

УДК 004.65

Л. В. РУДИКОВА Е. В. ЖАВНЕРКО

## О МОДЕЛИРОВАНИИ ДАННЫХ ПРЕДМЕТНЫХ-ОБЛАСТЕЙ ПРАКТИКО-ОРИЕНТИРОВАННОЙ НАПРАВЛЕННОСТИ ДЛЯ УНИВЕРСАЛЬНОЙ СИСТЕМЫ СКЛАДИРОВАНИЯ И ОБРАБОТКИ ДАННЫХ

*«Гродненский государственный университет имени Янки Купалы»*

*В статье рассматривается представления данных для предметных областей практико-ориентированной направленности, которые являются основой общей модели данных для создания единого хранилища данных. Приводится краткая характеристика предметных областей, связанных с различными направлениями деятельности людей на текущий момент времени. Предложены соответствующие модели данных, а также рассмотрена их взаимосвязь между собой в аспекте обработки данных и построения единой универсальной системы, базирующейся на информационной технологии складирования данных и которой характерны расширяемая комплексная предметная область, интегрированность данных, поступающих из различных источников, инвариантность хранимых данных во времени с обязательными темпоральными метками, относительно высокая стабильность данных, поиск необходимых компромиссов в избыточности данных, модульность отдельных блоков системы, гибкость и расширяемость архитектуры, высокие требования к безопасности хранимых данных. Предлагается общий подход к сбору и хранению данных, а также рассматриваются соответствующие модели данных с подробным описанием, которые в дальнейшем будут интегрироваться в единую схему базы данных, на основании чего будет формироваться обобщенная схема хранилища данных типа «созвездие фактов». Для получения моделей данных применяется структурная методология и учитываются общие принципы концептуального моделирования.*

*Разработка единой комплексной системы на основе полученных моделей предметных областей, способной работать сразу со многими информативными источниками и представлять данные в удобном для пользователей виде будет востребована в дальнейшем для проведения анализа данных выбранной предметной области и установления возможных межпредметных взаимодействий.*

**Ключевые слова:** хранилище данных, предметная область, универсальная система, схема данных, структурная методология, данные практико-ориентированной направленности

### О предметной области исследования, связанной с данными практико-ориентированной направленности

В настоящее время актуальным представляется разработка общей концепции и реализация системы складирования и обработки данных, связанных с различными видами деятельности людей, которая может быть рассмотрена в аспекте создания некоторого федерального хранилища данных.

Отметим, что исследование и обобщение предметных областей различной тематики является востребованным направлением исследований. Во-первых, в мире накоплены огромные массивы информации, часто достаточно разрозненной и воспринимаемой в отрыве от

общей картины жизнедеятельности общества. Во-вторых, разработаны и разрабатываются различные методы, алгоритмы и методологии анализа данных [1–3], которые используются для конкретной выборки или совокупности исследуемых данных. В-третьих, имеется достаточное количество необходимого программного обеспечения, технологий и методов формализации массивов данных, которые позволяют структурировать данные определенной тематики. С другой стороны, пример развития Всемирной сети показывает, что в едином информационном пространстве данные, практически, взаимосвязаны, порождены различными видами деятельности общества и следует рассматривать их в контексте цельного представ-

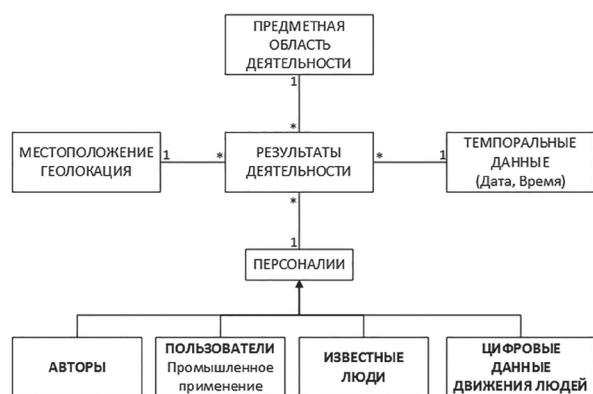


Рис. 1. Обобщенная схема предметных областей деятельности персоналий

ления, но в рамках проекций определенного вида деятельности или взаимодействий.

В рамках рассматриваемой тематики за основу хранимых и обрабатываемых данных при-

нимается информация, которая связана с конкретными людьми (персоналиями), их профессиональной (и иной) деятельностью, перемещениями в территориальном аспекте, их взаимосвязями и взаимодействиями с окружающим миром (рис. 1).

