

DOI: 10.18721/JE.12315

УДК 338.2; 004.652

## ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОГО ВЗАИМОДЕЙСТВИЯ АГЕНТОВ ЦИФРОВОЙ ЭКОНОМИКИ НА ОСНОВЕ ОНТОЛОГИЧЕСКОГО ОПИСАНИЯ ПРЕДМЕТНОЙ ОБЛАСТИ

**Ю.П. Липунцов**

Московский государственный университет имени М.В. Ломоносова,  
г. Москва, Российская Федерация

Организация информационного обмена в условиях активного производства данных является одной из основных задач цифровой экономики. Большая часть организаций использует информационные системы, призванные выполнять информационную поддержку деятельности, но не ориентированы на информационное взаимодействие. Создание системы взаимодействия должно опираться на содержательный анализ логики деятельности, представленной в системах источниках данных. Рассматривается вариант модели данных хранилища, в качестве методической базы которого использованы методы онтологического моделирования. На первом этапе создается онтология источников данных, а затем объединенная онтология. Схематичное представление онтологии предметной области предложено в форме иерархии, включающей две категории базовых компонент – первичные и производные. Такое представление предметной области позволяет организовать связь с моделью реального мира, адекватно отражающую основные особенности хозяйственной деятельности. В качестве основы для создания модели используются базовые объекты предметной области, а также транзакции, выполняемые с участием этих объектов. В создании таких моделей ведущую роль должны выполнять эксперты предметной области. Логическая модель хранилища строится с использованием метода Data Vault. Методика составления модели данных хранилища включает следующие этапы: анализ моделей данных источников, создание графического представления онтологии для источников, создание графического представления объединенной онтологии, создание модели данных хранилища. Отражены практические аспекты реализации информационного обмена, полученные в ходе выполнения проектов по теме в различных предметных областях. Представленная последовательность действий по созданию модели данных хранилища как основа для информационного обмена может быть использована для организации взаимодействия систем как в корпоративном, так и в государственном секторе. Особенностью выполнения проектов по интеграции данных в государственном секторе является ответственность производителей, в том числе персональных данных, за их сохранность. Представленная модель удовлетворяет предъявляемым со стороны таких заказчиков требованиям.

**Ключевые слова:** информационное моделирование, хранилище данных, информационный обмен, стандартизация данных, онтология предметной области, базовые компоненты, производные компоненты

**Ссылка при цитировании:** Липунцов Ю.П. Организация информационного взаимодействия агентов цифровой экономики на основе онтологического описания предметной области // Научно-технические ведомости СПбГПУ. Экономические науки. 2019. Т. 12, № 3. С. 172–186. DOI: 10.18721/JE.12315

## ORGANIZATION OF INFORMATION EXCHANGE OF DIGITAL ECONOMY AGENTS BASED ON ONTOLOGICAL DESCRIPTION OF SUBJECT AREA

**Y.P. Lipuntsov**

Lomonosov Moscow State University, Moscow, Russian Federation

Most sectors of the economy have gone through an elementary stage of informatization. Most organizations use information systems designed to carry out information support activities but are not focused on information interaction. Creating a system of interaction should be based on comprehensive analysis of the business logic presented in the systems of data sources. The paper considers a variant of the warehouse data model with ontological modeling methods used as a methodological base. The central role in constructing such models should be played by subject matter experts. At the first stage, an ontology of data sources is created, and then a combined ontology. A schematic representation of the domain ontology in the study is proposed in the form of a hierarchy that includes two categories of components: primary basic components and derived basic components. The logical model of storage is built using the Data Vault method. The methodology for compiling a data model of a repository includes the following steps: analyzing source data models, creating a graphical representation of ontology for sources, creating a graphical representation of a unified ontology, creating a data model of the repository. The paper reflects the practical aspects of implementation of the information exchange obtained in projects in various subject areas. The given sequence of actions for creating a data warehouse model as a basis for information exchange can be used to organize the interaction of systems in both the corporate and the public sector. A specific feature of projects for data integration in the public sector is the responsibility of data producers for personal data safety. The model presented meets the requirements.

**Keywords:** information modeling, data warehouse, information exchange, data standardization, domain ontology, basic components

**Citation:** Y.P. Lipuntsov, Organization of information exchange of digital economy agents based on ontological description of subject area, St. Petersburg State Polytechnical University Journal. Economics, 12 (3) (2019) 172–186. DOI: 10.18721/JE.12315

*Введение.* В цифровой экономике ведомства, коммерческие компании, организации вовлечены в информационный оборот, выступают в роли производителей и потребителей данных. Основой для организации деятельности экономических агентов, управления экономикой становится совокупность циркулирующих между агентами сведений.

В условиях активного информационного обмена одной из актуальных задач становится совершенствование методов взаимодействия. Участники обмена данными используют различные технологические решения. Основой для объединения существующего разнородного набора систем и создания устойчивого решения для их взаимодействия может выступать содержательный взгляд на выполняемую экономическую деятельность. Организация движения ин-

формации между существующими информационными системами отдельных экономических агентов предполагает активное участие экспертов предметной области, которые хорошо понимают суть экономической деятельности, в рамках которой предполагается организация взаимодействия информационных систем.

Результатом деятельности экспертов предметной области является онтологическая модель, представляющая собой формализованное отражение основных компонент предметной области и их соотношений. Онтологическая модель является методологической базой для создания информационной модели, на основе которой будет организовываться информационное взаимодействие.

С архитектурной точки зрения перспективным вариантом организации обмена является

создание центрального ядра, аккумулирующего атомарные данные источников. Этот вариант активно используется в практике информационного обмена в государственном секторе Японии [1], Евросоюза [2]. Создание центрального ядра, аккумулирующего данные локальных источников, можно реализовать различными методами. Здесь рассмотрен вариант, который позволяет организовать хранение данных локальных источников в одном хранилище, его модель данных отражает основные элементы бизнес-логики локальных систем и учитывает их пересечение.

Целью исследования является описание содержания этапа и представление последовательности шагов создания модели данных на основе онтологии. Организация взаимодействия агентов цифровой экономики с использованием хранилища данных предполагает выполнение следующих задач:

- разработку стандарта формата данных для обмена;
- создание модели данных для хранилища;
- разработку системы кодификации данных для обмена;
- разработку процедур трансформации данных для загрузки в хранилище.

Наиболее ответственный этап – создание модели данных, которая должна содержать все необходимые элементы данных из источников, а также корректно представлять области пересечения источников данных между собой.

*Методика исследования.* Рассматривается последовательность шагов для создания модели данных хранилища на основе формализованного представления предметной области. Обзор литературы по теме «Трансформация модели деятельности в модель данных», а также опыт по созданию приложений для интеграции показал несостоятельность методов автоматической генерации элементов модели данных хранилища на основе источников данных. Неудовлетворительные результаты получаются при применении методов агрегации моделей источников в модель данных хранилища путем построения традиционной реляционной модели. Рассмотрен вариант построения онтологии и реализации модели Data Vault на основе этой онтологии. Далее будут рас-

смотрены: *построение концептуальной модели хранилища в форме онтологии, создание онтологии на основе модели данных систем источников, с описанием шагов для анализа моделей данных источников, их графического представления в форме онтологии, и создание единого графического представления онтологии для объединенной предметной области.* Конечный этап исследования – построение логической модели данных хранилища на основе онтологии.

#### **Концептуальная модель хранилища в форме онтологии**

Сбор и совмещение в одном хранилище данных из нескольких локальных систем – задача с методической точки зрения не очень простая, поскольку для этого нужно представить информацию о различных разделах деятельности, отраженных в источниках данных, в единой модели [3].

