



СЕМАНТИЧЕСКАЯ ИНТЕРОПЕРАБЕЛЬНОСТЬ ГОСУДАРСТВЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

Слайд №1 (обложка). Введение

Доклад основан на выполненной в 2007–2008 гг. АНХ при Правительстве Российской Федерации и фирмой КТЕ Labs научно-исследовательской работе по заказу Минэкономразвития России. В работе были предложены проекты нормативного и методического обеспечения ведения единого реестра схем данных, позволяющие федеральным органам исполнительной власти использовать одинаковые схемы данных при информационном обмене. Также был разработан базовый набор схем данных в наиболее востребованных областях (описание физического лица, организации, документа, адресной информации и пр.).

Слайд 2. Интероперабельность информационных систем

При анализе проблем совместимости информационных систем на «верхнем» уровне классифицирования обычно выделяют три типа «совместимостей»: организационную, технологическую и семантическую интероперабельность. В докладе рассматриваются семантические аспекты совместимости, связанные с обеспечением единства (одинаковости) интерпретации информационными системами данных.

Слайд №3. Постановка проблемы. Ориентировочно: масштабы бедствия

В настоящее время способы машинного представления объектов и отношений между ними в пределах одной предметной области (например: адресная информация, названия ведомств, персональные данные граждан, данные кадастрового учета, данные учета юридических лиц и т. п.) различаются в различных органах власти. Более того, даже внутри одного ведомства подразделения зачастую используют различающиеся способы представления данных. Данное обстоятельство серьезно затрудняет взаимодействие государственных информационных систем, поскольку для каждого случая взаимодействия требуются обеспечивать преобразование данных путем доработки информационных систем (ИС).

В отсутствие нормативно-закрепленного способа представления данных в конкретных предметных областях ведомствам приходится заключать специальные соглашения о взаимодействии, число которых уже сегодня весьма значительно и будет увеличиваться с ростом использования ИС в госуправлении. На данный момент число государственных услуг, оказываемых в электронном виде, незначительно. Но с течением времени электронных госуслуг станет больше, кроме того, некоторые электронные формы отчетности становятся обязательными, как, например, сдача налоговой отчетности для налогоплательщиков с числом работников более 100 человек. Если сохранится сегодняшняя ситуация — когда ведомства принуждают взаимодействовать с ними с использованием несогласованных между собой ведомственных форматов данных — непроизводительные расходы граждан и хозяйствующих субъектов на поддержание различных способов представления данных будут расти. Впрочем, уже сегодня для предоставления налоговой, бухгалтерской и статистической информации, банковской и страховой отчетности в электронном виде



организациям приходится поддерживать несколько несогласованных форматов представления одних и тех же данных.

Слайд №4. Постановка проблемы. Что означает обеспечение семантической интероперабельности

Именно потому, что одни и те же типы данных используются разными участниками взаимодействия, то целесообразно иметь общие соглашения о том, в каком виде производится обмен. При этом не обязательно требовать хранить данные в ИС в одинаковых представлениях. Достаточно обеспечить поддержку нормативно закреплённых схем данных при взаимодействии ИС, при этом должно быть запрещено требовать от внешних ИС поддерживать нестандартизованные схемы данных.

Слайд №5. Схемы данных

Схемы данных — ключевое понятие, означающее стандартизованное машиночитаемое описание типового способа представления структурированных данных об определенной разновидности объектов и отношений между ними в определенной предметной области. Схема данных устанавливает формализованные требования к взаимному расположению и иерархии элементов данных, их кодировке, обозначениям, допустимым типам и значениям.

Например, одна из ключевых и наиболее простых схем, описывающих физическое лицо в объеме законодательных требований системы персонального учета населения (СПУН), содержит около 40 иерархически упорядоченных записей. Среди них — Идентификатор персональных данных; Первичные идентификационные данные (фамилия, имя, отчество, дата рождения, место рождения, пол); Основание занесения в СПУН; Данные о согласии субъекта на обработку данных; Лицо, давшее согласие на обработку; Данные, на обработку которых дано согласие; Документ, идентифицирующий личность; Цель обработки данных; Правовое основание обработки данных и др.

