

# Инфраструктура как среда интеграции единых справочников государственных информационных систем

Ю.П. Липунцов, Е.В. Стариков, В.А. Пшенкин, В.А. Булатов

Экономический факультет МГУ им. М.В.Ломоносова

lipuntsov@econ.msu.ru, eugeny@econ.msu.ru, pshenkin@econ.msu.ru, vlad@econ.msu.ru

## Аннотация

Использование информационных технологий в государственном секторе становится все более активным. Решаются вопросы законодательного характера, создаются условия для взаимодействия информационных систем ведомств между собой. Большинство процессов государственного управления носит межведомственный характер. Предоставление сервисов с использованием программных агентов требует активного взаимодействия информационных систем различных ведомств. Для эффективной организации межведомственного взаимодействия необходимо создание единого центра обмена данными – репозитория, реализующего две функции: сопоставление схем данных отдельных локальных подсистем и организацию хранилища данных для эталонных справочников и их связей со справочниками локальных информационных систем. Существует несколько методов интеграции подобных взаимосвязей. В статье представлено приложение, позволяющее реализовать интеграцию данных в системе реализована на основе модификации реляционного принципа организации данных. Ведутся работы по интеграции с использованием принципов онтологического моделирования.

## 1. Схема обмена данными между системами

С точки зрения архитектуры информационной системы взаимодействие двух систем осуществляется посредством адаптера, обеспечивающего технологическую совместимость систем для обмена данными и функциями. Если возникает необходимость

интегрировать несколько информационных систем (ИС), то вариант с реализацией адаптеров для каждой пары ИС становится очень затратным и не универсальным решением.

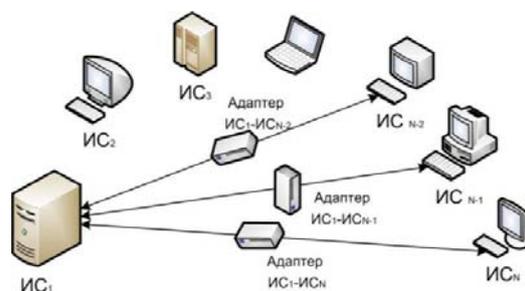


Рисунок 1. Схема интеграции посредством адаптеров

В этом случае возникает необходимость реализовать некую интеграционную шину (Enterprise Service Bus), с которой бы взаимодействовала каждая из систем. Такой вариант интеграции предполагает возможность обмена данными и функциями каждой ИС с любой другой ИС, помимо этого такая архитектура интеграции позволяют подключать новые информационные системы.

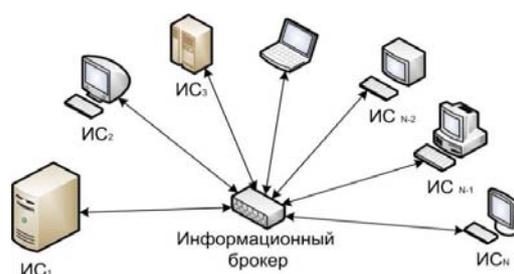


Рисунок 2. Схема интеграции посредством информационного брокера

Центральным звеном в такой архитектуре выступает информационный брокер, который позволяет организовать интеграцию разнородных приложений и данных путем предоставления следующих сервисов:

- пересылки сообщений и перемещение данных;
  - маршрутизации сообщений;
  - трансформирования.
- Выполнение заданной последовательности пересылок сообщений, их маршрутизация и

трансформирование могут быть доведены до автоматического выполнения – так формируются сервисы по управлению бизнес-процессами.

Отслеживание состояния автоматизированных бизнес-процессов позволяет реализовать на информационном брокере функции анализа операций.

Для реализации этих функций информационный брокер предполагает работу с адаптерами различных приложений и форматов данных:

- адаптерами к веб-сервисам;
- адаптерами к транзакционным системам;
- адаптерами к реляционным БД;
- API-адаптерами для популярных приложений.

## 2. Инфраструктура как элемент интеграции данных

Инфраструктуру электронного государства можно определить как создание единых для всех информационных систем элементов данных и технологий и обеспечение доступа граждан и бизнеса к услугам с помощью универсального интерфейса, независимо от того, предоставляются эта услуга одним или несколькими ведомствами. При оказании услуг предполагается межведомственное взаимодействие на вертикальном и горизонтальном уровне.