По сути создается новая модель с собственным взглядом на всю совокупность отражаемых в локальных системах процессов и данных о них. Модель предполагает соединение совокупности функций, поддерживаемых в системах источников, в объединенном виде. Такая модель может быть использована как для обмена данными между системами, так и для поставки данных для задач, решение которых предполагает анализ данных нескольких источников.

Для построения информационного обмена между совокупностью систем необходимо построение целостной информационной модели концептуального уровня, включающей основные разделы предметной области, отображенные в источниках данных. При создании информационных систем – источников данных происходит транслирование логики предметной области на элементы приложений – модели данных, процедуры обработки данных, различные интерфейсы и т. д. Появляется точка зрения на выполняемую деятельность с позиций информационной системы. Количество точек зрения, подлежащих объединению в модели данных хранилища, будет равно числу информационных систем, которые мы собираемся интегрировать. При этом точки зрения будут различными. Задача концептуальной модели состоит в выработке единой точки

зрения, желательно такой, которая сближала бы точки зрения экспертов предметной области и ИТ-специалистов.

Под концептуальной моделью будем понимать модель предметной области, независимой от конкретного приложения или контекста использования данных [4]. Поскольку нас интересует концептуальная модель хранилища данных, предназначенного для хранения атомарных данных источников, то концептуальная часть предполагает представление следующих аспектов:

- отражение данных всех источников;
- представление актуального состояния по каждому источнику данных;
- сохранение истории изменения данных в источниках.

Отражение в хранилище атомарных данных всех источников предполагает форму организации данных хранилища таким образом, чтобы, с одной стороны, все принципиальные элементы, описывающие экономическую деятельность в разных источниках данных были представлены в единой модели данных хранилища, и с другой – модель данных должна давать возможность сохранить разное описание одинаковых объектов для разных источников данных. Такая задача несколько отличается от большинства инструментов анализа данных, которые предполагают предварительную очистку данных перед загрузкой в аналитическую среду.

Для создания модели концептуального уровня [5, 6] необходимо описать модель структуры и модель поведения. Помимо этого, в [5] отдельным блоком представлен раздел метамодели, описывающий принципы метамоделирования. Единое описание различных аспектов хозяйственной деятельности, представленных в отдельных источниках данных, возможно путем построения онтологической модели, которая будет представлять собой модель структуры предметной области в форме совокупности компонент, связанных между собой. Динамика будет представлена в форме изменения этих компонент в ходе экономической деятельности.

В классическом понимании онтология воспринимается как знание о предельно общем [7]. В информатике под онтологией понимают попытку всеобъемлющего формализованного пред-

ставления некоторой области знаний с помощью определенной концептуальной схемы [8].

В нашем случае онтология будет представлять собой формализованное отражение предметной области посредством состава концептов и их зависимостей. Зависимости концептов представим в данном случае в форме иерархии, хотя возможен более сложный вариант зависимости – в форме графа.

При переходе на схематичное представление онтологии концепты заменяются на компоненты. Корнями иерархического представления онтологии будут выступать *базовые компоненты*, которые не зависят от других компонент. Ветви графа отражают *производные базовые компоненты*.

Базовая компонента – это основной элемент онтологической модели. Базовые компоненты представляют собой отражение объектов реального мира либо могут быть артефактами, отражающими совокупность фактов реального мира.

Производные базовые компоненты являются результатом транзакций, выполняемых в пределах предметной области.

Информационная модель является отражением слоя деятельности на информационный слой. Структурированное представление деятельности, например, в формате нотации IDEF0, трансформируется в модель данных, например, в формате ER-диаграммы.

Рассмотрим это на примере. Базовыми компонентами предметной области *управление кадрами* могут выступать *персона* и *вакансия подразделения*. В результате операции *прием на работу*, появляется производная базовая компонента *сотрудник*. В формате IDEF0 это действие отображено на рис. 1.

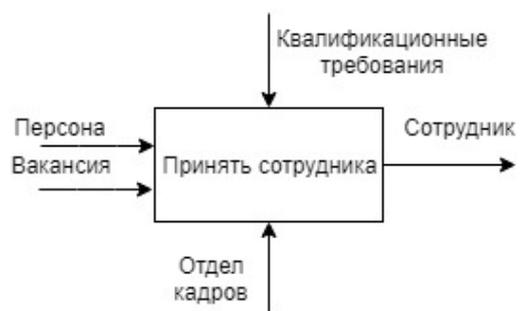
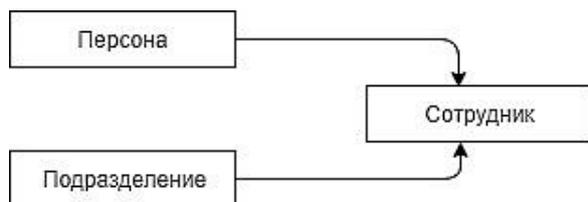


Рис. 1. Отображение действия *прием сотрудника*  
Fig. 1. Displaying the *Employee Reception* action

В графическом представлении онтологии это будет отражено в форме связи двух базовых компонент *персона* и *подразделение* с производной базовой компонентой *сотрудник* (рис. 2). Связи схематичного отражения онтологии можно интерпретировать как правила деятельности (бизнес-правила): «Персона, отвечающая квалификационным требованиям, может быть принята на вакансию подразделения организации».



**Рис. 2.** Графическое представление онтологии  
**Fig. 2.** Graphic representation of ontology

Эта транзакция может снабжаться такими характеристиками, как номер приказа, срок начала и окончания работы, оклад и пр. Можно сказать, что бизнес-правила, описывающие хозяйственные операции, формулируются с использованием терминов, обозначающих задействованные в этих операциях базовые компоненты. Описание этих терминов представляется в форме статей контролируемого словаря, являющегося частью документации проекта.

Вопросы перевода формального отражения бизнес-процессов в модель данных локальной системы описаны в [9, 10].

В нашем случае решается задача создания онтологии для модели данных хранилища на основе анализа существующих приложений. Для построения модели данных хранилища выполняется следующая последовательность действий:

- проведение анализа моделей данных локальных систем;
- построение графического представления для каждого источника данных;
- выявление общих сущностей;
- построение единого графического представления онтологии
- создание модели данных и диаграммы классов.

В результате построения концептуальной модели данных хранилища создается графическое представление онтологии, включающее все базовые компоненты хранилища.

Построение концептуальной модели хранилища является наиболее ответственным этапом создания хранилища. Изменение концептуальной модели после введения хранилища в эксплуатацию будет достаточно трудоемким мероприятием, поскольку может потребовать изменения схемы данных, многих процедур кодификации и преобразования данных. На основе концептуальной модели создается логическая модель, отражающая состав атрибутов сущностей и формы связи между сущностями.

Физическая модель данных источников будет меняться в силу развития организации. При этом концептуальная модель не должна претерпевать существенных изменений. Устойчивость концептуальной модели зависит от ее адекватности предметной области, что могут обеспечить эксперты предметной области, способные выделить в определенной сфере экономической деятельности существенные для информационного моделирования объекты.

#### **Создание онтологии на основе модели данных систем источников**

Основная часть бизнес-логики предметной области, использующей информационную систему, как правило, отражается в модели данных. Под моделью данных будем понимать совокупность объектов предметной области, а также действия с этими объектами, события, обстоятельства, в которых эти действия происходили, а также связи между ними. Зная модель данных источника и понимая основные принципы ее составления, можно выделить основные термины предметной области, на основе которых создается онтология.