Методология классификации схем данных

Очевидно, что задача построения универсальной и единой классификации схем данных «в общем виде» удовлетворительного решения не имеет, поскольку для ее решения требуется создание полной онтологической классификации сферы госуправления, что нереально. Это значит, что инструмента обеспечения семантической согласованности для любого взаимодействия и любой предметной сферы, мы не существуем.

Полностью противоположный подход — индивидуальное согласование каждого отдельно взятого взаимодействия информационных систем хотя и возможен (именно он и реализуется де-факто в России), но обладает очевидными недостатками. В общем случае сложность согласования семантических интерфейсов для N информационных систем пропорциональна $N!$

Соответственно, вместо создания универсальной системы классификации (и соответствующих схем данных «для всех случаев жизни») или ручного «индивидуального» согласования отдельных взаимодействий целесообразно выделить основные задачи, которые могут быть надежно типизированы и описаны с помощью



классификаторов, и ограничить область стандартизации этими границами. Вероятнее всего, определяющими признаками данных, схемы которых подлежат стандартизации, являются, во-первых, возможность разбиения данных на отдельные взаимно упорядоченные информационные элементы с определенным смысловым наполнением (семантикой) и, во-вторых, регулярность (повторяемость) использования одних и тех же структур данных при информационных взаимодействиях.

Слайд №6. Мировой опыт

В мире накоплен практический опыт семантической стандартизации. Тем не менее, при выработке российских методов классификации и учета схем данных необходимо учитывать, что прямое заимствование зарубежной практики, как правило, невозможно в силу правовых и культурных различий. Тем не менее, анализ мирового опыта дает много полезных данных, прежде всего он позволяет выделить черты, характерные для всех государственных проектов стандартизации данных:

- Использование структурированного языка разметки XML в максимально возможном количестве случаев взаимодействия и жесткая стандартизация конкретных форматов данных, где это нецелесообразно (мультимедиа, растровая графика, документооборот).
- Стремление к стандартизации и повторному использованию схем данных для типовых документов и информационных объектов.
- Стандартизация состава и семантики метаданных.
- Создание общедоступных репозиториях (реестров) схем данных, применяемых в государственном информационном обмене.

Слайд №7. Мировой опыт: Великобритания

Наиболее четкие и зрелые подходы к семантической стандартизации данных, используемых при взаимодействии с государственными ИС, демонстрируются в настоящий момент Великобританией (www.govtalk.gov.uk) и Германией (www.kbst.bund.de/xml-technologie).

Большинство государств, получивших на настоящий момент результаты в области учета схем данных, являются либо членами британского содружества, либо входят в зону английского влияния в области права и традиций госуправления (Новая Зеландия, Австралия, Гонконг). Как следствие, они воспроизводят английский подход к стандартизации ИТ (eGIF). Большинство европейских государств, напротив, ориентированы на немецкий опыт.

Европейский союз имеет планы создания общеевропейского репозитория XML-схем, однако вопрос стандартизации информационного обмена находится в исследовательской фазе. На уровне ЕС в основном обсуждается обмен неструктурированными электронными документами и бюджетно-закупочными данными с использованием формата ebXML.

Некоторым особняком в вопросе стандартизации схем данных стоят США. Вопросы установления обязательных требований к ИТ отнесены американским законодательством в основном к компетенции штатов, в подходах которых к стандартизации имеет место большой разноряд. В связи с этим попытки внедрения «Федеральной корпоративной архитектуры» (FEA) встретили активное сопротивление



даже на уровне федеральных ведомств. Федеральное правительство США влияет на стратегию развития межсистемного взаимодействия в основном через механизмы координации и консультирования, а также развитие репозитория для поиска и распространения документов, шаблонов и программного обеспечения XML через Интернет, см. деятельность Совета по ИТ (CIO Council), а также xml.gov.

