

Моделирование данных и стандартизация работы с данными в государственном секторе

Ю. П. Липунцов

Московский государственный университет им. М.В.Ломоносова
lipuntsov@econ.msu.ru

Аннотация

Успешность и эффективность проектов государственного управления во многом предопределяется информационной поддержкой деятельности. Информационное обеспечение административных процессов с учетом территориального и функционального разнообразия участников предполагает интеграцию информационных ресурсов. Значительные массивы информационных активов, накопленные в ведомствах, могут быть эффективно задействованы при выстраивании целенаправленной работы по интеграции информационных ресурсов. Качество модели интеграции предопределяется ее гибкостью, способностью к адаптации при миграции систем. Такие характеристики интеграции достигаются путем построения и поддержанием в актуальном состоянии модели данных. Для организации работы в этом направлении необходимо определить состав ролей по управлению данными и функции, закрепленными за этими ролями.

Система моделей

Составление моделей данных необходимо для эффективного интегрирования систем, обеспечения совместимости между системами на уровне слоя данных, управления доступом к данным, повторного их использования. Игнорирование слоя данных при широком применении сервисных технологий может обернуться серьезной проблемой при доведении системы до полноценного набора электронных услуг. Модели данных являются связующим звеном между системами.

Одной из проблем организации межведомственного взаимодействия является семантическая интероперабельность электронного правительства, поскольку с одним и тем же документом приходится работать многим пользователям: министерства, префектуры и муниципалитеты, организаций государственного сектора и т.д.

Традиционный вариант предоставления услуг предполагает стандартизацию документов в рамках ведомства, при переходе на электронный вариант количество пользователей документа становится гораздо больше. Перевод документов в электронный вид часто предполагает фиксацию тех полей, которые встречались в традиционном варианте. Поэтому при переводе в электронный вариант стандартизация данных сталкивается с проблемой моделирования данных, посредством которой реализуется инвентаризация и описание данных. Одной из особенностей государственных услуг является многообразие используемых форм документов. В бизнесе полный спектр операций охватывается с помощью нескольких типичных форм. Поэтому госсектору необходимо вырабатывать единый ресурс информационного контента государственных документов.

В связи с этим, необходимость создавать методологию моделирования, которая позволила бы стандартизировать и унифицировать документы административных процессов, создать условия для семантической совместимости. Такая методика послужит созданию, распространению и повторному использованию существующих схем данных и программных компонент.

В результате построения моделей данных происходит повышение согласованности и связанности данных в границах предметной области. Совокупность моделей позволяет определить возможности по созданию приложений класса Хранилищ данных или дополнительных узлов для работы с едиными справочниками, обеспечивает улучшенный обмен данными из-за семантической согласованности данных. Более эффективно развитие платформонезависимых моделей данных прикладных систем класса XSD, которые позволяют создать отчуждаемое определение данных ведомств, семантическую согласованность с целью повышения повторного использования данных.

На рисунке 1 представлены основные классы данных электронного государства верхнего уровня для определенной предметной области, где отражаются базовые информационные элементы следующих категорий:

- Информационные объекты, хранящиеся в базовых регистрах;
- В ведомственных системах, где будут отражаться транзакции, фиксируется информация

- о базовых элементах услуг для граждан, и услуг для бизнеса;
- Четвертая категория базовых элементов описывает информационные объекты, используемые в приложениях для поддержания деятельности чиновников;
- Отдельным блоком представлены данные для оценки результативности деятельности. Для представления состава информационных объектов предметной области формируется список информационных объектов.



Рис. 1. Классы информационных объектов модели данных предметной области государственного управления.
Источник [1]

Три категории моделей предметной области

Моделирование данных целесообразно проводить в рамках предметной области. Совокупность моделей предметной области должна давать полное представление обо всех базовых процессах, происходящих в рамках границ системы и подлежащих отражению в информационной системе. На рисунке 2 приведено три уровня моделей: модель предметной области, концептуальная модель данных ведомства (без атрибутов данных), логическая модель ведомства (с атрибутами данных).

Модель предметной области определяет основные классы данных.