Онтология, как формализованное отражение предметной области, может содержать термины, которые не используются в традиционном естественном языке. Например, термин «группа» из предметной области *высшее образование* часто ассоциируется с термином «академическая группа» студентов. Вместе с тем содержание термина «группа» на семинаре, на поточной лекции, на дисциплине по выбору, либо на производственной практике может сильно отличаться. В обычной практике используется понятие, близкое к понятию «академическая группа», но этого недостаточно для корректного отражения с точки зрения информационного моделирования. При построении онтологий могут появляться новые

термины, не свойственные для естественного языка, но необходимые для корректного формализованного отражения предметной области.

Все термины предметной области, в том числе и те, которые не встречаются в описании предметной области с использованием традиционного естественного языка, необходимо включать в качестве самостоятельных статей в контролируемый словарь документации проекта и раскрывать их содержание.

Модели данных источников строятся на основе методологии, основные принципы которой отображены у М. Блеха [11], где рассматриваются пять видов стандартных типов шаблонов, используемых для составления модели данных: дерево, направленный граф, ненаправленный граф, описание объектов и схема-звезда. Понимая эти принципы, эксперты предметной области могут воспроизвести основные элементы логики деятельности, запечатленные в модели данных. На этой стадии выполняется процедура реверс-инжиниринга процессов: извлечение из модели данных логики деятельности и формулирование набора терминов, на основе которых будет создаваться общая онтологическая модель.

Один из шагов работы с моделью данных источника является анализ схемы на предмет отражения отдельных категорий данных, перспектив. Для этого сущности схемы данных нужно разделить по перспективам. Возможные перспективы, соотносящиеся со столбцами матрицы Дж. Замана, представлены в [12]. Это:

- базовые информационные объекты предметной области;
- транзакции;
- роли участников;
- пространственное описание;
- временные характеристики;
- мотивация деятельности.

Для построения онтологической схемы из обозначенных перспектив наиболее важными являются базовые информационные объекты и транзакции. Роли представляют собой трансформацию организационной структуры и предназначены для предоставления полномочий по работе с данными в локальных системах. Пространственно-временной экстенс представляется либо в форме нормативно-справочной информации, например, в форме адресной системы, либо в виде

характеристик транзакций. Мотив представляет собой результат связи целей организации с обработкой данных операционной деятельности.

Для построения графического отражения онтологии на первом этапе выделяются первичные базовые сущности – сущности, отражающие основные объекты предметной области. Первичные объекты, участвуя в транзакциях, формируют производные базовые объекты.

Модели данных хранилища для отдельных предметных областей представлены в [13] – образование, [14] – фондовый рынок, [15] – нефтедобыча. В данном случае рассмотрим совокупность этапов по составлению графического представления онтологии для хранилища данных и созданию модели данных на условном примере. Задача предполагает создание модели данных хранилища, в котором будут собираться данные из трех информационных систем, используемых в различных видах деятельности: управление производством автомобилей, краткосрочная аренда автомобилей (carsharing) и управление отелем.

Рассмотрим последовательно отдельные шаги.

**Анализ модели данных источника.** Рассмотрим состав данных по источникам. Первый источник данных содержит информацию о производстве автомобилей. В модели (рис. 3) отражена иерархическая классификация комплектующих, отдельные из которых изготавливаются на собственном производстве, а остальные заказываются у поставщиков. Производственная часть включает иерархический процесс производства, для стадий производства указываются потребности в комплектующих. Отдельной сущностью представлен график производства. В производственном процессе особое место отводится последовательности выполнения операций, поэтому в качестве отдельной сущности представлен календарь.

Второй источник данных предполагает описание процесса сдачи в краткосрочную аренду автомобилей. Сущность «ВладелецКлиенты» содержит контракты клиентов с владельцами автомобилей. В схеме приведена категоризация автомобилей. Основная сущность «Аренда» отражает состояние автомобиля в текущий момент времени. На основе данных этой сущности клиенты могут видеть статус, местоположение автомобилей, заказать аренду (рис. 3).

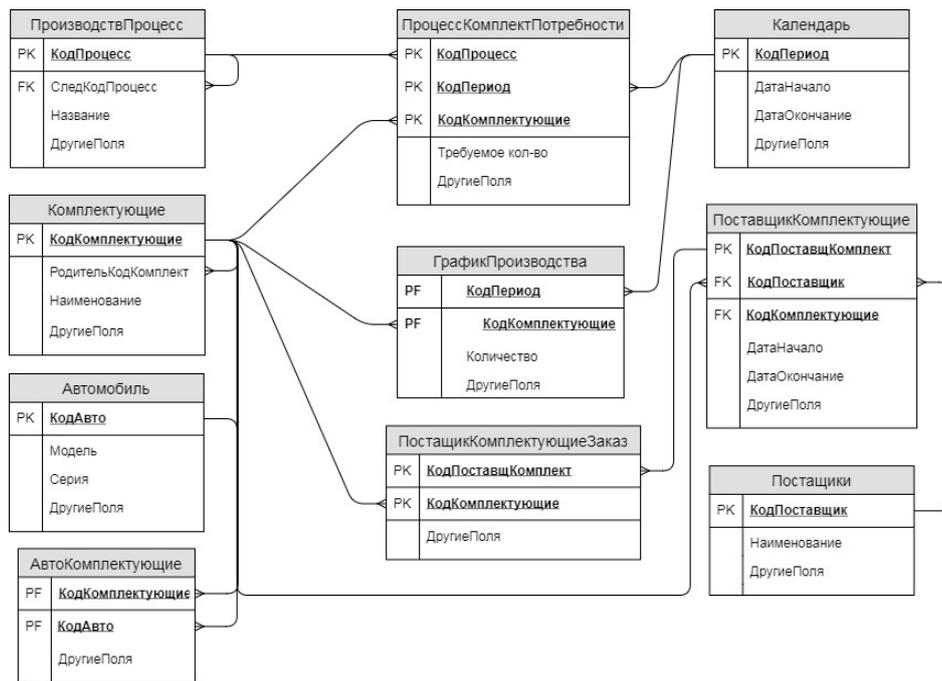


Рис. 3. Фрагмент модели данных системы управления производством автомобилей  
 Fig. 3. Fragment of the car production management data model

Последний, третий, источник данных описывает деятельность гостиницы: гость бронирует гостиницу, заселяется. В гостинице представлен ряд дополнительных услуг, которые оказывают сторонние компании (рис. 4).

Проанализируем модели данных каждого источника на предмет выделения базовых информационных объектов и транзакций с ними. В модели данных первой системы можно выделить следующие объекты, претендующие на отдельные базовые объекты:

- автомобиль,
- комплектующие,
- поставщик как юридическое лицо,
- календарь,
- этап производства.

В модели данных второй системы были определены следующие базовые объекты:

- автомобиль,
- клиент как физическое лицо,
- владелец как юридическое лицо,
- категория,
- характеристика,
- статус,
- местоположение.

В модели данных третьей системы выделены следующие базовые объекты:

- гость как физическое лицо,
- компания как юридическое лицо,
- услуги как перечень сервисов,
- иерархический справочник номеров гостиницы с отражением категории номеров.

Транзакции, в которых задействованы базовые объекты, выделим на этапе графического представления онтологии.

