

Интеграция гетерогенных сетей: опыт и проблемы

А. А. Грушо, М. М. Воробьев, А. В. Николаев, В. О. Писковский, В. В. Сенчило, Е. Е. Тимонина

Аннотация — Проблема интеграции гетерогенных информационных систем является актуальной в связи с созданием единого информационного пространства в цифровой экономике. В мире накоплен опыт создания и эксплуатации таких систем.

В статье проведен анализ некоторых примеров с целью выявления проблем, возникающих при эксплуатации и развития информационных систем государственного уровня. Исследование опирается на опыт Нидерландов и Эстонии.

Интеграция информационного пространства в Нидерландах основана на единых технических стандартах и стандартах обмена сообщениями. Язык представления данных, основанных на обмене сообщениями, является достаточно «богатым» для того, чтобы формулировать новые данные, новые запросы, и, тем самым, развивать информационное пространство.

Однако этот язык создает проблемы на стороне получателя сообщения. А именно, экстракция данных из сообщений различных форматов является трудной проблемой. Это связано с оценкой правильности идентификации требуемых данных, а также с оценкой полноты данных, получаемых из различных сообщений.

Согласование языков и форматов сообщений приводит к необходимости создания интеграционных шин, которые более или менее просты при «бедных» языках общения, т.е. все общение должно укладываться в простые шаблоны. Опыт Эстонии основан на использовании «бедных» языков.

Использование «бедных» языков создает проблемы при модернизации гетерогенных информационных систем. Для расширения языков общения в «бедных» языках часто начинают использовать семантические подходы и интеллектуальный анализ данных.

Существенную роль при интеграции гетерогенных распределенных систем играет проблема информационной безопасности. Это связано с тем, что децентрализованная система порождает множество различных политик

безопасности, совмещение которых не всегда возможно из-за их противоречивости.

Ключевые слова— гетерогенные распределенные информационные системы, интеграция распределенных систем, информационная безопасность, цифровая экономика.

I. ВВЕДЕНИЕ

Проблема интеграции гетерогенных информационных систем является актуальной в связи с созданием единого информационного пространства в цифровой экономике.

Один из подходов связан с объединением существующих информационных систем. В ряде стран этот подход реализован на уровне государственных информационных систем.

В статье рассматривается опыт Нидерландов и Эстонии, в которых государственные информационные ресурсы носят распределенный характер. Этот подход многие специалисты предлагают для построения цифровой экономики. В работе показано, что интеграция гетерогенных распределенных информационных систем обходится дороже. Это связано с необходимостью владельцев каждой информационной системы самостоятельно обеспечивать информационную безопасность.

Взаимодействие гетерогенных информационных систем часто упирается в вопросы согласования политик безопасности. Один из вопросов информационной безопасности – надежное резервирование, которое обходится достаточно дорого.

Согласование языков и форматов сообщений приводит к необходимости создания интеграционных шин [1], которые более или менее просты при «бедных» языках общения, т.е. все общение должно укладываться в простые шаблоны [2], [3]. Для расширения языков общения в «бедных» языках часто начинают использовать семантические подходы и интеллектуальный анализ данных.

В статье на примерах Нидерландов и Эстонии показано развитие и интеграция гетерогенных распределенных информационных систем. Эта информация подтверждает перечисленные выше проблемы.

II. СИСТЕМА ИНТЕГРАЦИИ ДАННЫХ В НИДЕРЛАНДАХ

A. История информатизации в Нидерландах

Цифровизация голландской экономики, как и во многих других европейских странах, началась в 80-е годы прошлого века с появлением надежных компьютеров, которые могли хранить значительные объемы

Статья получена 21 марта 2019.

Работа поддержана РФФИ (проект № 18-29-03081-мк).

А. А. Грушо, Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия (e-mail: grusho@yandex.ru).

М. М. Воробьев, Институт программных систем им. А. К. Айламазяна Российской академии наук, Москва, Россия (e-mail: mmv@rnt.ru).

А. В. Николаев, Институт химической физики им. Н. Н. Семенова Российской академии наук, Москва, Россия (e-mail: gentoogion@mail.ru).

В. О. Писковский, Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия (e-mail: vrvr80@yandex.ru).

В. В. Сенчило, Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия (e-mail: volodias@mail.ru).

Е. Е. Тимонина, Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия (e-mail: eltimon@yandex.ru).

информации на внешних носителях. На первых этапах предполагалось иметь реестры регистрации физических и юридических лиц в электронном виде. Кроме реестров лиц в дальнейшем были разработаны и внедрены реестры по кадастровой регистрации объектов недвижимости, сами земельные кадастры и другие необходимые системы.