При предоставлении услуг в форме электронных сервисов отдельными федеральными, региональными или местными субъектами возникают следующие трудности:

- проблема финансирования инновационных проектов в области государственного управления,
- отсутствие необходимой квалификации персонала в области поддержки инноваций,
- отсутствие надлежащей технической инфраструктуры.

Эти трудности сказываются и на уровне безопасности применяемых приложений, а безопасность в случае приложений для госсектора является одним из важных показателей.

Поэтому необходимо создавать мощную инфраструктуру, которая бы переносила центр тяжести с ведомственных информационных систем на элементы инфраструктуры. Предоставление услуг в такой форме аналогично передаче значительной части процесса предоставления услуг на аутсорсинг, когда отдельные этапы услуг выполняются внешними поставщиками. Это позволяет повысить внутреннюю эффективность деятельности госорганов. Помимо этого возрастают возможности мониторинга и контроля услуг, появляется возможность централизованной модификации, и технического обслуживания.

Инфраструктура выполняет операции в области хранения, преобразования данных и их обработки. Ведомственные приложения,

работающие в рамках инфраструктуры, выполняют специфические для ведомства функции, используя такие элементы инфраструктуры как единые справочники, стандартные веб-сервисы, электронно-цифровые подписи, платежные системы и т.д.

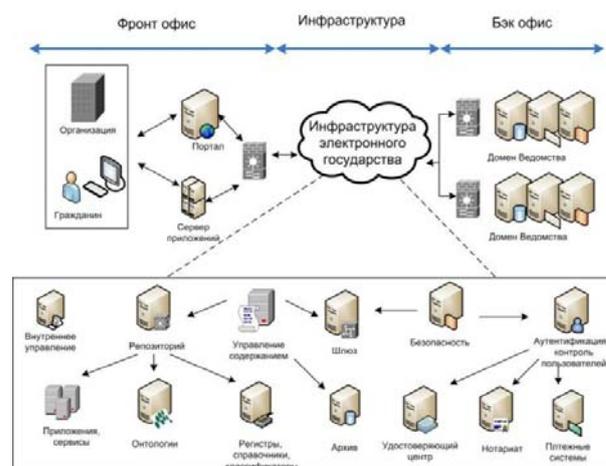


Рисунок 3. Элементы инфраструктуры электронных сервисов

Правительство РФ определило базовые государственные информационные ресурсы [1] которые будут использоваться при предоставлении муниципальных или государственных услуг. В тексте документа приведен список регистров, которые ведутся на федеральном уровне и предоставляющие доступ к регистрам.

К таким источникам сведений относятся, в частности:

- Единый государственный реестр налогоплательщиков;
- Государственный кадастр недвижимости;
- Единый государственный реестр юридических лиц;
- Единый государственный реестр индивидуальных предпринимателей;
- Единый государственный реестр прав на недвижимое имущество и сделок с ним;
- Реестр регистрации автотранспортных средств;
- Адресный реестр.

Согласно постановлению, информационные ресурсы должны содержать идентификаторы, которые позволят получить из регистра сведения о лице или объекте. Идентификаторами могут быть ИНН, номер государственной регистрации, номер страхового свидетельства, кадастровый номер объекта недвижимости и др.

Помимо источников указаны ведомства, уполномоченные поддерживать регистр в актуальном состоянии и предоставлять доступ к данным регистра. Среди уполномоченных органов значатся Пенсионный фонд РФ, Федеральная налоговая служба, Росреестр, Федеральная

миграционная служба, Министерство внутренних дел.

При разрозненной организации ведения регистров у пользователей возникает необходимость построения системы доступа к отдельному регистру. Настройка связей с регистром остается на ответственности локальной системы. При такой реализации межведомственного взаимодействия мы получаем схему, приведенную на рис. 1.

Для реализации интеграции по принципу интегрирующей шины необходима организация ведения справочников в едином центре, являющемся инфраструктурным звеном электронного государства. Вопрос физического места нахождения этих регистров, а также работа по поддержанию их в актуальном состоянии может остаться на ответственности указанных в постановлении ведомств. С изменением технологий будет происходить миграция схемы данных, протоколов доступа к данным и т.д. При существующей организации работы регистров за всеми эти изменения должны следить администраторы локальных систем. Это будет составлять существенную часть локальных ИТ-бюджетов, а также пострадает качество предоставляемых сервисов.