**Графическое представление онтологии источников данных.** Существует достаточно большое количество исследований, посвященных выделению набора терминов предметной области и созданию модели данных хранилища. Основная часть их посвящена теме разработки технологического инструментария для создания модели данных хранилища в автоматическом режиме на основе формального анализа модели данных источника [16–18]. В отдельных исследованиях состав терминов, включаемых в онтологию, определяется автоматически с использованием математического аппарата [19, 20]. Другая часть исследований посвящена определению состава атрибутов, описывающих сущности концептуальной модели и представленных в нескольких источниках в виде совокупности атрибутов одного класса [21].

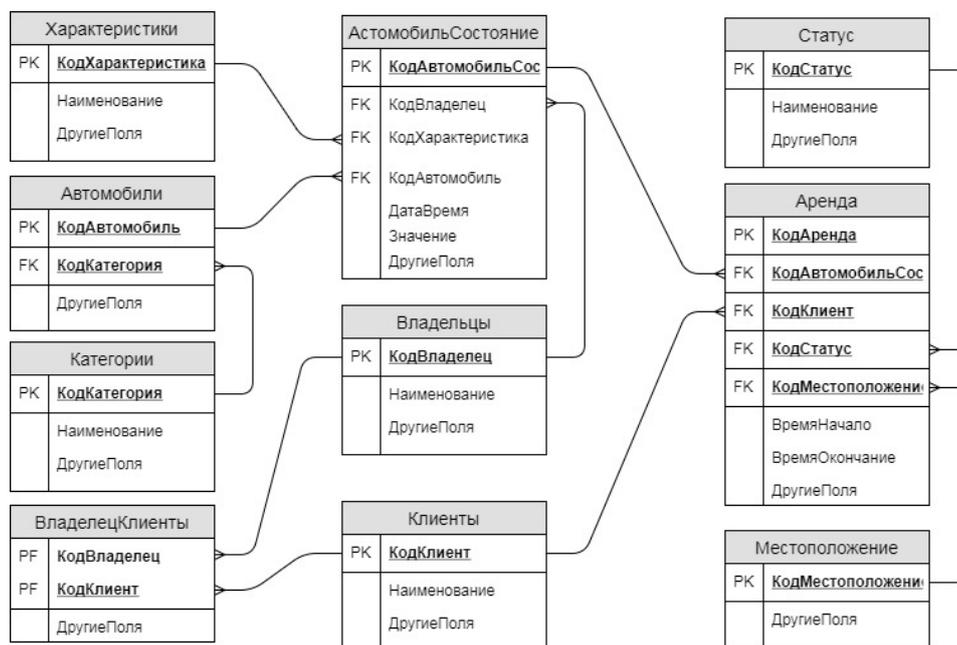


Рис. 4. Фрагмент модели данных системы управления краткосрочной аренды автомобилей

Fig. 4. A fragment of the data model of the car-sharing management system

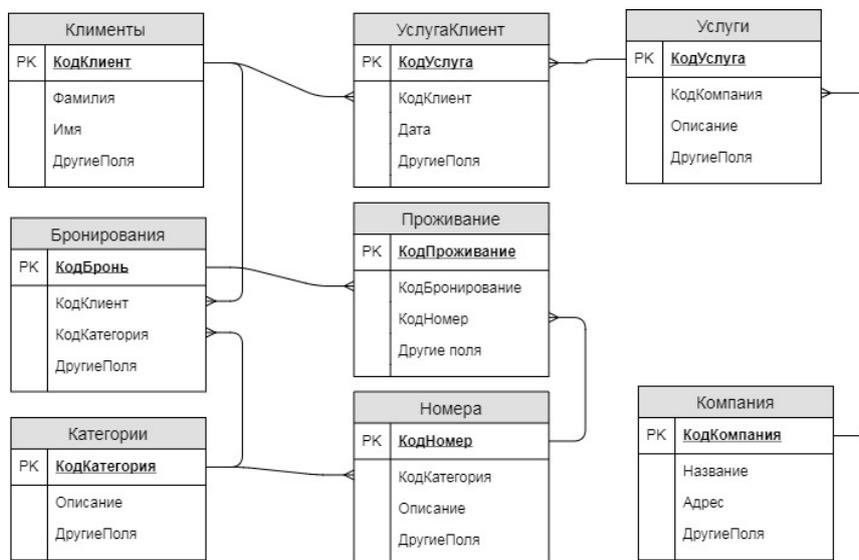


Рис. 5. Фрагмент модели данных системы управления гостиницей

Fig. 5. Fragment of the hotel management system data model

Основная часть бизнес-логики функциональности, для которой создается информационная система, воплощается в модели данных. При отражении деятельности реального мира в информационной модели происходят существенные изменения, которые привносятся

предпочтениями разработчиков, а также особенностями решаемых задач. Поэтому ориентация на совокупность сущностей, представленных в информационной системе, без глубокого содержательного анализа, не вполне обоснованна, а их автоматическая обработка может привести к се-

резному искажению картины реального мира, что повлечет ошибки логического и технического характера.

В проектах по созданию приложений, которые мы используем как источники данных, слой информации и данных является вспомогательным, поскольку основная задача таких проектов – создать приложение, которое будет поддерживать определенную функциональность, без акцента на поставку данных для внешних пользователей. В случае информационного обмена между системами слой данных является принципиальным поскольку взаимодействия происходит как раз посредством обмена информацией и данных, и задача слоя данных в случае проектов выходит на первый план.

В рассматриваемом варианте создания модели данных хранилища основной акцент сделан на участие экспертов предметной области в анализе моделей данных источников данных, создание онтологий источников и объединенной онтологии, на основе которой создается модель данных хранилища. Такой подход позволит воспроизвести модель деятельности источников данных и создать объединенную содержательную модель, которая будет основана не на технологических особенностях, представ-

ленных в источниках данных, а на управленческом, экономическом понимании выполняемой деятельности объединенной организации путем создания ее онтологии.

Графическое представление онтологии отражает логику деятельности в рамках предметной области: первичные базовые сущности в ходе деятельности объединяются и получают производные базовые сущности, такие как «Автомобиль», «Категория», «Характеристика» по второму проекту (рис. 6), производство автомобилей по первому проекту, проживание в гостинице и получение услуг по третьему проекту.

Состав сущностей моделей данных локальных систем не всегда содержит первичные сущности в явном виде. Первичные сущности таких перспектив, как «Время», «Местоположение», «Мотив», часто представлены в форме атрибутов других первичных сущностей. Эти перспективы в моделях данных хранилищ целесообразно выделять как самостоятельные первичные базовые сущности. Это позволяет не только проводить анализ данных по аналогичным объектам, но и анализировать объединенные данные по пространственно-временным срезам.

