схема единого пространства данных компании и приложений, работающих с его данными

Общий справочник	Название	Уровень иерар
<input type="checkbox"/> Классификатор	Огазіе	
<input type="checkbox"/> Соответствие	(4) Донбаэтрангаэ	Главный докум
<input type="checkbox"/> Список	Новопсковське ЛВУМГ	1-й уровень
<input type="checkbox"/> Папка	Новопсков	2-й уровень
	1 (КС 11)	3-й уровень
	2 (КЦ "Волна")	3-й уровень
	(10) Кнївтрангаэ	Главный докум
	Дїканьське ЛВУМГ	1-й уровень
	Дїканька	2-й уровень
	1 (4)	3-й уровень
	Лубенське ЛВУМГ	1-й уровень
	Гребїнкївська	2-й уровень
	1 (КС-33 I)	3-й уровень
	Сумське ЛВУМГ	1-й уровень
	Ромнї	2-й уровень
	2 (КС-3)	3-й уровень
	3 (КС-32 П)	3-й уровень
	(6) Черкаситрангаэ	Главный докум
	Барське ЛВУМГ	1-й уровень
	Бар	2-й уровень
	2 (КС 37)	3-й уровень
	Золотонївське ЛВУМГ	1-й уровень
	Софїївська	2-й уровень
	2 (КС 34 Б)	3-й уровень
	TWA	
	Телефонїй довідник	

Рис. 2. Пример справочника-классификатора

Главный документ	Объект	
	название	УМГ
Донбастрасгаз	значение	Донбастрасгаз
дата создания документа 04.07.2007 15:42	код в системе (если есть)	{<Раздел>=Укртрансгаз; /<КлассО>=УМГ;<ТипО>=-; <ЭкзО>=Донбастрасгаз;}
	источник	Oracle
	дата последней модификации	04.07.2007 15:43
	дата конца существования	

Рис. 3. Вариант формата документа обмена

После того как основные классификаторы, справочники, словари будут занесены в БД ССД, необходимо занести информацию в таблицу (таблицы) согласования, то есть определить связи между соответствующими различными классификаторами с учетом их различного описания в своих подсистемах. И только после этого возможно полноценное использование информации, хранящейся в ЕПДК. При этом по мере необходимости данные, находящиеся в БД ССД, могут и будут корректироваться.

Множество различных источников, образующих ЕПДК, обеспечивают хранение данных именно в тех местах, где они чаще всего используются. Информация в эти источники заносится, модифицируется, удаляется, считывается (при необходимости) с помощью соответствующих приложений, которых может быть много как по количеству, так и по типам (специализированных, широкоиспользуемых и пр.). На рис. 1 они представлены как приложения проектировщиков и пользователей БД с УМД, приложения работы с базами данных, не имеющих структуру УМД, приложения пользователей витрин и хранилищ данных, приложения иных источников данных, приложения, работающие с БД ССД.

На уровне подразделения компании его действующие подсистемы чаще всего работают в обычном “автономном” режиме. Определенное или полное согласование данных различных источников этих подсистем, как правило, требуется при проведении анализа и получения обобщенной информации (например, при составлении отчетности). В этом случае оно требует значительного времени и поэтому, согласно разработанным сценариям обмена данными, осуществляется в так называемом отложенном режиме. То есть откладывается до того самого момента проведения соответствующего анализа. Но если же требуется провести согласование данных, необходимое, например, для работы подсистемы подразделения, которое управляет некоторым процессом, то в этом случае ни о каком отложении согласования не может быть и речи. Оно проводится оперативно. За счет такого подхода обеспечивается возможность эксплуатации подсистем на существующем коммуникационном оборудовании.

Для обеспечения относительно свободного (с учетом прав) общего доступа к данным ЕПДК необходимы соответствующие интерфейсы, которые способствовали бы согласованию различных данных, имеющихся в разнотипных информационных источниках. Так, предлагаемые интерфейсы между БД с УМД и иными источниками информации (рис. 1) способствуют интеграции информации на уровне подразделения компании.

Интерфейсы между БД системы согласования и БД с УМД, между БД ССД и иными источниками информации обеспечивают возможность включения в единую БД ССД различных классификаторов, справочников и словарей соответствующих источников информации (подобных тем, что приведены на рис. 2) для их дальнейшего согласованного использования каждым из пользователей ЕПДК.