С развитием компьютерной техники и ее возможностей, в Нидерландах в 90-е и 2000-х тысячные годы уже функционировало большое количество разнообразных информационных систем, которые позволяли местным и центральным властям эффективнее управлять своим муниципалитетом и страной в целом. Перечислим некоторые из этих систем:

- BRS – базовая регистрация физических и юридических лиц;
- DIS – система учета и контроля документооборота с местными органами власти;
- VHS – система регистрации водных сооружений (дамбы, каналы и пр. объекты);
- SoZaWis – ведение дел на биржах труда;
- MILEAU – система контроля состояния окружающей среды;
- FMS – финансовый учет муниципалитетов.

Поражал тот факт, что даже велся учет, какие цветы на той или иной клумбе в городе высажены в текущем году. Фактически, вся необходимая информация в начале 2000-х годов уже находилась в базах данных и эффективно использовалась в экономике этой небольшой страны, ее территориальные и людские ресурсы аналогичны Московской области.

В. Концепция, подход

В 90-е годы прошлого столетия сеть Интернет еще только зарождалась, поэтому власти Нидерландов строили децентрализованную систему управления, иными словами распределенную систему управления экономикой. Это подразумевало следующее, в каждом муниципалитете, а их на тот момент было более 500, организовывался вычислительный центр на основе мощных компьютеров. Большинство муниципалитетов в те годы, да и сейчас, отдавали предпочтение Midrange Computer's компании IBM типа AS/400 (в дальнейшем получившие название iSeries). Но, муниципалитеты могли использовать и любую другую аппаратную платформу. В основном, муниципалитеты не приобретали вычислительную технику, а брали ее в аренду. Это позволяло оперативно обновлять аппаратную платформу с наименьшими затратами, так как уже в те годы шло бурное развитие и рост вычислительных возможностей компьютерной техники.

Указанный выше тип IBM компьютеров мог выполняться в различных модификациях, что позволяло использовать такие компьютеры и в совсем маленьких муниципалитетах с числом граждан от нескольких тысяч и в больших городах, где количество населения приближалось к миллионам граждан.

Главной особенностью техники IBM было наличие встроенной на аппаратном уровне мощной СУБД DB/400. Приобретая такой компьютер с необходимым программным обеспечением, муниципалитет получал полноценный автоматизированную систему управления

своей деятельности. Кроме того, IBM поставляла и продвинутые системы автоматического резервного копирования, которые в ночное время ежедневно вели резервное копирование важных баз данных и электронных документов.

Кроме использования техники IBM, было решено использовать независимую выделенную сеть, которая получила название DataNet1 [4]. Все муниципальные системы имели подключение к этой сети и могли обмениваться любой необходимой информацией между собой.

Функционирование сети DataNet1 осуществлялось основной телекоммуникационной компанией Нидерландов РТТ. Эта компания обеспечивает граждан и компании почтовой, телефонной и телеграфной связью внутри Нидерландов. Технически никто другой не мог подключиться к сети DataNet1, это обеспечивало нужную защиту персональных данных, которые циркулировали в информационном обмене между муниципалитетами и другими гос. структурами (налоговая служба, различные министерства и ведомства и др.).

С. Обмен информацией

По законам Нидерландов, любой гражданин, меняя место жительства, должен в 5-дневный срок посетить муниципалитет, куда он прибыл, и представить информацию кто он и откуда прибыл сюда. На основании этой информации новый муниципалитет отправляет запрос в муниципалитет прежнего проживания и все досье гражданина в электронном виде пересылается в новый муниципалитет.

Чтобы такое взаимодействие между муниципалитетами и другими гос. структурами стало возможным, были разработаны стандарты обмена сообщениями. Кроме того, досье гражданина имело в базах данных гибкую структуру, эта структура была классифицирована и представляла собой набор рубрик. Все типы межсистемных сообщений также была классифицированы и стандартизированы. Перед передачей сообщений вычислялись контрольные суммы, которые проверялись в точке приема сообщения и если они не совпадали, это давало возможность автоматически запросить повторную передачу данных.

За обмен информационными сообщениями отвечал специальный интерфейсный модуль. Т.е. любой модуль такой информационной системы периодически общался с интерфейсным модулем и получал-отправлял «свою» информацию в сеть. Это позволяло не иметь постоянную он-лайн систему, обмен информационными сообщениями происходит в пакетном режиме и у администраторов систем есть возможность наблюдения за прохождением-поступлением сообщений для того или иного модуля системы управления муниципалитетом.

Д. Разработчики ПО

ПО для муниципалитетов могут разрабатывать практически любые компании и их достаточно много, но единое требование – сертифицированная система должна удовлетворять всем требованиям и стандартам информационного обмена GBA [5], [6] (Gemeentelijke Basis Administratie - Муниципальная

Основополагающая Администрация), которые позволяют любой такой системе полноправно использоваться в любой из муниципалитетов Нидерландов. Все подобное ПО проходит предварительное тщательное тестирование перед тем, как начнется его рабочая эксплуатация.

Так как вся информация хранится децентрализованно, а во многих случаях центральным властям требуется какая-то информация в разрезе всей страны. Для получения такой агрегированной информации используется та же система обмена сообщениями, что и описано выше.