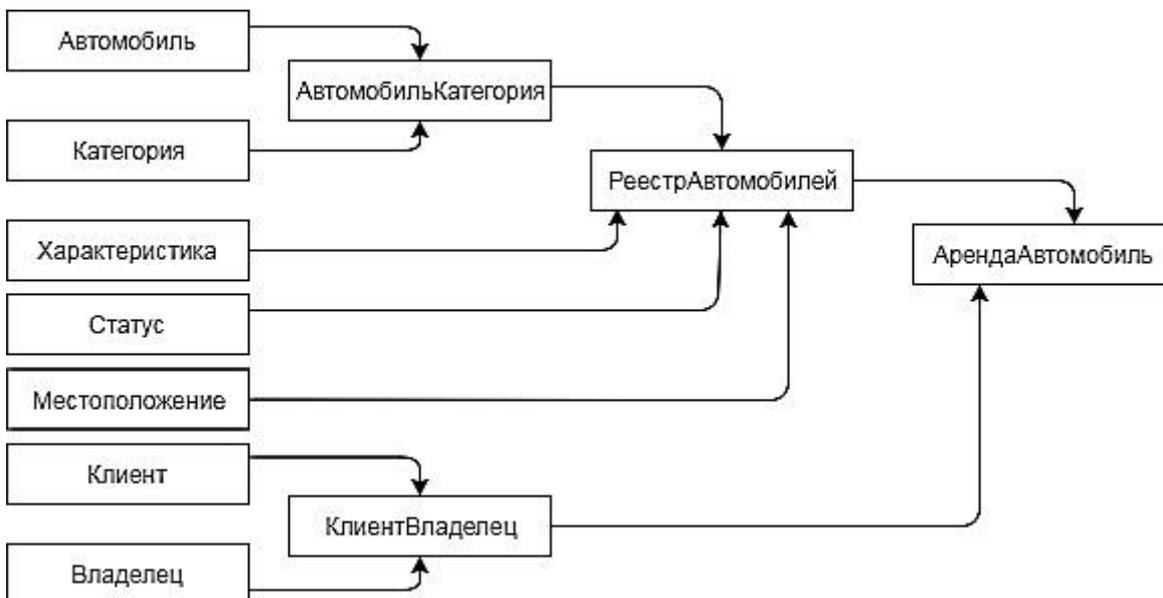


Рис. 6. Графическое представление онтологии краткосрочной аренды автомобилей  
 Fig. 6. Graphic representation of the ontology of Carsharing

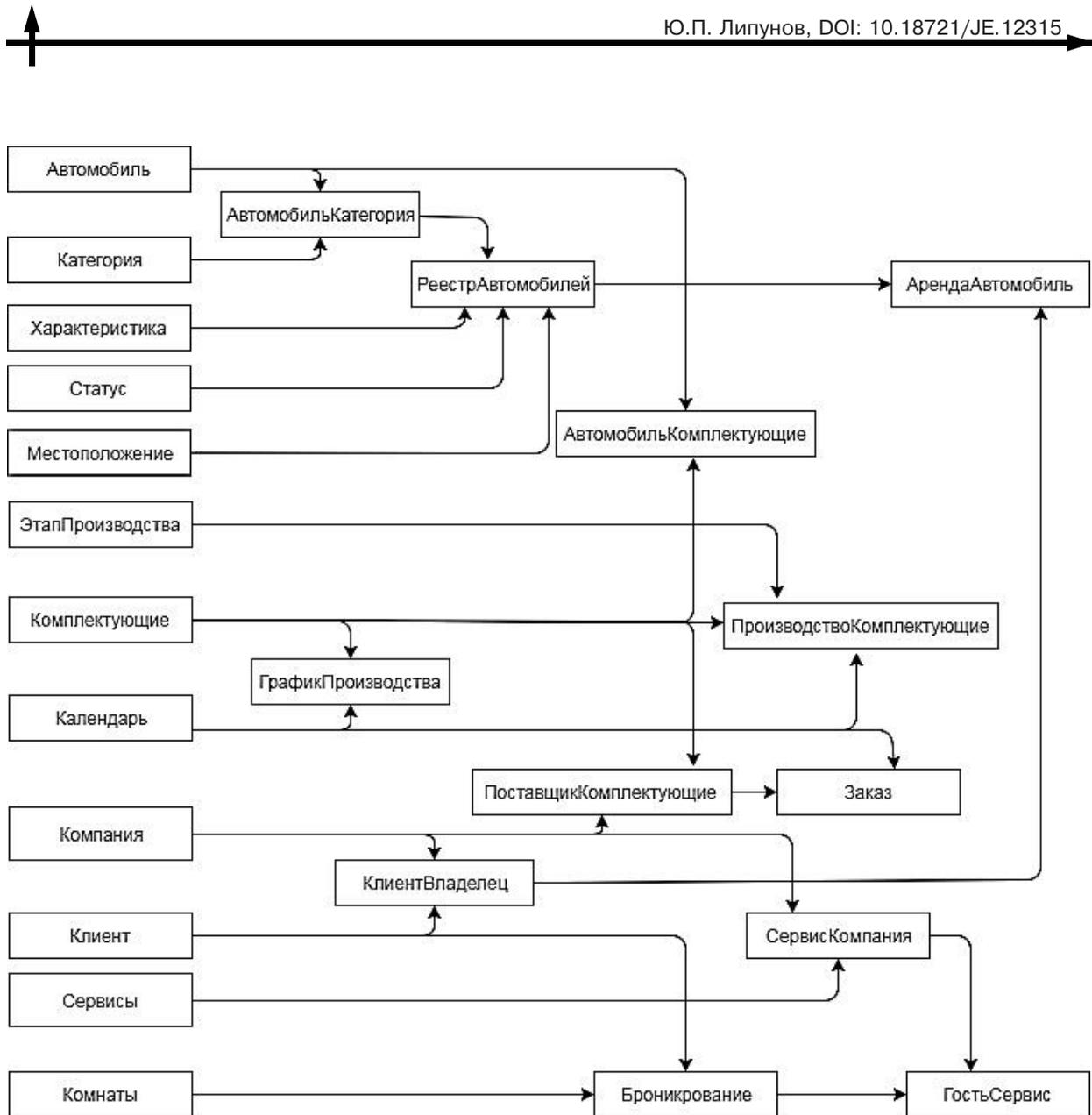


Рис. 7. Графическое представление объединенной онтологии

Fig. 7. Graphic representation of a unified ontology

**Создание единого представления для хранилища данных.** Для создания объединенной онтологии на основе разрозненных графических представлений необходимо воспроизвести единое представление. Поскольку рассматриваемый пример предполагает объединение разных предметных областей, то в данном случае будут пересекаться только базовые сущности, а все производные сущности будут различными, поскольку различен характер транзакций. На этапе анализа выявлено, что в проектах представлены юридические и физические лица, и в объединенном представлении возможно их унифицированное отражение. Помимо этого для де-

монстрации были объединены автомобили. В результате наложения локальных схем друг на друга получилось следующее представление (см. рис. 6).

Имея объединенное представление предметной области, соединяющей три различных источника данных, можно перейти к созданию модели данных хранилища.

#### Построение логической модели данных хранилища на основе онтологии

В логической модели хранилища можно выделить две категории компонент, информация о которых загружается в хранилище: первичные

базовые компоненты и производные базовые компоненты. Логическую модель хранилища будем создавать по реляционному принципу с использованием трех категорий сущностей: «Базовые сущности», «Сущности пересечений», «Сущности хранения». Принцип построения модели интеграции данных с использованием трех категорий сущностей представлен в работе W.H. Inmon и D. Linstedt [22]. Состав полей для базового варианта модели Data Vault представлен в [23]. Шаблон модели данных хранилища представлен на рис. 8.

Путем использования трех категорий сущностей создается модель, позволяющая корректно собирать атомарные данные из внешних систем, а также отслеживать историю их изменений.

Физическая модель данных хранилища реализуется в виде трех категорий таблиц: *H*-таблиц, *L*-таблиц и *S*-таблиц. *H*-таблицы предназначены для хранения кодов объектов, *L*-таблицы отражают связи между базовыми объектами, *S*-таблицы являются местом хранения данных систем-источников (рис. 9). На рисунке отражены поля, участвующие в формировании первичных и внешних ключей, и не представлены атрибуты описания объектов.

