

Репозитарии документов как подход к интеграции информационного наполнения автоматизированных систем

Л. А. Рейнгольд, О. А. Славин

В статье рассмотрена система понятий, позволяющая отражать новые явления в социально-экономической сфере, позволяющая на основе расширительного толкования понятия репозитария предложить подходы к интеграции информации об однородных объектах взаимодействующих автоматизированных систем. Обсуждается применение понятий «репозитарий» и «социально-экономические технологии» в автоматизированных информационных системах.

В настоящее время происходит повсеместное внедрение информационных технологий в сферу управления и повседневную жизнь. Все объекты при этом получают информационное отражение, модель в информационных системах различного назначения. Все существенные стороны каждого значимого в жизни общества объекта получают отражение в автоматизированных информационных системах (АИС) соответствующих социальных институтов. Формирующаяся информационная модель, включаем в той или иной форме знания всех поставщиков информации о ней. Это описание постоянно совершенствуется с использованием новых технологических решения, осуществляется в режиме мониторинга — то есть осуществляется не однократно, а систематически. Эта модель должна быть применима для решения типовых задач с использованием типовых инструментов моделирования.

Эти модели первоначально формируются независимо, однако фактически сразу после появления значимых взаимосвязей между информацией АИС, работающих с одними и теми же объектами, а также технической возможности интеграции данных, возникает задача информационной интеграции систем, которые разрабатывались поначалу независимо. То есть осуществляется совмещение различных информационных моделей одних и тех же объектов. Это один из наиболее сложных процессов в процессе информатизации. Действительно — сложности интеграции не ограничиваются сиюминутным поиском взаимосвязей в информации об объектах — это лишь одна из частных задач. Возникают проблемы семантических несоответствий в данных из различных источников,

необходимости учета неодинаковой компетентности поставщиков информации об интересующем объекте, проблемы разграничения доступа к информации и предотвращения негативных социально-экономических последствий от ее неправомерного и злонамеренного использования.

Множество источников и форм представления данных об одном и том же представляющем социально-экономический интерес объекте приводит к множеству несовместимых информационных моделей этого объекта. Такие модели могут существовать как в виде совокупности традиционных документов, так и в виде взаимодействующих и территориально распределенных АИС [1]. Даже простая идентификация объекта является непростой задачей, не говоря о подробном описании его свойств.

Кроме того, совместимая информационная модель значимого объекта нужна не сиюминутно, а должна периодически обновляться, совершенствоваться, становиться более соответствующей запросам потребителя информации.

Можно говорить о различных фазах интеграции информационных моделей объектов в рассматриваемом контексте. Обычно интеграция начинается с простейшего нерегулярного обмена данными, позволяющего облегчить ввод данных во взаимосвязанные АИС. В идеале интеграция должна приводить в процессе развития системы к формированию прозрачной, унифицированной модели каждого интересующего объекта, вся информация о котором явится доступной любому ее потребителю в соответствии с его полномочиями и практическими интересами.

Например, информация о человеке в различных контекстах поначалу накапливается в разрозненных системах, однако постепенно становится все более необходимым ее согласованное получение и использование, обусловленное как чисто техническими причинами — устранения избыточности и повышения достоверности данных, так и организационными причинами, поскольку трудно обеспечить контроль за правомерным использованием, санкционированным получением информации, не имея полных сведений об «информационном образе» человека в АИС и методах его применения в различных ситуациях.

Можно сделать вывод, что локальная информационная база разрозненных информационных систем постепенно преобразуется в интегрированную среду хранения, обработки, санкционированного нормативными документами распространения данных, использующую тот уровень технологий, который доступен в сложившейся инфраструктуре. Технологии интеграции информации находятся в стадии становления, в этой сфере происходят технологические изменения, существующие организационные структуры подстраиваются под новые реальности. Этот

процесс нуждается в методологическом осмыслении и оптимизации, что позволит снизить действующие негативные факторы.

Назовем некоторую интегрированную среду хранения информации об объектах, в которой информация накапливается непротиворечиво и использованием сведений поставщиков данных с наиболее высокой возможной компетентностью *репозитарием* данных.

То есть *репозитарий* (repository) — хранилище информации об объектах, в котором данные представлены в семантически согласованном, сопоставимом виде. Подспудно такое хранилище информации предполагает наличие механизмов обеспечения взаимной информационной согласованности данных всех информационных систем, которые работают с некоторым множеством однотипных объектов. Эта сопоставимость может достигаться различными технологическими способами. Это или четко регламентированное взаимодействие между независимыми взаимодействующими АИС, или единая независимая система, интегрирующая частные модели соответствующих объектов, формируемые взаимодействующими с ней системами. Важной особенностью репозитария является не его технологическая реализация, а то, что информационная модель объекта видна потребителю как согласованное единое целое, несмотря на возникновение этой информации в различных АИС и в территориально распределенном режиме.