Модель данных схематично отображает предметную область, и строится с целью содействия взаимодействию управленческого звена с техническими специалистами, улучшению коммуникаций и накоплению знаний. Помимо схем модель сопровождается документацией, в которой приводится описание модели.

Модель строится с целью содействия взаимодействию управленческого звена с техническими специалистами, улучшению коммуникаций и накоплению знаний. Помимо схем, модель сопровождается документацией, в которой приводится описание модели.

В модели предметной области, пример которой приведен на рис.3., нет привязки к определенному

ведомству. На схеме видно, что к модели помимо медицинских учреждений имеют отношения исследовательские институты, учреждения образования, производители и поставщики материалов. Таким образом, создается круг пользователей информационного пространства предметной области.

Создание полноценной модели процесс итеративный, который предполагает поддержание модели данных в актуальном состоянии, делая все больший акцент на специфику модели, определяемую предметной областью. Такая модель будет необходимым элементом при решении многих задач, связанных с информационной поддержкой управленческих задач, а также технических решений, например, при создании Системы электронного документооборота (СЭДО).

Детализация отдельных блоков, представленных в модели предметной области, реализуется на концептуальном уровне. Для этого уровня характерно большое количество сущностей, взаимосвязи которых также отражаются. Характер отдельных связей может носить контекстный характер, то есть в одном случае эти связи будут использоваться, в другом они не принципиальны. На рис. 4 приводится фрагмент модели описания местоположения, которое используется для указания места дислокации отдельных сущностей предметной области: материалов, учреждений, клиентов и прочего. На концептуальной модели приводится детальное представление, поэтому в ней должны быть представлены все возможные варианты описания мест расположения разных категорий сущностей. Для одних вариантов будет актуален почтовый адрес, для записи клиентов на прием через сайт поликлиники необходим адрес сайта в виртуальном пространстве, особо выделено положение мобильных пользователей.

Физическая реализация информационной системы осуществляется на основе логической модели данных, в которой прописываются сущности, их атрибуты и связи между ними. На рис. 5 приведена схема модели данных адресной системы РФ КЛАДР. Пример адресной системы показывает, насколько техническая реализация адресной системы отличается от традиционного представления потового адреса. В основу модели положено универсальное описание стандартных уровней адресной системы (сущности Level, Socrbase). В этих сущностях описывается семь уровней классификации: первый уровень включает объекты федерального значения, второй уровень — районы объектов второго уровня, третий уровень классификации включает города регионального и районного подчинения, сельсоветы и т. д. Таким образом, на седьмом уровне детализации представлены квартиры, офисы, частные дома.



Рис. 2. Уровни модели данных ведомства
Источник адаптировано: [4]

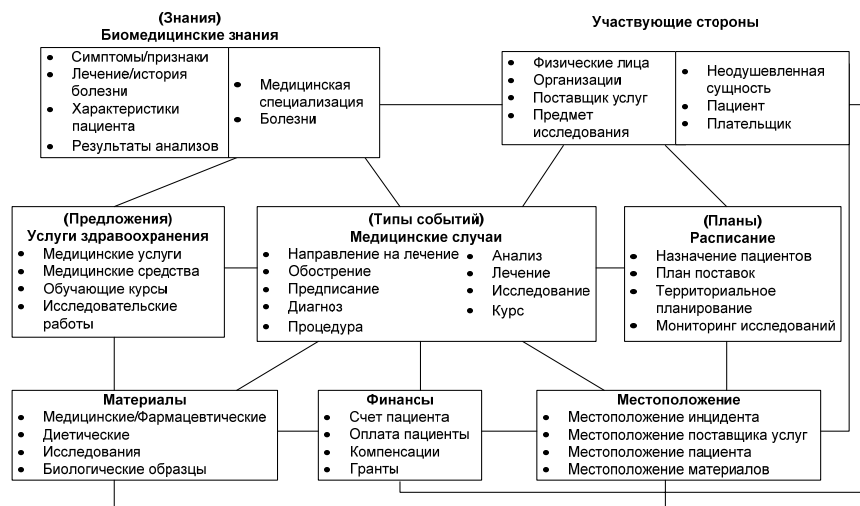


Рис. 3. Пример модели предметной области (здравоохранение, фрагмент)
Источник: адаптировано [4]

Наиболее сложным аспектом разработки модели является достижение семантической однозначности. Одним из этапов в этом направлении является работа с наименованием данных, в том числе работа с синонимами (слов разных по звучанию и написанию, но имеющие одинаковое или близкое значение) и омонимами (слова, разные по значению, но одинаковые по написанию), которые должны быть по возможности ликвидированы или отмечены в метаданных систем.