Все организации, которые заинтересованы в получении агрегированной информации (это могут быть не только министерства и ведомства, например, религиозные организации имеют право на запрос некоторой информации) регистрируются в центральной системе и получают разрешения на получение той или иной информации – data authorization (DA). А все такие организации называются Afnemer (или клиент, получатель информации).

Таким образом, какая-нибудь религиозная организация может подать заявку на участие в информационном обмене и ей могут предоставить право с определенной периодичностью получать через систему информационного обмена в такой гетерогенной среде, скажем, информацию о количестве мусульман определенного возраста мужского пола, которые проживают в той или иной местности Нидерландов.

Такой запрос будет доставлен во все муниципалитеты и каждый муниципалитет сделает необходимую выборку из локальной базы данных и ответит организации, которая инициировала этот запрос. А уже ПО такого Afnemer принимает информацию от каждого муниципалитета и может агрегировать ее по своему усмотрению.

III. ИНФРАСТРУКТУРА E-GOVERNMENT ESTONIA. ТЕКУЩЕЕ СОСТОЯНИЕ

A. EEBone, начало

Эстония целенаправленно начала развивать свою сетевую правительственную инфраструктуру в середине 90-х. Сначала принимались законодательные акты, утверждались концепции, а с 1998 года работа началась в практической плоскости с создания широкополосной сети EEBone, соединяющей компьютеры во всех правительственных учреждениях [7]. После запуска единой правительственной сети работа продолжилась в направлении создания систем эстонского электронного правительства (e-gov далее).

Следующие постулаты считаются универсальными в современном подходе к e-gov [8]:

- 1) Если данные не могут быть найдены и индексированы в сети, то это равносильно тому, что данные не существуют.
- 2) Если данные не представлены в открытом машинно-читаемом формате, то их невозможно использовать повторно.
- 3) Если какие-либо законодательные акты ограничивают возможность копирования этих данных, то их полезность сильно ограничена.

B. X-Road, концепции

С конца 2001 поверх EEBone был запущен X-Road (X-Tee) – протокол/уровень/шина формализованного обмена данными между правительственными учреждениями (далее будет использован термин "шина"). Шина X-Road заслуживает особого рассмотрения, поскольку в ее дизайне отражается философия эстонского подхода к построению e-gov и даже в какой-то мере современная западная идеология. Шина X-Road является одним из основных компонент e-gov и предлагает следующие сервисы: аутентификацию, авторизацию, мини-портал, регистрацию простых запросов, запросы фиксированной структуры к различным базам данных и регистрам, возможность заполнять регистры, пересылку больших объемов данных по Internet, безопасный обмен данными, запись событийных журналов и поиск по ним, портал для граждан и для операторов, централизованный и локальный мониторинг, описание сервисов в специальной базе данных.

На одном из ресурсов [9] Information System Authority Эстонии (далее RIA) приводится следующий идеологический конструкт как философское обоснование для архитектуры e-gov:

We probably should start stating that knowledge is power. The more data, the bigger power. The main task of a democracy, however, is to distribute power. That leads us to the logical notion that in a democratic society, the databases probably shouldn't grow too large. Each office should have its own database but, leaving entirely aside the issue of pirated copies of illegally collected data, they should not have the copies of the states' remaining databases.

Суть концепции состоит в том, что каждое подразделение правительства должно иметь свою и только свою базу данных и никто не должен иметь полную базу данных (или даже некоторый snapshot) со всех баз данных подразделений правительства, поскольку демократия стремится к распределению власти, а концентрация данных (и связанных с ними знаний) с неизбежностью приведет к концентрации власти. Для работы с удаленными базами пользователи могут использовать заранее согласованный и утвержденный набор запросов фиксированного формата [9]. Запросы свободного формата не позволены.

В технологии X-Road цитированный выше идеологически мотивированный конструкт приводит к нескольким последствиям. Первое и очевидное из них, что каждый департамент, предоставляющий свою БД другим, должен обеспечивать ее доступность (подключенность и работоспособность), иначе, согласно постулату 1) их просто не существует. Это налагает высокие требования на качество и конфигурацию оборудования, квалификацию обслуживающего персонала и т.д. Открытым вопросом остается возможна ли какая-то форма делегирования функции поддержки работоспособности некой структуре или даже размещение на ее площадке серверов, содержащих целевую БД и/или ее копию, снапшот. Если буквально следовать приведенному выше идеологическому конструкту, то такое делегирование невозможно. Однако технические и административные реалии вполне могут сдвинуть границы «идеологически приемлемого», как неоднократно демонстрировала история. В любом

случае, даже если с соблюдением постулата 1) департаменты и получают помощь от какой-то централизованной структуры, то соблюдение постулатов 2) и 3) будет возлагаться на департамент – владелец данных. А значит, в департаменты должны будут привлекать специалистов, способных упаковывать первичные данные в машинно-читаемые форматы и осуществлять их юридическое сопровождение для гарантии возможности их копирования. Для различных департаментов, имеющих дело с разными данными, все три задачи хоть и выглядят типовыми, тем не менее их корректные решения могут кардинально отличаться. Следовательно, департаменты-владельцы столкнутся с проблемой "высокой стоимости решений", а возможность распространения типовых успешных практик между разными профильными департаментами будет не очевидна.