Вопрос о создании единых справочников для государственных нужд рано или поздно возникнет, поскольку все регистры названные в постановлении носят ведомственный характер, то есть предназначены для решения ведомственных задач. Например, Регистр избирателей – как наиболее полная на сегодняшний день агрегированная на федеральном уровне база данных по персональным лицам не может выступать полноценным регистром населения, поскольку изначально эта подсистема была ориентирована на формирование списка избирателей и в регистр попадали лица после получения паспорта. Помимо этого, могут существовать технические ограничения на работу системы, специфичные для ведомственной задачи, например в ПРИУР не может быть фиксирован гражданин без адреса места жительства, поскольку он не может быть привязан в избирательному участку. Похожие проблемы могут возникать и в случае других регистров.

### **3. Справочники государственных информационных систем.**

В системе государственных информационных систем используются два типа справочников: линейные и иерархические. Два типа классификатора различаются друг от друга принципами распределения полномочий по ведению классификатора между операторами разных уровней. Рассмотрим это на примере классификаторов подсистемы Регистр

избирателей, участников референдума (ПРИУР) «ГАС Выборы».

Классификаторы фамилий, имен, отчеств и т.д. является примером реализации линейного справочника. Полномочиями на ведение линейных классификаторов обладает комплекс системы автоматизации Центральной избирательной комиссии (КСА ЦИК) – самый верхний уровень информационной системы. КСА нижних уровней могут только добавлять так называемые «неподтвержденные» значения, которые может впоследствии утвердить КСА ЦИК. Неподтвержденные значения имеют код КСА принадлежности того КСА, который их породил, и не распространяются на все КСА ГАС «Выборы» до их подтверждения КСА ЦИК.

Классификатор территорий ГАС «Выборы» КЛАСТЕР представляет собой иерархический классификатор. Полномочия на ведение значений КЛАСТЕР распределяются между КСА по ветвям классификатора. Основное правило ведения этого справочника следующее: КСА наделен полномочиями на работу со значениями ветви классификатора, если узел, образующий эту ветвь, связан с учреждением, обслуживаемым данным КСА и другие КСА, лежащие ниже по иерархии ГАС «Выборы», не наделены полномочиями на работу с данной ветвью КЛАСТЕР.

Необходимость создания единого универсального классификатора подтверждает ситуация с адресным классификатором ГАС «Выборы», при реализации которого возникла проблема. С 2003 года министерство налогов и сборов организовало и поддерживало в актуальном состоянии систему адресации КЛАДР. Поэтому при формировании законодательных актов на создание регистра избирателей эта система адресации была включена как составная часть регистра. Вместе с тем эта адресная система не была расширяемой и не смогла выполнить необходимые для регистра функции, в частности для ПРИУР необходимо распределение адресов по учреждениям, которые были задействованы в формировании базы данных регистра, таких как избирательные комиссии, призывные пункты военкоматов, суды, медицинские учреждения и т.д. Поэтому в регистре сформирован собственный блок ответственный за формирование адреса и привязку адреса к совокупности учреждений.

Каждый регистр федерального уровня является достаточно сложной системой. Если посмотреть на концептуальную схему данных ПРИУР (рис. 3), то интеграция ведомственной информационной системы с такой подсистемой на предмет получения данных является достаточно сложной задачей.

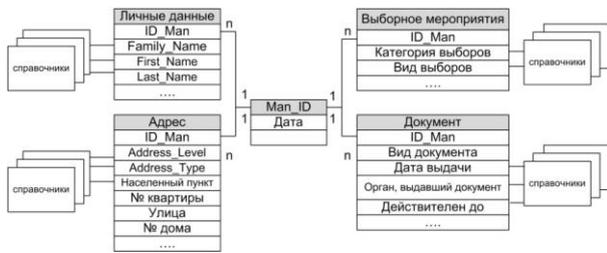


Рисунок 4. Концептуальная схема данных ПРИУР

Это не просто получение двух полей: идентификационного номера и текста, как в случае элементарного справочника. Для применения на практике отдельного регистра необходимо установление однозначной связи между источником и системой его использования. Применение таких идентификаторов как номер страхового свидетельства другими ведомствами, например миграционной службой, будут вызывать проблемы, поскольку на все изменения в первоисточнике нужны репликации на сопредельные системы. В предложенной постановлении правительства версии организации регистров заинтересованными лицами в репликации являются локальные системы, а лицо, ответвленное за ее реализацию, отсутствует. Поэтому целесообразно выстраивать среду для поддержания единых справочников.