Второе направление деятельности в области семантики - создание глоссария терминов сервисов. При реализации концепции архитектурного подхода создаются несколько моделей – модели содержательного слоя, модели данных, модели технологий. Все модели имеют дело с одной

предметной областью, и возникает значительное количество артефактов, связанных с осуществлением одних и тех же действий, но на разных уровнях архитектуры или на разных уровнях декомпозиции модели. Поэтому однозначность толкования терминов является необходимым элементом стандартизации для обмена данными и взаимодействия систем.

Модели данных должны быть импортированы в среду управления метаданными, которая активно задействуется в построении системы идентификации и безопасности, обеспечивающих выполнение нормативных требований о неприкосновенности частной жизни и конфиденциальности персональных данных.



Рис. 4. Пример концептуальной модели данных для описания местоположения ресурсов
Источник: адаптировано [4]

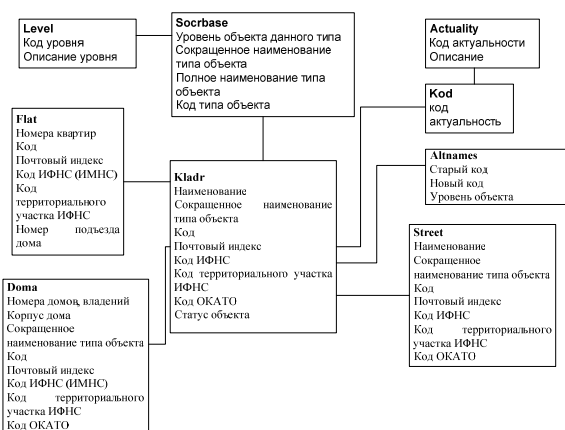


Рис. 5. Пример логической модели данных для описания Почтового адреса
Источник: схематично на основе описания [2]

Сущности модели данных или отдельные атрибуты, которые содержат персональные данные, конфиденциальную информацию ведомства, а так же данные, использование которых регулируется специальными документами, при построении модели данных должны быть классифицированы в соответствии с режимом доступа. Эта классификация отображается в модели данных.

Управление стандартизацией данных

Совокупность данных, которые имеют непосредственное отношение к государственному управлению, распределены между ведомствами, агентствами. Для эффективного использования накопленных данных, упрощения процедур обмена необходимо составлять модели данных, сопровождать модели документацией. Для подготовки документации нужны стандарты моделирования, стандарты сопроводительной документации, а также описание процедур утверждения подготовленных документов. Эти разделы управления данными находятся в ведении деятельности по управлению стандартизацией данных.

В разделе стандартизации определяется состав ролей по управлению данными и их обязанности (рис.6). Для того чтобы выделить основную деятельность по управлению данными и определить участие специалистов соприкасающихся областей, рассмотрим категории ролей начиная с само-

го верхнего уровня, где представлено деление сектора ИКТ на слои архитектуры.

Слой архитектуры	Роли
Функции государственного управления в форме процессов	Эксперт предметной области (Subject Matter Expert)
Информация (данные + контекст)	Комитет по управлению данными Руководитель управлением данными (Chief Information Officer) Архитектор данных Куратор данных Владелец данных
Приложения	Руководитель управлением приложениями (Chief Applications Officer) Администратор базы данных
Оборудование и системное ПО	Руководитель управлением мощностями (Chief Technology Officer)

Рис. 6 Категории ролей по слоям архитектуры

Формализованное представление деятельности ведомства в форме административных процессов находится в ведении эксперта предметной области, который является связующим звеном между специалистами ключевых процессов и технологическим подразделением. В разделе управления данными роль эксперта сводится к анализу существующего положения в сфере информационной поддержки, включая анализ неравномерности в поставке информации, определяет потребность в дополнительных данных.

К основным ролям в области управления данными относятся руководитель управления данными, архитектор данных, владелец данных и куратор данных.