Менее очевидной, но даже более сложной для преодоления, является проблема создания и актуализации форм сложных запросов к таким профильным БД. При развитии БД естественным образом будет эволюционировать ее структура и смысловые связи в ней. Поэтому, если однажды департамент А согласовал набор фиксированных запросов к БД департамента Б, то при эволюции БД последнего возникнет необходимость в пересогласовании форм запросов. Но в то же время организационно департаменты А и Б независимы, могут находиться в различном состоянии (штатная работа, реорганизация, слияние, расширение), что немедленно отразится на процессах согласования форм запросов.

Если же запросы к БД департамента Б происходят не просто от департамента А, а являются частью сложного запроса от департамента В, который запрашивает департамент А, которому в свою очередь для выполнения запроса надо запросить БД департамента Б, то в такой ситуации отчетливо появляется перспектива поддержания множества фиксированных форм запросов комбинаторной сложности, что быстро приведет к неуправляемой ситуации. Данная проблема далее будет упоминаться как "проблема сложных запросов".

Таким образом, на взгляд авторов, из-за описанных выше проблем "высокой стоимости решений" и "сложных запросов" данная система приемлема только для относительно простой структуры e-gov.

C. ID карта

ID карта физически является основным документом, идентифицирующим граждан Эстонии не только в пределах страны, но и в Евросоюзе [10]. Ее функциональность, сходна с функциями отчуждаемых ключевых носителей (безопасная аутентификация, возможность использования данных при генерации электронной цифровой подписи, шифрование-дешифрование передаваемых данных), выдаваемых гражданам в Российской Федерации для работы с ЕСИА и юридически значимого взаимодействия с государственными структурами.

Практика использования ID карт активно внедряется в Эстонии с 2002 года – ими пользуется почти, что у 90% населения страны. На момент

написания настоящей статьи она может применяться в следующих целях:

- электронное голосование (с 2005 года);
- верификация банковских транзакций;
- получение доступа к медицинской карте гражданина (с 2010 года);
- регистрация бизнес-структур;
- использование карты в качестве виртуального билета.

Правительство Эстонии также поощряет использование встроенных в ID карту механизмов для юридически значимого B2B взаимодействия.

При генерации электронной цифровой подписи также используется технология простановки штампов времени (OCSP).

ID карта выдается гражданам, достигшим возраста 15 лет. При выдаче карты гражданину заводится постоянный адрес электронной почты. Срок действия карты – 5 лет. По истечении этого времени гражданин получает новую ID карту. Чип карты содержит в себе следующую информацию:

- файл с персональными данными;
- сертификаты открытых ключей аутентификации (с адресом электронной почты) и цифровой электронной подписи (срок действия сертификатов – также 5 лет);
- соответствующие закрытые ключи, защищенные PIN-кодом.

Кроме ID карты для выполнения аутентификации и работы с электронной цифровой подписью с использованием ЭВМ разработан дополнительный продукт *digi-ID* – отдельная смарт-карта, не являющаяся физическим документом. Для продукта разработаны приложения, функционирующие под распространенными операционными системами (включая мобильные ОС).

Учитывая все вышеизложенное, следует отметить, что частота обновления ключевой информации в Эстонии на отчуждаемых ключевых носителях в отличие от Российской Федерации ниже – это сделано для снижения нагрузки на государственные организации и удобства граждан, но при этом несколько снижает уровень безопасности инфраструктуры электронного правительства в случае выявления уязвимостей в используемых криптографических алгоритмах в ключевых областях.

D. Mobile ID

На базе ID карт в Эстонии широко применяются технологии *smart-ID* и *mobile-ID*, обеспечивающие аутентификацию пользователей онлайн и использование электронной цифровой подписи с мобильного телефона [10]. Технология на данный момент позволяет пользователям:

- производить вход в интернет-банки без ввода паролей;
- подписывать электронные документы с использованием электронной цифровой подписи;
- оплачивать билеты общественного транспорта и парковку.

Для обеспечения работы технологий часть информации (такой, как сертификаты открытых ключей

пользователей) загружаются в память SIM-карты телефона и требуют дополнительных расходов на поддержку со стороны оператора сотовой связи в сотрудничестве с удостоверяющим центром. При использовании технологий для произведения действий пользователь должен ввести PIN-код.

Технологии обеспечивают дружелюбность к пользователю, однако вводят в процессы дополнительного участника – оператора сотовой связи для обеспечения ее работы.