При разработке архитектуры системы интеграции используются два подхода, которые принято называть Global as View и Local as View.

Первый из этих подходов (Global as View) используется, когда известны все источники данных. Доступ к данным локальных систем в этом случае может быть реализован либо через отображение запроса пользователя во множество подзапросов к локальным источникам данных либо как запрос к интегрированным данным, если система интеграции предназначена для поддержки полного физического представления интегрируемых данных в хранилище данных.

При втором архитектурном подходе (Local as View) допускает динамичность состава множества источников данных, когда новый источник данных может подключаться к системе.

Рассмотрим основные элементы среды интеграции. К числу основных средств, используемых для обеспечения интеграции информационных ресурсов, относятся инструменты интеграции базы данных метаданных и инструменты интеграции непосредственно данных в форме хранилища данных (рис. 5). К таким инструментам относятся адаптеры (Wrappers), посредники (Mediators) и средства интеграции схем или интеграции онтологических спецификаций. Результаты интеграции представляются пользователю в виде доступа метаданным и доступа к данным.

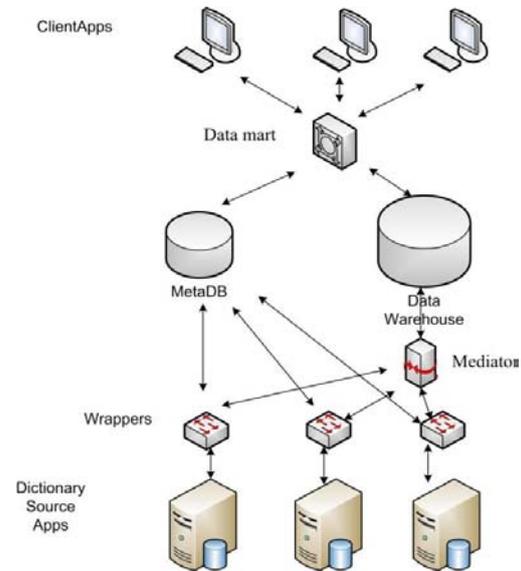


Рисунок 5. Две части системы интеграции: Метаданные и Хранилище данных

Процесс интеграции можно представить в виде следующей последовательности действий [3].

1. Подготовить схемы данных локальных систем
2. Сгруппировать схем в кластеры на основе сообществов по интересам (определить какие источники будут использованы для формирования единого справочника).
3. Для каждого кластера схем создается схема интеграции
  - 3.1 Интеграция схем
    - 3.1.1 Парные сравнения входных схем
      - анализ конфликтов имен
      - анализ конфликтов структур
    - 3.1.2 Подготовка поправок схемы
    - 3.1.3 Подготовка интегрированной схемы
    - 3.1.4 Описание атрибутов схемы интеграции.
  - 3.2. Реализация схемы на физическом уровне

Существенную часть в интеграции данных занимает интеграция метаданных. Одной из задач интеграции метаданных является задача интеграции схем данных.

При интеграции данных ключевой вопрос процесса (мэпинга – mapping) сопоставления схем данных исходной и целевой системы – для каждого источника данных необходимо определить совокупность правил по которым будут преобразовываться данные при переходе от одной системы к другой.

Трудности ее решения в конкретных ситуациях могут быть связаны с наличием конфликтов [7], например:

- конфликтов неоднородности (используются различные модели данных для различных источников);
- конфликтов именования (в различных схемах используется различная терминология, что приводит к омонимии и синонимии в именовании);

- семантических конфликтов (выбраны различные уровни абстракции для моделирования подобных сущностей реального мира);
- структурных конфликтов (одни и те же сущности представляются в разных источниках разными структурами данных).

Другая типичная задача - интеграция онтологических спецификаций информационных ресурсов.