Основная функция руководителя управления данными сводится к адаптации стандартов данных, поддержание их в актуальном состоянии. Поскольку именно посредством стандартов данных устанавливается межведомственное и межуровневое взаимодействие, а также взаимодействия с системами вспомогательных процессов и с системами третьих лиц.

В задачи архитектора данных входит разработка и публикация модели данных. Региональные модели, модели предметной области разрабатываются в координации с архитекторами и кураторами данных ведомств.

Кураторы данных координирует создание и рецензирование стандартов данных. Владелец данных – обеспечивает соответствие стандартам данных для данных поставляемых за пределы ведомства. Руководитель управления приложениями следит за соответствием стандартам данных используемых и разрабатываемых приложений.

При построении организационной структуры управления данными на уровне государства (рис. 7) сохраняются основные роли и функции. В дополнение к основным ролям появляется Комитет по управлению данными, который выполняет аудит межведомственного и межуровневого взаимодействия на уровне данных. Функции по составлению модели данных остаются за архитекторами данных, куратор данных реализует в рамках своих полномочий процесс разработки и

адаптации стандартов данных. Реализация модели данных на уровне приложений осуществляется администраторами баз данных. Помимо этого администратор способствует разработке модели

обмена данными. Архитектор данных ведомства содействует архитектору данных верхнего уровня в реализации политики управления данными.

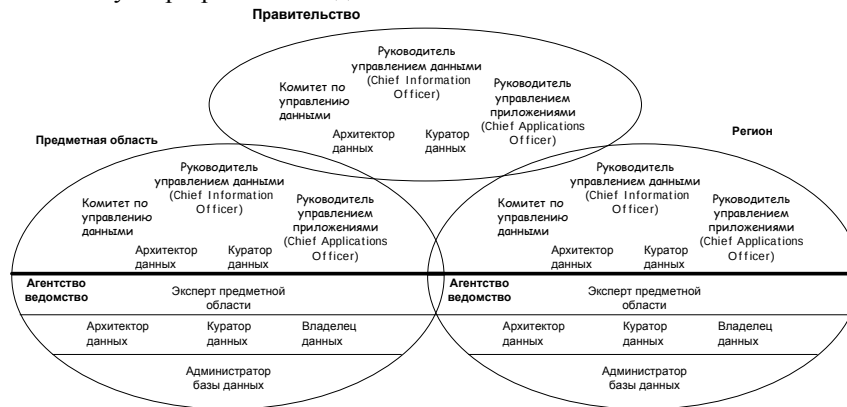


Рис. 7. Схема организационной структуры управления данными электронного государства

Реализация политики стандартизации данных включает этапы обследования, формирование требований, анализ межведомственного и межуровневого обмена данными и приведение стандартов в соответствие с требованиями.

Этап 1 Обследование. На этапе обследования производится инвентаризация источников информации с целью определения информационных потребностей, определяются недостатки информационной поддержки в отдельных областях, и производится планирование деятельности с учетом выявленных недостатков.

В результате инвентаризации формируется совокупность документов, которые включают модели данных, диаграммы, документы, описывающие модели, XSD, схем процессов. В более зрелых организациях документы могут включать схемы для последовательного обмена данными. Для анализа потребностей в данных процессы обмена представляются в форме документов, что позволяет координировать интересы участников информационного обмена.

Отправной точкой для анализа существующей ситуации является совокупность управленческих задач, для решения которых выявлена необходимость совершенствования работы с данными, в том числе по обмену данными. Эти потребности оформляются в виде заявок со стороны владельцев ключевых процессов с указанием недостатков и возможностей для улучшения. На основе этих документов производятся модернизация административных процессов, обмен информацией, доработка и разработка приложений. Особое внимание уделяется потребностям, которые могут быть удовлетворены за счет стандартизации данных.

Разработка стандартов данных начинается с выявления информационных потребностей, которые описаны в формализованном виде.

Этап 2 Формулирование требований. Цель второго этапа стандартизации заключается в построении первоначального варианта модели данных и стандартов данных с целью инициировать обсуждение разработки стандартов.