Если посмотреть на пример справочника-классификатора, представленного на рис. 2, то можно заметить, что он очень похож на концептуальное описание метаданных соответствующей предметной области в виде связанного иерархического дерева модели “объект-событие”. И это справедливо, так как этот вариант справочника был получен путем его извлечения из схемы БД, построенной на основе УМД и помещенная в БД ССД через интерфейс “БД с УМД ↔ БД ССД”.

Теперь если в ином подразделении компании имеются подсистемы, работающие с данными той же ПрО, и в них есть подобные элементы, но которые определены иначе, то благодаря данному классификатору, таблице (таблицам) согласования с соответствующими документами обмена можно организовать корректное использование таких данных в соответствующих приложениях согласуемых подсистем. Для этого, задействуем различные интерфейсы обмена и воспользуемся несколькими способами согласования.

Рассмотрим два возможных сценария согласования и использования различных интерфейсов обмена: а) требуемые для согласования данные, используемые соответствующими подсистемами того или иного подразделения, находятся не в БД с УМД; б) данные помещены в БД с УМД.

Первый сценарий предполагает два следующих способа согласования.

1. Данные из различных источников всевозможных информационных систем подразделения компании через соответствующие интерфейсы (“БД ↔ БД с УМД” и/или “источник инф. ↔ БД с УМД” – рис. 1) помещаются в согласующую БД, построенную на основе УМД (где предварительно согласуются и очищаются), а затем через интерфейс “БД с УМД ↔ БД ССД” (рис. 1) заносятся в БД ССД. Работа по очистке и согласованию проводится специалистами ПрО с помощью специального программного обеспечения в отложенном режиме (не оперативно). При этом процесс согласования и очистки данных протоколируется так, что пользователь может знать, какой источник был выбран, каким образом и когда производились согласование и очистка данных. Затем уже интегрированные данные помещаются в БД ССД в виде справочника-классификатора, подобного приведенному на рис. 2.

2. Данные из различных источников всевозможных информационных систем подразделения компании через соответствующие интерфейсы (“БД ↔ БД с ССД” и/или “источник инф. ↔ БД с ССД” – рис. 1) и с помощью соответствующего программного обеспечения заносятся в БД ССД. При этом эта работа также выполняется

экспертами ПрО заранее, до ее использования другими приложениями и пользователями, в отложенном режиме с занесением соответствующих данных в таблицу (таблицы) согласования. При этом формат документа обмена будет аналогичен формату, приведенному на рис. 3.

Когда и какой способ использовать зависит, во-первых, от возможностей, знаний, желания разработчиков и пользователей тех или иных подсистем, а, во-вторых, от того, как в дальнейшем эти данные могут или будут использоваться в данном подразделении компании, то есть будет ли необходима их дальнейшая интеграция для проведения анализа и получения отчетных документов. Отчетные документы можно достаточно просто получить с помощью специальных приложений как из согласующей БД, в роли которой выступает БД с УМД, так и из витрины данных, рассматриваемой ПрО подразделения, куда эти данные, уже очищенные и согласованные, оперативно поступают из БД с УМД.

При этом обратно из БД с УМД согласованные, обобщенные данные могут передаваться в обратном направлении в соответствующие источники данных подразделения компании через интерфейсы (“БД↔БД с УМД” и/или “источник инф.↔БД с УМД” – рис. 1). Как один из возможных вариантов можно рассматривать стандартные процедуры экспорта/импорта, имеющиеся в распоряжении соответствующих администраторов (АБД, системных администраторов) информационных систем (подсистем) подразделения компании.

В случае второго сценария, когда данные уже находятся в БД с УМД, вариант согласования один – в БД ССД заносятся данные справочника-классификатора подобно приведенному на рис. 2. При этом основным согласующим элементом документа обмена (рис. 3) является строка метаописания ЯМД. Используя таким способом строки метаописаний ЯМД, ССД обеспечивает стандартизацию, консолидацию и согласование определенных текущих и исторических данных. При этом метаданные, подлежащие согласованию, централизованы (в том смысле, что хранятся в единой БД ССД), а данные могут быть распределены (по источникам ЕПДК).

Таким образом, в случае получения нужной информации из ЕПДК пользователь через соответствующий интерфейс обращается к БД ССД, а специальные сервисные программы (программы-агенты [14]), используя данные таблиц согласования БД ССД, помогают ему установить связь с требуемым источником.