В материалах по открытым источникам данных в области государственного управления приводятся мнения, о том, что интеграция открытых данных может быть выполнена на уровне метаданных [2,5]. Для этого предлагается описать источник данных в виде онтологической модели. При этом роль репозитория сводится к работе на уровне метаданных.

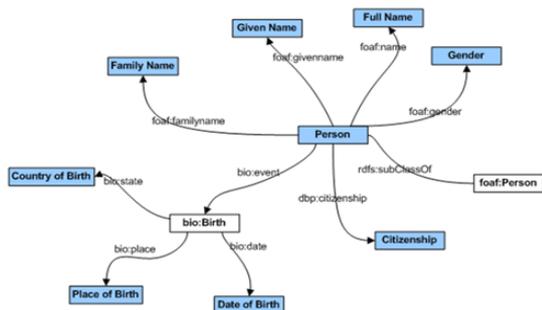


Рисунок 6. Модель идентификации физических лиц федерального репозитория ЕС. Источник [5]

#### 4. Среда интеграции

В ходе проекта по интеграции ряда информационных систем Экономического факультет МГУ им. М.В.Ломоносова была создана среда, в которой нашли практическую реализацию принципы интеграции систем. Схема основных компонент среды интеграции приведен на рис. 7.

Основная задача проекта по созданию среды интеграции была формирование эталонных справочников, которые в последующем могли бы стать основой для интеграции существующих и присоединения новых приложений организации. Основу приложения составляет интегратор данных и сервер приложений. При реализации среды приложений список приложений был заранее известен, при этом не один из источников не мог служить полноценным поставщиком отдельного справочника. Таким образом, каждый справочник формировался из нескольких источников. Для реализации этой функции в среде интеграции был разработан инструмент, который позволяет определить тип справочника и источник - локальное приложение. Для каждого справочника и источника были реализованы процедуры сопоставления локальных справочников и эталонным, а в случае необходимости добавления записей в эталонный

справочник. При добавлении записей в эталонный справочник отсылается уведомление администраторам локальных систем о произведенных изменениях.

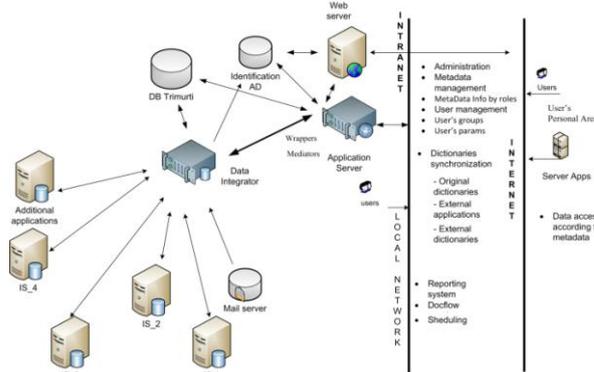


Рисунок 7. Схема среды интеграции

Соединение эталонных справочников со справочниками локальных систем осуществляется через таблицу X-Link (рисунок 8).

Для каждой записи отдельного справочника создается атом в таблице X-Atom. Эта запись связана с записями локальных справочников локальных систем через таблицу X-Link, в которой указывается приложение, справочник и запись справочника.

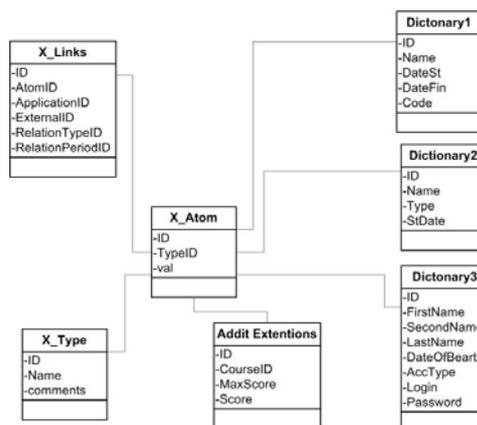


Рисунок 8. Концептуальная схема данных среды интеграции

Детализация связи определяется в таблице X-type, где проставляется время установления связи эталона с локальным справочником и прочие необходимые характеристики. Такая схема связи эталонных справочников позволяет выстроить гибкую работу с наборами данных.

Взаимодействие сервера приложений с веб-сервером реализуется с использованием передачи метаданных в формате XML.