E. DigiDoc

Для обеспечения функционирования механизма электронной цифровой подписи в перечисленных выше технологиях используется семейство форматов DigiDoc – BDOC, DDOC, CDOC [10]. Семейство форматов DigiDoc использует XML и XSD, и основано на соответствующих международных стандартах и RFC, модифицированных под нужды электронного правительства. Для нужд разработчиков предоставляются библиотеки на языках C и Java.

F. Портал eesti.ee

Взаимодействие граждан с государственными службами с использованием перечисленных технологий производится через веб-портал eesti.ee, предоставляющий сервис, схожий с сервисом российского веб-портала gosuslugi.ru.

Пользователи, аутентифицировавшиеся на портале, могут:

- просмотреть и проверить свои персональные данные;
- завести адрес электронной почты в указанном выше домене;
- настроить получение уведомлений электронной почты и SMS;
- производить оплату счетов;
- заполнять обращения;
- получать доступ к реестру сервисов для более, чем 20 национальных баз данных и т.д.

Портал является точкой сбора данных от различных государственных учреждений и с учетом распространенности механизмов электронного правительства его значимость существенно выше по сравнению с российским порталом «Госуслуги».

G. eProcurement Estonia

Процедура закупок автоматизирована с использованием другого веб-портала Procurement Register. Данный портал является аналогом российской Единой информационной системы в сфере закупок и позволяет объявлять тендеры, назначать исполнителей контрактов, производить закупки и проводить электронные аукционы, а также осуществлять проверку участников аукционов путем автоматических запросов в соответствующие государственные ведомства.

H. Document Exchange Centre (DVK)

Взаимодействие государственных организаций друг с другом осуществляется через общую центральную систему обмена документами Document Exchange Centre, поддерживающую различные корпоративные системы управления контентом и системы документооборота [10]. Система использует для

передачи документов транспортную инфраструктуру X-Road и обеспечивает кратковременное хранение документов и обработку документов.

Для резервирования и обеспечения бесперебойности работы государственной информационной системы Эстония разрабатывает и планирует реализовать концепцию «посольства данных», заключающуюся в безопасном размещении копий всей информации и критичных для функционирования государства информационных систем на территории других государств с использованием облачных технологий. По мнению разработчиков, реализация данной концепции сделает разработанную инфраструктуру более устойчивой во время природных катастроф или кризисов, вызванных человеческим фактором.

I. Перспективы семантических технологий в e-gov

Идея применения семантических технологий в глобальной сети, вероятно, была высказана еще до широкого распространения интернета. В 1998 году Тим Бернерс-Ли предложил термины "семантическая паутина" (semantic web) или гигантский глобальный граф (giant global graph, or GGG akin WWW). Реализация данной концепции позволит использовать формализованные, а значит хорошо алгоритмизируемые методы для анализа разнородных данных из разных источников. Естественно, что для решения задачи построения хорошо функционирующего и высокоавтоматизированного e-gov наличие таких технологий принципиально.

Интерес к технологиям искусственного интеллекта (ИИ) и машинного обучения в последние годы только добавляет актуальности семантическим технологиям т.к. широкое внедрение этих технологий позволит использовать системы ИИ для анализа практически полезных данных, собираемых государственными структурами различных стран.

К сожалению, опыт показал, что из-за необходимости реверсивного инжиниринга, документирования элементов данных, устранения ошибочных данных, соотнесения схем данных и конкретных их экземпляров требуются существенные вложения в создание Linked Data Sets (LDS) и Linked Government Data (LGD) в частности [11]. Таким образом, алгоритмизация и автоматизация работы по созданию LDS является актуальной проблемой больших информационных систем длительного использования.

Закономерно, что специалисты RIA не обошли вниманием семантические технологии как таковые, сделав в своих работах упор на применение их в эстонском e-gov [12]. Стоит отметить, что применение семантических технологий для эстонского варианта e-gov является не опцией, а ключевой необходимостью, т.к. архитектура и технические аспекты функционирования базовых подсистем подчинены политическим идеологемам. Разделение баз данных и административное предотвращение концентрации информации под управлением даже сугубо технической правительственной службы ведет к практической невозможности выстраивать взаимосвязи между разнородными данными, собираемыми государством, согласно единой политике в предсказуемом и запланированном порядке с помощью известных

технологий. Поэтому, практически единственной надеждой на реализацию многих возможностей, необходимых для продолжительного функционирования e-gov, является внедрение программ, реализующих семантические технологии. Например, изменение онтологических связей между большими объемами децентрализованных данных в ручном режиме с помощью известных методов приведет к непредсказуемому (скорее хаотическому) поведению системы, находящейся в неких промежуточных состояниях в процессе трансформации.

Возможные альтернативы, впрочем, не более привлекательны – придется либо отключить все вовлеченные правительственные информационные системы на период внесения изменений, либо создать полностью дублированные системы одну оставив работать, а вторую модифицировать, тестировать, запускать в рабочий режим. После запуска модифицированной системы в нее надо будет внести изменения, которые накопились за время эксплуатации дублирующей системы. При наличии программ, реализующих семантические технологии, указанный выше переход между онтологическими схемами может быть реализован без выведения системы из эксплуатации либо при кратковременном ее выключении.