Архитектор данных предметной области или региона в координации с архитектором данных ведомств разрабатывают и представляют на обсуждение шаблон модели данных и сопроводительную документацию. На этом этапе делаются первые шаги по направлению координации деятельности по выработке стандартов данных. Куратор данных, в ведении которого находится процесс разработки стандартов, отвечает за проведение предварительного согласования с владельцами административных процессов, администраторами баз данных и прочими участниками. Процесс стандартизации данных итеративный процесс, который предполагает постоянное взаимодействие с заинтересованными сторонами.

Взаимодействие с участниками строится на обсуждении логических моделей данных, которые содержит предлагаемый стандарт.

Архитектор данных предприятия помимо разработки модели данных может инициировать процесс перехода на новый стандарт.

Этап 3 Проведение формального анализа межведомственного и межуровневого обмена данными. На этом этапе производится анализ поступивших предложений на предмет необходимости внесения изменений в межведомственный обмен данными.

Анализ производится при активном участии Архитектора данных и Куратора данных, которые отслеживают наличие данных, корректность представления метаданных, соотношение логической модели и административных процессов и прочие моменты.

Этап 4 Изменение стандартов с учетом требований. На основе замечаний и рекомендаций, полученных в ходе анализа межведомственного взаимодействия, формируется пакет изменений стандартов. Прежде чем определенные предложения попадут в пакет изменений, проверяются все возможности по реализации запросов пользователей в рамках существующих стандартов.

Перед окончательным утверждением изменений проводится презентация Консультационному Совету, Руководителю управления

Выводы

Работу в области управления данными целесообразно выстраивать по предметным областям, в которые могут включаться участники разных секторов деятельности государственного управления, но имеющие отношение к единому информационному пространству. О построении универсальной модели данных, охватывающей все разделы деятельности в области государственного управления говорить пока рано.

В рамках предметных областей создается совокупность моделей данных, которые позволяют использовать в качестве инструмента интеграции не только готовые взб-сервисы, но и SQL-сервисы, позволяющие обращаться напрямую к данным. Для реализации такой схемы необходимо стандартизировать работу с данными: определить состав ролей и функции, закрепленные за этими ролями. В статье приведен обзор зарубежного опыта на эту тему.

Литература

- [1] Business Integration Driven by Business Lines: A perspective on the Data Reference Model as it relates to Cross Agency Challenges. Standards Based Architecture to Support Federated Data Management Industry Advisory Council 2003
- [2] Классификатор адресов России (КЛАДР) http://www.gnivc.ru/inf_provision/classifiers_reference/kladr/
- [3] B. Niemann A New Enterprise Information Architecture and Data Management Strategy for the U.S. EPA and the Federal Government http://semanticcommunity.info/Niemann_Family/Brand_Niemann#Online_Book.3a_A_New_Enterprise_Information_Architecture_and_Data_Management_Strategy_for_the_U.S._EPA_and_the_Federal_Government
- [4] Enterprise Data Architecture. Strategic Assessment Architecture and Roadmap [http://semanticcommunity.info/EPA/EPA_Data_Architecture/Enterprise_Data_Architecture_Strategic_Assessment_Architecture_and_Roadmap#Enterprise_Data_Architecture_\(EDA\).2f_Enterprise_Information_Management_\(EIM\)](http://semanticcommunity.info/EPA/EPA_Data_Architecture/Enterprise_Data_Architecture_Strategic_Assessment_Architecture_and_Roadmap#Enterprise_Data_Architecture_(EDA).2f_Enterprise_Information_Management_(EIM))
- [5] XRepository <https://www.xrepository.deutschland-online.de/xrepository/help.xhtml>
- [6] Data Standardization Process Enterprise Information Architecture.. Version 1.1 Commonwealth of Virginia October, 2009 <http://www.vita.virginia.gov/oversight/default.aspx?id=10342>

Data modeling and standardization for e-Government

Y. Lipuntsov

The quality of information support for public administration processes based on integration the information resources. A lot of information assets accumulated in government agencies can be effectively involved in integration process. The quality of integration model is predetermined by flexibility and adaptability of information systems. These characters of systems are achieved through set of data models: general model, model of domain, conceptual model, logical and physical models. The creation of data models is necessary to define the roles of the management of data and functions assigned to these roles.