## 5. Дальнейшие направления разработки

Большинство информационных систем, работающих в области государственного управления, разработаны с использованием реляционных баз данных. Предметная область в таких БД описывается языком модели данных на основе ER-диаграмм. В ER-модели бизнес-модель трансформируется в набор сущностей, связанных отношениями и описанных атрибутами. Семантическая модель предметной области может быть воспроизведена отдельно и содержит общую и постоянную информацию (например, «гражданин имеет ровно один адрес жительства») в противоположность конкретным фактам («Петров зарегистрирован по адресу г. Орел ул. Самолетная д.15»). Эта семантическая модель вводит термины для описания предметной области и определяет их значения путем задания взаимосвязей и ограничений. Этот уровень представления данных наиболее близок онтологическому представлению.

В [8] приводится описание того, как элементарные понятия реляционной модели данных, такие как сущности, отношения и атрибуты, могут быть выражены на языке дескриптивной логики (DL) ориентированной на онтологическое описание. По ER-схеме  $S$  создается база знаний  $\phi(S)$ , которая определяется так:

- множество атомарных концептов  $\phi(S)$  состоит из множества сущностей и доменных символов  $S$ ;
- множество атомарных отношений  $\phi(S)$  получается из множества отношений и атрибутивных символов  $S$ ;
- множество аксиом включения  $\phi(S)$  состоит из утверждений формализующих иерархические зависимости между сущностями и отношениями, ограничения на сущности имеющие атрибуты и/или связанные отношениями и ограничения кардинальности для ролей каждого из отношений.

Следует отметить, что онтологическое моделирование как метод интеграции разнородных ресурсов многими авторами рассматривается как перспективное.

## Литература

- [1] Распоряжение Правительства РФ от 15.04.2011 № 654-р «О базовых государственных информационных ресурсах».
- [2] Cyganiak R. State of Play in Linked Open Data [Электронный ресурс]. — Режим доступа: [http://www.slideshare.net/init\\_brussels/cyganiak-richardstateofplaylod](http://www.slideshare.net/init_brussels/cyganiak-richardstateofplaylod)
- [3] Interoperabilidade, Ulrich Schiel, COPIN - Coordenação de Pós-Graduação em Informática [www.slidefinder.net/I/Interop/30827070](http://www.slidefinder.net/I/Interop/30827070)
- [4] Kondylakis H., Flouris G., Plexousakis D. Ontology & Schema Evolution in Data Integration: Review and Assessment [Электронный ресурс]. — Режим доступа: [http://disi.unitn.it/~p2p/RelatedWork/Matching/Kondylakis\\_ODBASE09.pdf](http://disi.unitn.it/~p2p/RelatedWork/Matching/Kondylakis_ODBASE09.pdf)
- [5] Peristeras V., SEMIC.EU: Towards Linked Government Metadata, [Электронный ресурс]. — Режим доступа: [http://www.slideshare.net/init\\_brussels/peristera-svassiliostowardslinkedgovernmentmetadata](http://www.slideshare.net/init_brussels/peristera-svassiliostowardslinkedgovernmentmetadata)
- [6] Peter Mc., Brien A. P Data Integration by Bi-Directional Schema Transformation Rules [Электронный ресурс]. — Режим доступа: <http://portal.acm.org/citation.cfm?id=1783777>
- [7] Когаловский М.П. Методы интеграции данных в информационных системах. [Электронный ресурс]. — Режим доступа: <http://www.cemi.rssi.ru/mei/articles/kogalov10-05.pdf>
- [8] Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В. Онтологии и тезаурусы. Учебное пособие. Казань, М., 2006.

### Infrastructure as reference data integration environment

Yuri Lipuntsov, Eugeny Starikov,  
Vladimir Pshenkin, Vlad Bulatov

Information technology in the public administration is widely used today. Most processes of public administration are cross-departmental and providing e-services requires high level interoperability. For efficient interagency connections it is necessary to create a separate Data Exchange Center - a repository for implementation of two functions: data schema integration and data warehouse for original dictionaries integration and links to local information systems dictionaries. There are some methods to integrate systems. The article presents an application for dictionary data integration. The application is based on modification of ER models. A next step of work is data integration by ontological methods.