Работы, демонстрирующие возможность автоматического создания LDS, авторам не известны, однако многие аспекты проблемы сейчас активно изучаются. Например, предложены методы автоматизированного поиска ключей для создания связей между данными [13], автоматического связывания объектов онтологии и имен атрибутов (колонок) баз данных [14]. Таким образом, один из ключевых компонентов эстонского e-gov пока не имеет законченного технологического решения.

В качестве организационной альтернативы предлагается подход в чем-то схожий с идеологией открытого программного обеспечения. В рамках предлагаемого подхода пользователи вовлекаются в кооперативную работу над разнородными данными в рамках TWC LOGD (Linked Open Government Data) экосистемы [15]. Предоставляя платформу, производительные инструменты для работы с данными и средства для публикации своих наборов данных пользователям, владельцы порталов не только открывают новые возможности для бизнеса, создания аналитических продуктов в нужных областях, но также накапливают определенный багаж семантических описаний представленных данных, который они уже могут использовать в своих целях. Чтобы пользоваться спросом у конечных пользователей, подобная система должна поддерживать ключевые современные семантические технологии: RDF/RDFS (англ. Resource Description Framework/Resource Description Framework Schema), OWL (Web Ontology Language), SPARQL (рекурсивный акроним от англ. SPARQL – Protocol and RDF Query Language) [8]. Некоторым аналогом TWC LOGD портала [15] является проект SKAN [16], получивший довольно широкое распространение в последние годы как ПО для организации OGD площадок отдельных городов и регионов. В рамках SKAN, в свою очередь, стимулируется написание разнообразных плагинов, расширяющих возможности

стандартного ПО SKAN. В качестве одного из примеров ПО, базирующегося на SKAN, можно привести проект архитектуры LIRE (Linked Relations) [17]. LIRE позволяет пользователям искать, управлять, интегрировать, публиковать и повторно использовать установленные связи между базами данных. По-видимому, именно SKAN и, возможно, LIRE рассматривается RIA для нужд проекта эстонского e-gov как перспективная LOGD платформа.

Результаты анализа широко разрекламированного проекта эстонского e-government показывают, что базовые технологии взаимодействия гражданина с электронным правительством уже отработаны, и предоставляют удобные для гражданина методы взаимодействия с государственными службами. Безусловно, существуют некоторые проблемы безопасности, скорость внедрения и распространения передовых решений оставляет желать лучшего. Однако, в общем и целом процесс запущен и можно с уверенностью утверждать, что технология позволяет всеобъемлющее внедрение электронных идентификационных документов гражданина, возможность его удаленного взаимодействия с электронным правительством.

В тоже время технологии работы с данными внутри правительственных структур пока еще не отработаны в должной мере. А некоторые технологии, такие как автоматическая интеграция разнородных данных, автоматическое динамическое изменение онтологических связей между данными еще не реализованы. Безусловно, семантические технологии будут играть ключевую роль в решении обозначенных проблем.

IV. ПРОБЛЕМЫ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В РАЗЛИЧНЫХ ИНТЕГРАЦИОННЫХ ПОДХОДАХ

Безопасность интегрированных информационных систем, рассмотренных выше, основана на механизмах аутентификации и авторизации обмена данными.

Формализация шаблонов в эстонской системе (бедный язык) гарантирует ограничение доступов к данным в распределенной системе. Вместе с тем, сбор персональных данных вполне допустим в системе Эстонии и, тем более, в системе Нидерландов. Полная ответственность за данные в информационных системах в условиях децентрализации, как было отмечено выше, создает повышенную нагрузку на бюджет собственников этих систем. Отсюда следует, что интегральная информационная безопасность распределенной системы не может не обладать уязвимостями. Даже в условиях резервирования уничтожение информации муниципалитета или региона с точки зрения информационной безопасности является критическим моментом.

Вместе с тем, чем более богат язык обмена сообщениями, тем больше проблем возникает в обеспечении информационной безопасности. Это связано с противоречиями в политиках безопасности и возможностями совмещения требований по безопасности изолированных информационных систем.

V. ЗАКЛЮЧЕНИЕ

Интеграция информационного пространства в Нидерландах основана на единых технических стандартах и стандартах обмена сообщениями. Язык представления данных, основанных на обмене сообщениями, является достаточно богатым для того, чтобы формулировать новые данные, новые запросы, и, тем самым, развивать информационное пространство.

Однако этот язык создает проблемы на стороне получателя сообщения. А именно, экстракция данных из сообщений различных форматов является трудной проблемой. Это связано с оценкой правильности идентификации требуемых данных, а также с оценкой полноты донных, получаемых из различных сообщений.

Существенную роль играет проблема информационной безопасности в едином информационном пространстве при обмене сообщениями. Это связано с тем, что децентрализованная система порождает множество различных политик безопасности, совмещение которых не всегда возможно из-за их противоречивости.