В условиях современной ИТ-инфраструктуры в идеале любая информация об объекте должна находить свое место в той или иной информационной системе и размещаться в ней в связи с уже имеющейся об этом объекте информацией.

Репозитарий при таком подходе — универсальная, независимая от конкретной технологической реализации абстракция, которая позволяет отделить функциональность по обработке информационного наполнения АИС от ее конкретной технологической реализации.

Понятие репозитария применимо как связующий концептуальный элемент между несовместимыми технологическими решениями, выделяющий аспект содержания, унификации, а также независимости информации от технологии ее хранения, обработки и распространения.

Репозитарий — средство «свертывания» объема и уровня сложности информационного наполнения АИС за счет многократного использования наиболее полных и эффективных моделей объектов и устранения необходимости в многократном хранении данных об одних и тех же объектах и отсутствия необходимости в применении средств для приведения данных к сопоставимому виду.

К задаче формирования репозитариев (в широком смысле) сводятся многие задачи, решаемые АИС социально-экономического профиля.

Новая технология обработки информации в этом контексте возникает не сама по себе, а как новая возможная функциональность репозитария, способ сокращения различных видов затрат на его ведение и практическое использование.

Репозитарии в той или иной форме занимают существенное место в разрабатываемых подходах к построению АИС, в государственных программах России и зарубежных стран. Хотя называться АИС, реализующие функциональность репозитария могут по-разному, во всех реализациях присутствуют появление в том или ином виде первоисточника для каждого элемента информации и ее отчуждение для совместного использования потребителями в пределах установленных нормативными документами полномочий.

То есть в основе любой технологии, реализуемой АИС, лежит тот или иной способ объединения и предоставления потребителям информации об объектах, то есть репозитарий в рассматриваемом контексте, реализованный с той или иной степенью полноты. На основе информационных потоков в обществе формируются *социально-экономические технологии* (СЭТ), реализующие некоторую потребность в использовании информации, размещенной в репозитарии. Понятие СЭТ подробно рассмотрено в работе [2]

Репозитарий, как универсум данных, достижимое на данном технологическом уровне представление информации о значимых объектах — основа для формирования СЭТ. Новый уровень технологии ведения информации с использованием концепции репозитариев может обеспечить — новый уровень СЭТ применительно к некоторой предметной области.

Подход с использованием понятия репозитария целесообразен в процессе анализа и проектирования АИС на различных уровнях внедрения информационных систем, в том числе при решении задач автоматизации на региональном, муниципальном уровне управления, а также в процессе формирования информационного наполнения территориально распределенных, концептуально и технологически отличающихся систем управления ведомств на уровне государства.

Использование понятия репозитария в рассматриваемом контексте позволяет на новом уровне решать задачи управления социально-экономическими системами. Так как результатом использования этого подхода является семантически согласованные в момент времени и в процессе изменения данные, они могут применяться для решения социально-экономических задач в различных приложениях.

В том числе может использоваться семантическая согласованная информация в задачах моделирования социально-экономических процессов. При рассмотрении информационного отражения объекта во вза-

имодействующих АИС как единого целого, возрастает информационная обеспеченность задач моделирования. Облегчается верификация информации в процессе ее получения и использования.

Обобщая рассмотренные в статье положения, можно сделать следующие основные выводы:

Каждый объект в социально-экономической системе описывается с различных сторон. Применительно к системе социально-экономического управления — каждая организация на своем уровне и в своих интересах формирует собственную модель объекта.

В то же время для каждого объекта в каждом контексте существует и должен быть доступен для заинтересованных пользователей «первоисточник» данных, информация об объекте которого является для них первичной. То есть, его информация должна оперативно учитываться в процессе формирования «информационного образа» объекта в автоматизированных системах, которые с ним работают.

Для отражения перечисленных выше тенденций предлагается использовать термин *репозитарий* в несколько более широком смысле, чем это обычно принято. То есть под репозитарием предлагается понимать виртуально или реально интегрированную среду обработки данных, обеспечивающую оперативный доступ всем участникам формирования информации об объекте к первоисточнику по каждому элементу этой информации.

Такой подход позволит повысить качество информации, формирующейся в системе, уменьшить издержки на обработку информации, а также сформировать новые подходы к разработке, нормативному и организационному обеспечению автоматизированных систем.

Литература

1. Рейнгольд Л. А. Структурирование информации: системный подход. М.: Наука, 2004. 200 с.
2. Рейнгольд Л. А., Славин О. А. Социально-экономические технологии как обобщение тенденций социально-экономического развития // Труды ИСА РАН. Интеллектуальные информационные технологии. Прикладные аспекты. М.: 2005. С. 40–55.
3. Информационная технология. Комплекс стандартов на автоматизированные системы. Техническое задание на создание автоматизированной системы: ГОСТ 34.602–89. М.: Издательство стандартов, 1990. 13 с.