Модель данных хранилища предполагает сохранение данных источников и отражение данных по аналогичным объектам в сопоставимом

виде. Например, если клиент компании краткосрочной аренды воспользовался сервисами отеля, то хранилище данных позволит реализовать просмотр всех действий с участием этого физического лица. Имея единый код, физическое лицо может иметь разное описание, загруженное из разных систем. Данные о физлицах из двух источников сохраняются в двух разных таблицах: *S\_Person\_SHAR*, *S\_Person\_HOT*. При этом физическое лицо может участвовать в разных категориях действий: бронировать автомобиль, отель и т. д. Код физического лица будет являться составной частью суррогатного кода сущностей пересечений, описывающих транзакции. Это позволит получить прозрачную картину об участии отдельных объектов в различных действиях. Аналогичный пример можно привести с юридическими лицами, например, компания, осуществляющая услуги по краткосрочной аренде, может выполнять услуги по перевозке проживающих в гостинице туристов.

Созданная модель данных хранилища предоставляет возможность по обмену информацией между системами, например, для поставщика комплектующих интересен парк арендуемых автомобилей, которым необходимы обслуживание и ремонт. Такая организация данных в хранилище позволяет решать разного рода аналитические задачи, в случае если произошло объединение этих предприятий.

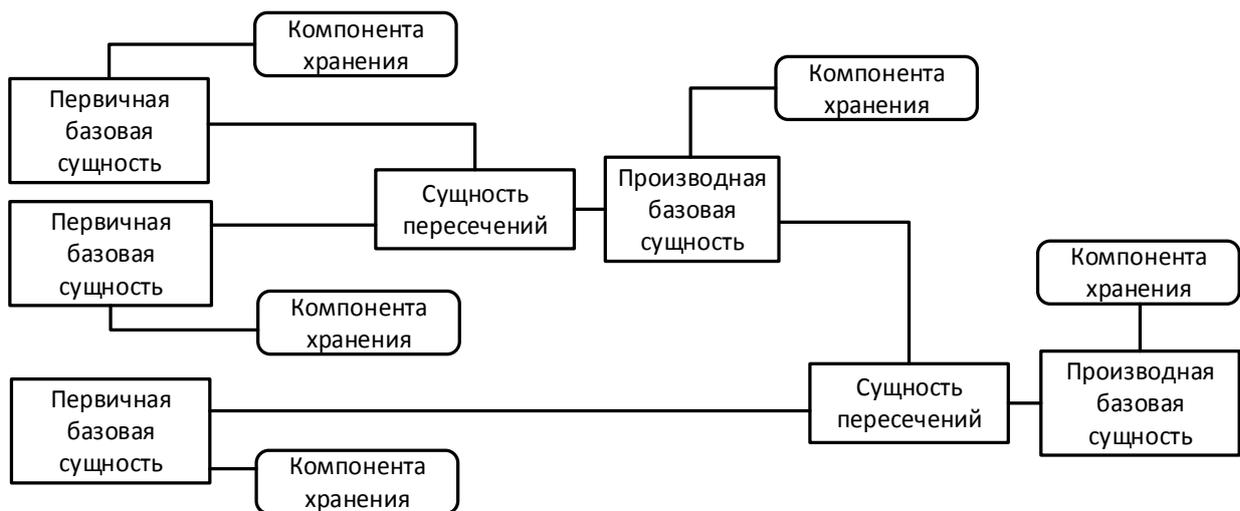


Рис. 8. Совокупность сущностей и компонент модели данных хранилища и их связи  
 Fig. 8. The set of entities and components of the data warehouse model and relationships

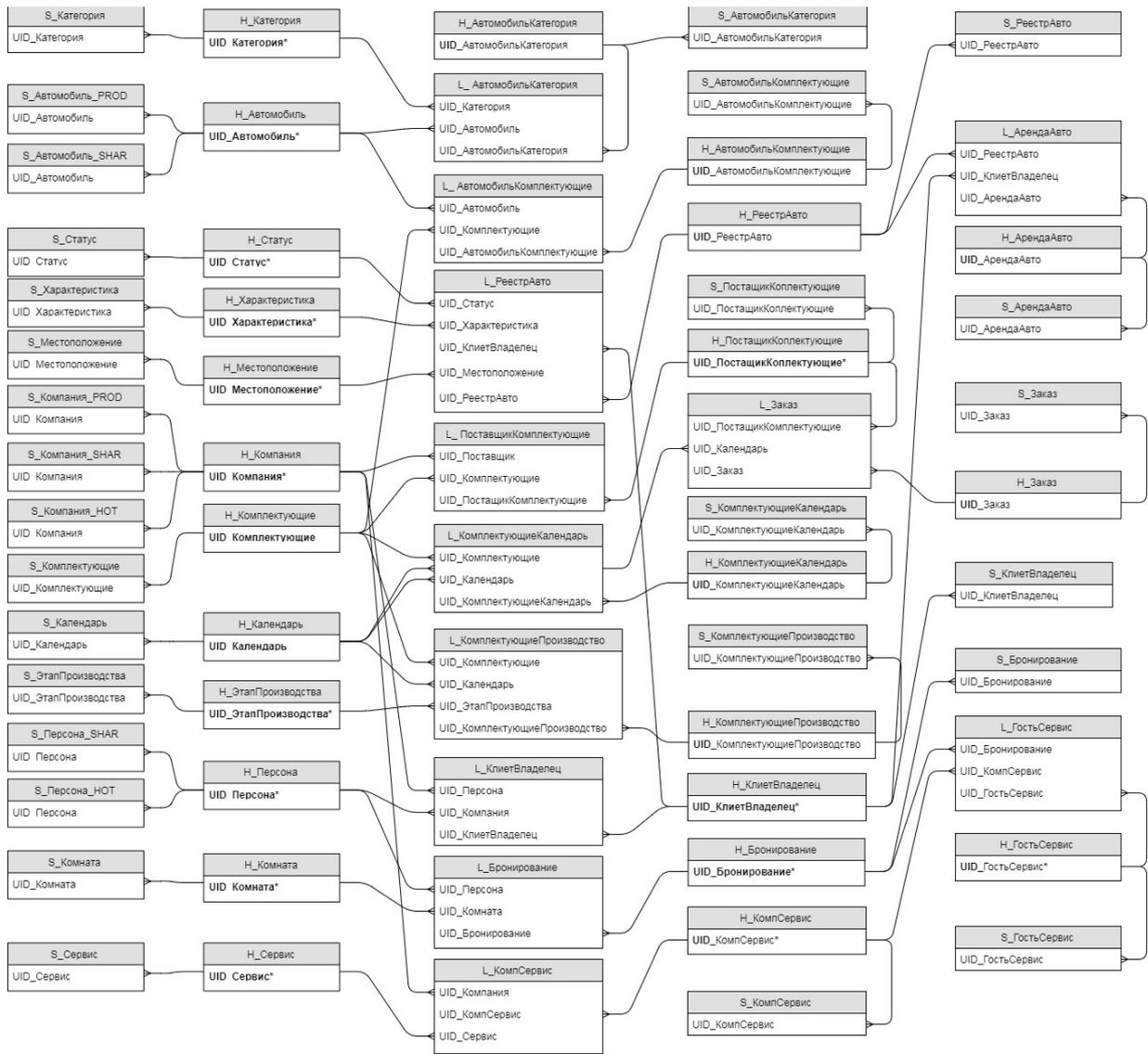


Рис. 9. Модель данных хранилища (отражены поля, участвующие в формировании ключей)  
 Fig. 9. Warehouse data model (keys fields)

В данном случае представлена модель данных, которая реализуется в СУБД, и поддержание ее в актуальном состоянии путем выполнения процедур загрузки, трансформации может выполняться средствами СУБД либо внешним приложением, в котором будет реализована логика преобразования данных. В этом случае технологиями реализации могут выступать объектно-ориентированные программные среды, в том числе те, которые ориентированы поддержку онтологических методов [24].