Предлагаемая система согласования данных, основа которой строится на платформе *Lotus Notes/Domino*, позволяет также значительно упростить решение задач по обеспечению целостности распределенных данных и гарантированной их доставки путем репликаций потребителю независимо от качества коммуникационных каналов. А так как использование репликаций сегодня является альтернативным и более простым подходом к созданию любых распределенных систем, в виду того что [15]: а) реплицированные данные могут быть сконфигурированы как альтернативный источник данных; б) если основная БД выходит из строя, то может использоваться БД, в которую реплицировались данные; г) можно снизить сетевой трафик, разделяя выполнение приложений по региональным БД; д) в нужный момент времени можно провести соответствующую синхронизацию источников и так далее, то такая способность ССД как нельзя лучше удовлетворяет насущные проблемы пользователей различных распределенных систем, к которым можно отнести, в том числе, и рассматриваемое ЕПДК.

При этом регламент реплицирования в рассматриваемом методе определяется динамикой изменений в справочниках.

В случае использования во многих различных подразделениях компании баз данных с УМД, входящих в ЕПДК, следует отметить еще более простую организацию механизма репликаций, а именно:

- для объектов, отличных от обычных данных (индексов, хранимых процедур, триггеров), репликация не требуется (они уже имеются в схеме БД с УМД);
- вместо моментальных снимков таблиц (*snapshot*) асинхронно распространяются строки метаописаний ЯМД из специальной таблицы – журнала измененных данных, что позволяет снизить нагрузку на систему и ее трафик. При этом в таблицу – журнал измененных данных целевой БД с УМД записываются все реплицируемые данные таблицы – журнала измененных данных исходной БД с УМД. Пример заполнения основных полей таблицы – журнала БД с УМД, содержимое которой реплицируется, приведен в табл. 1;
- автоматически решается проблема целостности транзакций при репликациях (внесение изменений в реплицируемую БД с УМД осуществляется в ней же самой путем исполнения переданных строк метаописаний ЯМД).

Таблица 1

Пример реплицируемой таблицы-журнала

Системное имя	IP-адрес компьютера	Имя пользователя	Имя БД	Строка метаописания ЯМД	Данные или метаданные	Операция*	Время записи в текущую БД
VIT\vitaly	195.110.100.1	HEMA_1	Ora_umd	{<Раздел>=Нов_1;}	метаданные	insert	16.09.2006 10:00:01
Nik\mycmp	101.63.102.28	HEMA_2	spumd	{<Раздел>=Хар_УГЭС;/<КлассО>=Тех. средство;<ТипО>=Вольво;<ЭкзО>=111;<УдалитьО>=;}	данные	delete	26.07.2011 17:41:48

* – операция чтения (*select*) не отслеживается.

Эффективность таких действий во многом будет зависеть от правильного понимания и описания на ЯМД той предметной области, которая моделируется в схеме БД с УМД или переносится из других источников информации в ее схему (или наоборот), а также программного обеспечения, которое организует обмен данными между источниками информации через специально разработанные интерфейсы.

Итак, предлагаемый метод управления информацией в ЕПДК позволяет сделать общедоступными данные, поддерживаемые каждым из существующих подразделений, обеспечив при этом их хранение именно в тех

местах, где они чаще всего используются. Подобный подход расширяет возможности совместного использования информации, одновременно повышая эффективность доступа к ней.

Выводы. Предлагаемое построение и использование единого пространства данных компании позволяет пользователю “размещаться” над существующими системами баз данных, файловыми системами и прочими источниками информации и рассматривать его как единый источник данных. А предлагаемый метод управления информацией в пространстве данных, основанный на использовании универсальной модели данных и платформы для автоматизации совместной деятельности рабочих групп, обеспечивает в отличие от многих имеющихся подходов:

- семантическую интероперабельность пространства данных компании (благодаря использованию общего понятийного пространства ПрО в виде метаонтологий модели “объект-событие”, нормативно-справочной информации и системы классификации и кодирования);
- техническую интероперабельность пространства данных компании (благодаря использованию разработанных стандартных интерфейсов и механизмов доступа к информационным ресурсам, обеспечивающих свободный обмен между разнородными источниками информации ЕПДК, переносимость, масштабируемость и т. д.);
- синтаксическую интероперабельность пространства данных компании (путем использования данных, приведенных к формату представления в документальной системе платформы для автоматизации совместной деятельности рабочих групп и в универсальной модели);
- балансировку нагрузки на телекоммуникационные и вычислительные ресурсы распределенной системы при обработке запросов, разделяя выполнение приложений по автономным базам данных (после соответствующих репликаций в них данных), а также за счет наиболее простой организации механизма репликации в случае использования баз данных с УМД;
- поддержку глобальной идентификации ресурсов (путем использования в справочниках, документах обмена, таблицах согласования данных об этих источниках-ресурсах).