Слайд № 8. Мировой опыт: Великобритания

Рассмотрим для примера используемую в британском eGIF одноуровневую тематическую классификацию схем. Все схемы в библиотеке GovTalk, как правило, представлены в нескольких форматах, снабжены указаниями по применению и подробно документированы. Публикуются как утвержденные схемы, так и черновые схемы — для их публичного обсуждения. Рубрикатор включает следующие разделы (в скобках указано количество зарегистрированных схем данных):

- Адреса и персональные данные (11)
- Архивы и документооборот (6)
- Департамент окружающей среды, продовольствия и сельского хозяйства (1)
- Образование (6)
- здравоохранение (2)
- Информация о выборах, на базе стандартизуемого OASIS языка EML (5)
- Администрирование государственного шлюза данных (4)
- Безопасный обмен через государственный шлюз (4)
- Географическая информация (2)
- Службы местного самоуправления (27)
- Метаданные (2)
- Планирование (7)
- Закупки (1)
- Собственность (20)
- Транспорт (4)

Слайд №9. Внедрение учёта схем данных

Таким образом в качестве основного механизма решения задачи семантической совместимости государственных ИС предлагаются следующие механизмы:

- Стандартизация представления в электронном виде наиболее часто используемых в информационном обмене объектов и отношений между ними (схем данных) и использование для их представления технологий, основанных на XML.
- Создание единого общедоступного реестра стандартизованных схем данных.

Основными задачами государственного реестра схем данных (далее — РСД), являющегося единым нормативным источником информации о способах представления электронных данных (схемах данных), применяемых при взаимодействии государственных ИС друг с другом, а также с системами граждан и организаций, является снижение затрат на организацию информационного взаимодействия за счет *упрощения доступа* к информации о семантике используемых структур данных, *исключения повторных расходов* на создание схем данных для однотипных объектов, *возможность автоматизации* процедур разработки конверторов и обработчиков данных. Также немаловажно, что стандартизация схем



данных снижает зависимость от разработчика ИС как для государственного заказчика, так и для граждан и организаций, взаимодействующих с государственными ИС (при интеграционных работах отсутствует необходимость обращаться к разработчику для преобразования данных).

Создание, согласование и регистрация схем данных

Объектом учета в РСД являются стандартизованные способы представления структурированных данных при информационном взаимодействии с участием государственных ИС. Способы представления данных (грамматики электронных документов) должны описываться в соответствии с формальными правилами, позволяющими осуществлять их машинную обработку и анализ. Для этого данные должны представляться при информационном обмене в виде электронных документов, размеченных в соответствии со спецификацией XML.

Для формализованного описания таких данных далее применяется термин «схема данных», причем в состав объекта учета помимо собственно схемы, описывающей структуру XML-документов, включается комплекс документов, необходимых для правильной интерпретации и генерации электронных документов. Среди них: схемы, на которые ссылается регистрируемая схема; словари и классификаторы предметной области, которые используются в схеме; словесное (человекочитаемое) описание схемы; набор типовых трансформаций (XSLT) и стилевых описаний (XSL), необходимых для представления данных схемы в удобной для восприятия человеком виде; набор трансформаций (XSLT), обеспечивающих совместимость схемы с ее предыдущими версиями или с унаследованными системами; стандартизованное метаописание схемы, используемое для ее поиска и классификации в реестре схем.

Слайд № 10. Статус реестра схем данных

Реестр схем данных является официальным централизованным информационным ресурсом, аккумулирующим информацию о семантических соглашениях (стандартах), касающихся взаимодействия государственных ИС. С момента полного введения регулирования в области применения схем данных во вновь внедряемых или модернизируемых системах использование для межведомственного информационного обмена, а также при обмене данными с гражданами и негосударственными субъектами не допускается применение незарегистрированных схем данных, а также использование иных технологий представления структурированных данных, чем XML.

Ведение РСД осуществляется уполномоченным государственным органом (оператором реестра). Для технической поддержки информационной системы РСД и для проведения предусмотренных регламентами технических экспертиз могут привлекаться подрядные организации (в порядке, установленном законодательством о закупках для государственных нужд).

Вероятнее всего, функции по ведению РСД необходимо возложить на Федеральное агентство по информационным технологиям Российской Федерации, соответствующие решения должны быть приняты на уровне Правительства Российской Федерации и Министерства связи и массовых коммуникаций Российской Федерации. Эффективная деятельность Общероссийского государственного



информационного центра (ОГИЦ) и достижение практических результатов при переходе к оказанию государственных услуг в электронном виде невозможны без формирования при Центре соответствующих инфраструктур. Не последней среди них является РСД (к числу иных необходимых для деятельности ОГИЦ инфраструктур, вероятно, также относится реестр технических стандартов).