Для такой небольшой страны, как Нидерланды, эти проблемы являются трудными, но преодолимыми. Однако в масштабах Российской Федерации такой подход к интеграции может привести к хаосу, или к появлению и распространению недостоверных данных.

Анализируя опыт Эстонии, как было отмечено выше, приходим к проблеме трудности модернизации системы, основанной на бедных языках. Подход, основанный на семантическом расширении формализованного языка, быстро приводит к тем же проблемам, которые описаны выше в связи с интеграцией на базе сообщений. Более того, использование онтологий – трудно автоматизируемый процесс, и требует доказательства полноты и непротиворечивости в реализации такого подхода.

Таким образом, семантическое расширение бедных языков не упрощает, а усложняет проблему интеграции и проблему обеспечения информационной безопасности.

БИБЛИОГРАФИЯ

[1] Забейко М. И. *Банковский бизнес в России: индустрия или искусство?* Волгоград: Волгоградское научное изд-во, 2009.

- [2] А. А. Грушо, Н. А. Грушо, Е. Е. Тимонина, С. Я. Шоргин, “Возможности построения безопасной архитектуры для динамически меняющейся информационной системы,” *Системы и средства информатики*, Т. 25, № 3, С. 79–93, 2015.
- [3] А. А. Грушо, Д. В. Смирнов, “Защита бизнес-логики от атак нулевого дня,” *Системы и средства информатики*, Т. 26, № 3, С. 61–73, 2016.
- [4] Wide Area Networks en Datanet 1, Samsom BedrijfsInformatie, EAN: 9789014044484, 1992.
- [5] Specificatie datacommunicatienetwerk GBA, GBA-projectbureau, 's-Gravenhage, 1990.
- [6] Logisch Ontwerp GBA ver.3.9, Available: <https://www.rvig.nl/binaries/rvig/documenten/richtlijnen/2015/01/15/1ogisch-ontwerp-gba-3-9/lo-gba-3-9.pdf>.
- [7] K. Paide, “Analysing Estonian X-Road as a collaboration tool for public-private partnerships,” M.S. thesis, Dept. Software Science, Tallinn Univ. of Technology, Tallinn, Estonia, 2017.
- [8] P. Milić, N. Veljković and L. Stoimenov, “Semantic Technologies in e-government: Toward Openness and Transparency,” in *Smart Technologies for Smart Governments*, R. Bolivar, Ed. Springer, Cham, 2018, pp 55–66.
- [9] A. Veldre (2016, February). Introduction of X-tee [Online]. Available: <https://www.ria.ee/en/state-information-system/x-tee/introduction-x-tee.html>.
- [10] Digital Agenda 2020 for Estonia. Ministry of Economic Affairs and Communications [Online]. Available: https://www.itl.ee/public/files/DigitalAgenda2020_Estonia_ENG.pdf
- [11] Putting the frontline first: smarter government. UK Government policy, 2009 [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/228889/7753.pdf.
- [12] Estonian Information Society YEARBOOK 2011/2012. Ministry of Economic Affairs and communications publication [Online]. Available: <https://www.digar.ee/arhiiv/ru/download/214871>.
- [13] M. Achichi, M. Ellefi, D. Symeonidou, K. Todorov, “Automatic Key Selection for Data Linking,” in *Knowledge Engineering and Knowledge Management*, E. Blomqvist, P. Ciancarini, F. Poggi, F. Vitali, Ed. Springer, Cham, 2016, pp. 3–18.
- [14] S. Profliet, R. Ichise, “Automated mapping generation for converting databases into linked data,” in *Proc. 2010 Int. Conf. on Posters & Demonstrations Track*, Shanghai, pp.173–176, 2010.
- [15] L. Ding, T. Lebo, J. Erickson, D. DiFranzo, A. Graves, G. Williams, X. Li, J. Michaelis, J. Zheng, Z. Shangguan, J. Flores, D. McGuinness, J. Hendler “TWC LOGD: A Portal for Linked Open Government Data Ecosystems,” *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 9, no, 3, pp, 325–333, 2011.
- [16] CKAN User Guide [Online] Available: <https://docs.ckan.org/en/latest/user-guide.html>
- [17] P. Milić, N. Veljković and L. Stoimenov, “Linked Relations Architecture for Production and Consumption of Linksets in Open Government Data,” *Open and Big Data Management and Innovation*, M. Jassen, M. Mäntymäki, J. Hidders, B. Klievink, W. Lamersdorf, B. Loenen, A. Zuiderwijk Ed. Springer, Cham, 2015, pp. 212–222.

Integration of Heterogeneous Networks: Experience and Problems

A. A. Grusho, M. M. Vorobiev, A. V. Nikolaev, V. O. Piskovski, V. V. Senchilo, E. E. Timonina

Abstract — The problem of integration of heterogeneous information systems is relevant in connection with creation of a common information space in digital economy. In the world an experience of creation and operation of such systems is accumulated.