Литература

1. Зиндер Е. З. Проектирование баз данных: новые требования, новые подходы [Электронный ресурс] / Зиндер Е. З. – Режим доступа : <http://www.citforum.ru/database/kbd96/41.shtml>
2. Когаловский М. Р. Перспективные технологии информационных систем / Когаловский М. Р. – М. : ДМК Пресс ; М. : Компания АйТи, 2003. – 288 с. – (Серия : ИТ-Экономика).
3. Когаловский М. Р. Методы интеграции данных в информационных системах [Электронный ресурс] / Когаловский М. Р. – Режим доступа : <http://www.cemi.rssi.ru/mei/articles/kogalov10-05.pdf>
4. Сорока Л. С. Реинжиниринг существующих баз данных в базу с универсальной моделью данных и возможностью распараллеливания процессов / Л. С. Сорока, В. И. Есин, М. В. Есина : сборник научных трудов “Параллельная компьютерная алгебра” : по материалам всероссийской научной конференции с элементами научной школы для молодежи, г. Ставрополь, 11–15 октября 2010 г. / Ставропольский государственный университет. – Ставрополь : Фабула, 2010. – С. 298–304.
5. Калиниченко Л. А. СИНТЕЗ: язык определения, проектирования и программирования интероперабельных сред неоднородных информационных ресурсов / Калиниченко Л. А. – М. : ИПИ РАН, 1993. – 121 с.
6. Франклин М. От баз данных к пространствам данных: новая абстракция управления информацией [Электронный ресурс] / Франклин Майкл, Хэлеви Элон, Майер Дэвид ; [пер. с англ. С. Кузнецова]. – Режим доступа : http://citforum.ru/database/articles/from_db_to_ds/
7. Интеграция метаданных Единого научного информационного пространства РАН [Электронный ресурс] / [А. А. Бездушный, А. Н. Бездушный, В. А. Серебряков, В. И. Филиппов]. – Режим доступа : <http://www.icsti.ru/portal/rus/newproblem/index.php?m=14>
8. Есин В. И. Семантическая модель данных “объект-событие” / В. И. Есин // Вісник Харківського національного університету. – 2010. – № 925. – С. 65–73. – (Серія : Математичне моделювання. Інформаційні технології. Автоматизовані системи управління ; вип. 14).
9. Есин В. И. Язык для универсальной модели данных / В. И. Есин, М. В. Есина // Системи обробки інформації. – 2011. – № 5(95). – С. 193–197.
10. Старых В. А. Спецификация и форматы обмена данными в разнородных информационных системах на базе XML-технологий [Электронный ресурс] / Старых В. А., Дунаев С. Б., Коровкин С. Д. – Режим доступа : <http://citforum.ru/internet/xml/xmltech/>
11. Совместное использование учетных систем и технологии OLAP [Электронный ресурс]. – Режим доступа : http://citforum.ru/database/articles/olap_oltp.shtml
12. Когаловский М. Р. Энциклопедия технологий баз данных / Когаловский М. Р. – М. : Финансы и статистика, 2002. – 800 с.
13. Есин В. И. Формирование аналитической отчетности ДК “Укртрансгаз” / В. И. Есин, Ю. А. Пергаменцев / Проблеми нафтогазової промисловості : зб. наук. праць. – К., 2006. – Вип. 4. – С. 194–201.
14. Линд Д. Lotus Notes и Domino 5/6. Энциклопедия программиста / Дебби Линд, Стив Керн. : пер. с англ. – 2-е изд. перераб. и доп. – К. : ДС, 2003. – 1024 с.
15. Стэнглэнд Д. Расширенная репликация данных [Электронный ресурс] / Стэнглэнд Давид. – Режим доступа : <http://baks.gaz.ru/oradoc/ora/ora009.htm>