*Результаты исследования.* Представленная методика создания онтологии для хранилища

данных включает следующие шаги: построение концептуальной модели предметной области, создание онтологического представления для источников данных, создание объединенного онтологического представления, построение модели данных хранилища на основе онтологии. Создание онтологии является одним из основных шагов для создания формата данных, на основе которого организуется информационное взаимодействие между системами. В области создания онтологий реализуется большое количество проектов, обзоры по онтологиям для отдельных сфер экономики приведены в [25] – сельское хозяй-

ство, [26] – железнодорожный транспорт, [27] – строительство. Несмотря на то, что сегодня создано большое количество онтологий для различных предметных областей, отдельные из которых представлены в форме международных стандартов<sup>1</sup>, для большинства сфер деятельности их приходится разрабатывать самостоятельно. Методика разработки собственных онтологий особенно актуальна для государственного сектора, который работает в рамках регламентов, и совокупность поставляемых для обмена сведений не может предопределяться внешними стандартами.

*Выводы.* Данные становятся существенным активом многих организаций. В организационных структурах государственных корпораций и компаний с государственным участием вводится позиция «Руководитель по цифровой трансформации» (Chief Data Officer, CDO)<sup>2</sup>, правительственные и межправительственные документы предполагают развитие работы с данными<sup>3</sup>. Что касается данных, большинство

<sup>1</sup> ISO 12006-3:2007. ISO 12006-3:2007 Building construction – Organization of information about construction works – Part 3: Framework for object-oriented information. ISO/TC 59/SC 13 Organization of information about construction works, 2007. URL: <https://www.iso.org/standard/38706.html> (дата обращения: 12.12.2017); ISO 16739: 2013. ISO 16739:2013 Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries. ISO/TC 184/SC 4 Industrial data, 2013. URL: <https://www.iso.org/standard/51622.html> (дата обращения: 12.12.2017); ISO. Building information models -- Information delivery manual. ISO/TC 59/SC 13 Organization of information about construction works, 2016. URL: <https://www.iso.org/standard/60553.html> (дата обращения: 12.12.2017).

<sup>2</sup> Методические рекомендации по разработке стратегий цифровой трансформации государственных корпораций и компаний с государственным участием. 17.05.2019 / Минкомсвязь России [Интернете].

<sup>3</sup> Положение о модели данных Евразийского экономического союза / Евразийская экономическая комиссия. М., 2017; О первоочередных мерах, направленных на создание государственной информационной системы «Единая информационная среда в сфере систематизации и кодирования информации» / Правительство РФ. М., 2016; Проект Федерального закона «О систематизации и гармонизации информации в Российской Федерации». 02.07.2018 / Минфин РФ. URL: <https://www.consultant.ru/law/hotdocs/54165.html/>

организаций корпоративного и государственного сектора основное внимание уделяют их анализу, поиску решений для оперативной или управленческой деятельности. Анализ данных строится на информационной базе, и качество анализа предопределяется возможностями информационной модели. Здесь был представлен подход к созданию собственного решения по интеграции данных на основе создания онтологии. Такой подход позволяет минимизировать отклонение технологической точки зрения, используемой при создании приложений, от содержательного, управленческого взгляда на деятельность. Близость двух точек зрения позволит сократить число логических и технических ошибок, сделает систему более адаптируемой к происходящим изменениям.

На основе изложенных подходов можно разработать среду, которая позволит описывать структуру данных источников, моделировать объединенную онтологию и создавать модель данных хранилища. Для реализации такого подхода необходимо представить модель данных хранилища в виде классов с описанием обязательных атрибутов и программным выполнением таких операций, как хеш преобразование кодов объекта и значений набора атрибутов, описывающих объекты. Такие инструменты созданы для представления данных источника в виде стандартного формата данных<sup>4</sup>, но не позволяют решить поставленную здесь задачу.

Среда взаимодействия разнородных систем предполагает создание большого количества компонент, отдельные из которых можно стандартизировать и осуществить типовое решение. Примером такого решения является компонента проект OpenXDX<sup>5</sup>, позволяющая поставлять данные, отвечающие стандарту обмена для модели обмена данными NIEM.

<sup>4</sup> Content Assembly Mechanism. 02.2016 / OASIS. URL: [https://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=cam](https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=cam) (дата обращения: 05.2019).

<sup>5</sup> VerifyXML.org. CAM XML validation. URL: <http://www.verifyxml.org/OpenXDX-page.html> (дата обращения: 23.12.2017).

## СПИСОК ЛИТЕРАТУРЫ

- [1] IPA, «Infrastructure for Multilayer Interoperability», IPA, 2015. URL: <http://imi.ipa.go.jp/IMIOverview-En.html> (дата обращения: 05.05.2017).
- [2] ISA2, «ADMS». ISA, 2016. URL: <https://joinup.ec.europa.eu/asset/adms/home> (дата обращения: 15.04.2017).
- [3] Hughes R. Traditional Data Modeling Paradigms and Their Discontents // Agile Data Warehousing for the Enterprise. 2016. P. 249–291.
- [4] OMG, Semantic Information Modeling for Federation (SIMF RFP). 2011. URL: <http://tinyurl.com/SIM-Frpf> (дата обращения: 12.08.2014).
- [5] Olive A. Conceptual Modeling of Information Systems, Berlin Heidelberg. Springer-Verlag, 2007.
- [6] Wieringa R. Real-World Semantics of Conceptual Models, The Evolution of Conceptual Modeling, Berlin Heidelberg: Springer-Verlag, 2011.
- [7] Кольчев П.М. Релятивная онтология. СПб.: СПбГУ, 2006.
- [8] Лапшин В.А. Онтологии в компьютерных системах. М.: Научный мир, 2010.
- [9] Benker T. A Generic Process Data Warehouse Schema for BPMN Workflows // Business Information Systems. BIS 2016. Lecture Notes in Business Information Processing, Cham, 2016.
- [10] Wu N., Simpson A. Formal relational database design: an exercise in extending the formal template language // Form Asp Comp. 2014. Vol. 1231, no. 26.
- [11] Blaha M. DataModelPatterns.
- [12] Hay D. Data Model Patterns: A Metadata Map, Morgan Kaufmann, 2006.
- [13] Липунов Ю.П. Принципы создания онтологии для хранилища данных. Пример: высшее образование // Бизнес. Образование. Право. Вестник Волгоградского института бизнеса. 2016. Т. 36. С. 72–79.
- [14] Lipuntsov Y., Beatch R., Collier I. Financial Markets Data Collection Using the Information Model of Interagency Cooperation and the International System of Codification of Financial Instruments // Communications in Computer and Information Science. 2017. Vol. 745.
- [15] Kharlamov E., Hovland D. Ontology Based Data Access in Statoil // Journal of Web Semantics. 2017. Vol. 44. P. 3–36.
- [16] Pourtalebi S., Horváth I. Information schema constructs for defining warehouse databases of genotypes and phenotypes of system manifestation features // Frontiers of Information Technology & Electronic Engineering. 2017. Vol. 17, no. 9. P. 862–884.
- [17] Boyd M., McBrien P., Tong N. The AutoMed Schema Integration Repository // Advances in Databases. 2002. Vol. 2405.
- [18] Kang I., Na S., Lee J. Conceptual Schema Approach to Natural Language Database Access // Proceedings of the Australasian Language Technology Workshop 2003, Melbourne, 2003.
- [19] Schmitt I., Saake G. A comprehensive database schema integration method based on the theory of formal concepts // Acta Informatica. 2005. Vol. 41, no. 7/8. P. 475–524.
- [20] Preis M., Seitz J. A Hybrid Approach of Data Warehouse Integration Based on New Storage Technologies // International Journal of Advances in Computing and Management. 2012. Vol. 1, no. 1. P. 40–46.
- [21] Szymczak M., Bronselaer A., Zdrożny S., De Tré G. Content Data Based Schema Matching // Challenging Problems and Solutions in Intelligent Systems, 2016.
- [22] Inmon W., Linstedt D. Data Architecture: A Primer for the Data Scientist. Morgan Kaufmann, 2014.
- [23] Linstedt D. DV Modeling Specification. URL: <http://danlinstedt.com/allposts/datavaultcat/standards/dv-modeling-specification-v1-0-8/>.
- [24] Bellatreche L., Ameer A., Papadopoulos G. Models and data engineering // Future Generation Computer Systems. 2017. Vol. 70. P. 1–15.
- [25] Куприяновский В., Липунов Ю., Намиот Д., Гринько О. Агрокультура 4.0: синергия системы – систем, онтологии, интернета вещей и космических технологий // International Journal of Open Information Technologies. 2018. Vol. 6, no. 10. P. 46–67.
- [26] Куприяновский В., Покусаев О., Волокитин Ю., Намиот Д., Петрунина И., Зажигалкин А. Формализованные онтологии и сервисы для высокоскоростных магистралей и цифровой железной дороги // International Journal of Open Information Technologies. 2018. Vol. 6, no. 6. P. 69–86.
- [27] Куприяновский В., Сиягов С., Липунов Ю. ВМ и инженерные формализованные онтологии на цифровой железной дороге Европы в объединении EULYNX – экономика данных // International Journal of Open Information Technologies. 2018. Vol. 6, no. 8.