In article the analysis of some examples for the purpose of identification of the problems arising at operation and development of information systems of the state level is carried out. The research relies on experience of the Netherlands and Estonia.

Integration of information space in Netherlands is based on uniform technical standards and standards of message exchange. Language of data view, based on message exchange, is rather "rich" to express new data, new requests, and, thereby, to develop information space.

However, this language creates problems on the party of the message recipient. Namely, extraction of the different formats given from messages is a complex problem. It is connected with assessment of correctness of identification of required data and also with assessment of completeness data, which were received from different messages.

Coordination of languages and the message formats results in need of creation of integration buses, which are more or less simple at "poor" languages of communication, i.e. all communication should keep within simple templates. Experience of Estonia is based on use of "poor" languages.

Usage of "poor" languages creates new problems at upgrade of heterogeneous information systems. For expansion of languages of communication in "poor" languages often begin to use semantic approaches and intelligent data analysis.

An essential role at integration of heterogeneous distributed systems is played by a problem of information security. It is connected with the fact that the decentralized system generates a set of different security policies for which their union is not always possible because of their inconsistency.

Keywords — heterogeneous distributed information systems, integration of distributed systems, information security, digital economy.

REFERENCES

- [1] M. I. Zabezhylo, *Banking business in Russia: industry or art?* Volgograd: Volgograd scientific publishing house, 2009.
- [2] A. Grusho, N. Grusho, S. Shorgin and E. Timonina, "Possibilities of Secure Architecture Creation for Dynamically Changing Information Systems," *Systems and means of informatics*, vol. 25, no. 3, pp. 78–93, 2015.
- [3] A. A. Grusho, D. V. Smirnov, "Protection of business logic against zero day attacks," *Systems and means of informatics*, vol. 26, no. 3, pp. 61–73, 2016.
- [4] Wide Area Networks en Datanet 1, Samsom BedrijfsInformatie, EAN: 9789014044484, 1992.
- [5] Specificatie datacommunicatienetwerk GBA, GBA-projectbureau, 's-Gravenhage, 1990.
- [6] Logisch Ontwerp GBA ver.3.9, Available: <https://www.rvig.nl/binaries/rvig/documenten/richtlijnen/2015/01/15/ogisch-ontwerp-gba-3-9/lo-gba-3-9.pdf>.
- [7] K. Paide, "Analysing Estonian X-Road as a collaboration tool for public-private partnerships," M.S. thesis, Dept. Software Science, Tallinn Univ. of Technology, Tallinn, Estonia, 2017.
- [8] P. Milić, N. Veljković and L. Stoimenov, "Semantic Technologies in e-government: Toward Openness and Transparency," in *Smart Technologies for Smart Governments*, R. Bolivar, Ed. Springer, Cham, 2018, pp 55–66.
- [9] A. Veldre (2016, February). Introduction of X-tee [Online]. Available: <https://www.ria.ee/en/state-information-system/x-tee/introduction-x-tee.html>.
- [10] Digital Agenda 2020 for Estonia. Ministry of Economic Affairs and Communications [Online]. Available: https://www.itl.ee/public/files/DigitalAgenda2020_Estonia_ENG.pdf
- [11] Putting the frontline first: smarter government. UK Government policy, 2009 [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/228889/7753.pdf.
- [12] Estonian Information Society YEARBOOK 2011/2012. Ministry of Economic Affairs and communications publication [Online]. Available: <https://www.digar.ee/arhiiv/ru/download/214871>.
- [13] M. Achichi, M. Ellefi, D. Symeonidou, K. Todorov, "Automatic Key Selection for Data Linking," in *Knowledge Engineering and Knowledge Management*, E. Blomqvist, P. Ciancarini, F. Poggi, F. Vitali, Ed. Springer, Cham, 2016, pp. 3–18.
- [14] S. Profliet, R. Ichise, "Automated mapping generation for converting databases into linked data," in *Proc. 2010 Int. Conf. on Posters & Demonstrations Track*, Shanghai, pp.173–176, 2010.
- [15] L. Ding, T. Lebo, J. Erickson, D. DiFranzo, A. Graves, G. Williams, X. Li, J. Michaelis, J. Zheng, Z. Shanguan, J. Flores, D. McGuinness, J. Hender "TWC LOGD: A Portal for Linked Open Government Data Ecosystems," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 9, no, 3, pp, 325–333, 2011.
- [16] CKAN User Guide [Online] Available: <https://docs.ckan.org/en/latest/user-guide.html>.
- [17] P. Milić, N. Veljković and L. Stoimenov, "Linked Relations Architecture for Production and Consumption of Linksets in Open Government Data," *Open and Big Data Management and Innovation*, M. Jassen, M. Mäntymäki, J. Hidders, B. Klievink, W. Lamersdorf, B. Loenen, A. Zuiderwijk Ed. Springer, Cham, 2015, pp. 212–222.