Основными документами, на основании которых осуществляется деятельность реестра, являются нормативный акт, устанавливающий статус реестра, определяющий ответственный за его ведение орган; регламент ведения реестра; технические требования к регистрируемым схемам данных.

Таким образом, с нормативной точки зрения реестр является:

- Официальным источником требований по семантической совместимости информационных систем.
- Средством контроля над соответствием информационных систем государства требованиям по совместимости на уровне данных.

На переходный период деятельность по ведению РСД может не носить нормативного характера. Ведение РСД может быть увязано с работами по разворачиванию ОГИЦ, для чего необходимо сформировать необходимую организационную и технологическую инфраструктуру, что возможно сделать примерно **за 4-6 месяцев.**

Возникновение обязанностей по регистрации схем

Обязательной регистрации в РСД подлежат схемы, которые в ходе предусмотренных регламентами взаимодействий пересекают границы ведомства, в том числе передаются в ИС другого ведомства, предоставляются гражданам, организациям или принимаются от них в машиночитаемом виде. Регистрация схем внутриведомственного распространения рекомендована, но остается на усмотрение самого ведомства.

Ответственность за своевременную регистрацию вновь создаваемых схем данных несет владелец ИС. Ввод в промышленную эксплуатацию систем (в том числе после модернизации), использующих для вневедомственных взаимодействий схемы данных, не зарегистрированные в реестре, не допускается.

Слайд № 10. Обязанности оператора реестра

- Экспертиза разрабатываемых схем на соответствие техническим требованиям, предъявляемым к регистрируемым схемам данных, организация общественного обсуждения и межведомственного согласования, а также итоговая регистрация схем.
- Определение потребности в создании наиболее общеупотребительных схем данных, размещение заказа на их разработку.
- Прием обращений о нарушениях порядка регистрации схем данных при использовании ИС, проверка фактов, изложенных в заявлениях, инициирование устранения нарушений.
- Ведение справочников и классификаторов РСД, методическая поддержка ведомств.
- Экспертиза вводимых в промышленную эксплуатацию ИС, взаимодействующих с гражданами и организациями или иными ведомствами.



Слайды № 11,12. Сервисы реестра схем данных

Также возможна реализация публичных практических сервисов, в том числе:

- Преобразование (трансформация) XML-документов пользователя, соответствующих зарегистрированной в реестре схеме, в удобную для просмотра и чтения форму.
- Сервис заполнения форм (формирование валидных электронных документов на основании введенных пользователем данных по указанной им схеме).
- Предоставление инструментальных средств для проектирования и анализа схем.

Практическая классификация схем

Для упрощения пользования РСД предлагается разработать три различных способа классификации всех регистрируемых в реестре схем:

1. Классификация *по типу взаимодействующих субъектов*, например, Схемы первичных объектов и универсальные схемы; Публичные схемы; Межведомственные схемы; Ведомственные схемы; Унаследованные схемы данных. При этом чем больше субъектов используют схему данных, тем более строгие требования к ней предъявляются.
2. Классификация *по иерархии объектов и отношений*. Иерархическая классификация должна быть таксономической, описывающей иерархию классов и подклассов. Предлагается использовать три группы «верхнего» уровня иерархии:
 - Схемы объектов (классифицируются по их правовой природе), например, Человек, Государство, Субъект Федерации, Орган государственной власти, Орган местного самоуправления, Организации и объединения, Объекты собственности и др.
 - Схемы правовых отношений между объектами, например, Административные, Гражданские, Налоговые, Бюджетные отношения, Конституционные права и др.
 - Прочие схемы (необходимость стандартизации которых очевидна, но которые не могут рассматриваться как объект правоотношений), например, адреса, номера средств связи, географические координаты и т. п. реквизиты, схемы классификаторов и словарей, используемых другими схемами и т. п.
3. *Тематическая (отраслевая) классификация*, в основу которой должна быть положена структура органов государственной власти, Положения о ведомствах, их полномочия и обязанности.