## REFERENCES

- [1] IPA, «Infrastructure for Multilayer Interoperability,» IPA, 2015. Available: <http://imi.ipa.go.jp/IMIOverview-En.html>. [Data obrashcheniya: 05 05 2017].
- [2] ISA2, «ADMS». ISA, 2016. URL: <https://joinup.ec.europa.eu/asset/adms/home> (accessed April 15, 2017).
- [3] **R. Hughes**, Traditional Data Modeling Paradigms and Their Discontents, *Agile Data Warehousing for the Enterprise*, (2016) 249–291.
- [4] OMG, Semantic Information Modeling for Federation (SIMF RFP), 2011. URL: <http://tinyurl.com/SIMFrfp> (accessed August 12, 2014).
- [5] **A. Olive**, Conceptual Modeling of Information Systems. Berlin Heidelberg: Springer-Verlag, 2007.
- [6] **R. Wieringa**, Real-World Semantics of Conceptual Models, *The Evolution of Conceptual Modeling*, Berlin Heidelberg: Springer-Verlag, 2011.
- [7] **P.M. Kolychev**, Relyativnaya Ontologiya. SPb.: SPbGU, 2006.
- [8] **V.A. Lapshin**, Ontologii v kompyuternykh sistemakh. M.: Nauchnyy mir, 2010.
- [9] **T. Benker**, A Generic Process Data Warehouse Schema for BPMN Workflows, *Business Information Systems. BIS 2016. Lecture Notes in Business Information Processing*, Cham, 2016.
- [10] **N. Wu. A. Simpson**, Formal relational database design: an exercise in extending the formal template language, *Form Asp Comp*, 1231 (26) 2014.
- [11] **M. Blaha**, DataModelPatterns.
- [12] **D. Hay**, Data Model Patterns: A Metadata Map, Morgan Kaufmann, 2006.
- [13] **Yu.P. Lipuntsov**, Printsipy sozdaniya ontologii dlya khranilishcha dannykh. Primer: vyssheye obrazovaniye, *Biznes. Obrazovaniye. Pravo. Vestnik Volgogradskogo instituta biznesa*, 36 (2016) 72–79.
- [14] **Y. Lipuntsov, R. Beatch, I. Collier**, Financial Markets Data Collection Using the Information Model of Interagency Cooperation and the International System of Codification of Financial Instruments, *Communications in Computer and Information Science*, 745 (2017).
- [15] **E. Kharlamov, D. Hovland**, Ontology Based Data Access in Statoil, *Journal of Web Semantics*, 44 (2017) 3–36.
- [16] **S. Pourtalebi, I. Horváth**, Information schema constructs for defining warehouse databases of genotypes and phenotypes of system manifestation features, *Frontiers of Information Technology & Electronic Engineering*, 17 (9) (2017) 862–884.
- [17] **M. Boyd, P. McBrien, N. Tong**, The AutoMed Schema Integration Repository» *Advances in Databases*, 2405 (2002).
- [18] **I. Kang, S. Na, J. Lee**, Conceptual Schema Approach to Natural Language Database Access, *Proceedings of the Australasian Language Technology Workshop 2003, Melbourne, 2003*.
- [19] **I. Schmitt, G. Saake**, A comprehensive database schema integration method based on the theory of formal concepts, *Acta Informatica*, 41 (7/8) (2005) 475–524.
- [20] **M. Preis, J. Seitz**, A Hybrid Approach of Data Warehouse Integration Based on New Storage Technologies, *International Journal of Advances in Computing and Management*, 1 (1) (2012) 40–46.
- [21] **M. Szymczak, A. Bronselaer, S. Zadrozny, G. De Tré**, «Content Data Based Schema Matching,» *v Challenging Problems and Solutions in Intelligent Systems*, 2016.
- [22] **W. Inmon, D. Linstedt**, *Data Architecture: A Primer for the Data Scientist*, Morgan Kaufmann, 2014.
- [23] **D. Linstedt**, DV Modeling Specification. URL: <http://danlinstedt.com/allposts/datavaultcat/standards/dv-modeling-specification-v1-0-8/>
- [24] **L. Bellatreche, A. Ameer, G. Papadopoulos**, Models and data engineering, *Future Generation Computer Systems*, 70 (2017) 1–15.
- [25] **V. Kupriyanovskiy, Yu. Lipuntsov, D. Namiot, O. Grinko**, Agrokultura 4.0: sinerhiya sistemy – sistem, ontologii, interneta veshchey i kosmicheskikh tekhnologiy, *International Journal of Open Information Technologies*, 6 (10) (2018) 46–67.
- [26] **V. Kupriyanovskiy, O. Pokusayev, Yu. Volokitin, D. Namiot, I. Petrunina, A. Zazhigalkin**, Formalizovannyye ontologii i servisy dlya vysokoskorostnykh magistralei i tsifrovoy zheleznoy dorogi, *International Journal of Open Information Technologies*, 6 (6) (2018) 69–86.
- [27] **V. Kupriyanovskiy, S. Sinyagov, Yu. Lipuntsov**, BIM i inzhenernyye formalizovannyye ontologii na tsifrovoy zheleznoy doroge Yevropy v obyedinenii EULYNX – ekonomika dannykh, *International Journal of Open Information Technologies*, 6 (8) 2018.

LIPUNTSOV Yuri Y. E-mail: lipuntsov@econ.msu.ru