Так, под хранимыми в системе данными о персоналиях подразумевается информация следующего вида:

- информация об авторах материального наследия (например, художники, писатели, ученые и др.);
- данные об известных людях, внесших существенный вклад в экономическую (реформа экономики, сельского хозяйства, иных отраслей), социальную (положительный опыт страны в здравоохранении, соцобеспечении, образовании, в какой-либо области науки, культу-

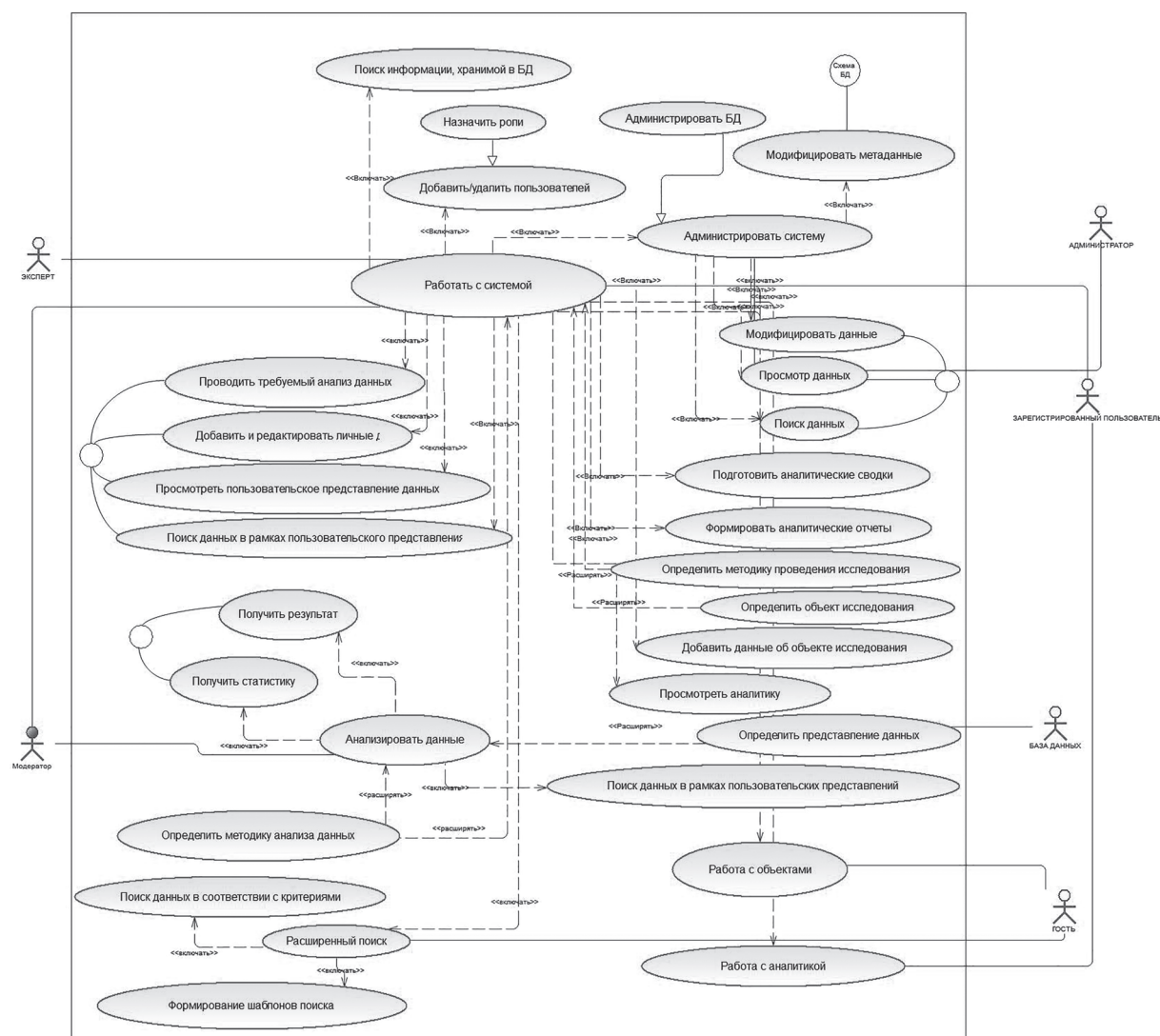


Рис. 2. Диаграмма вариантов использования для универсальной системы складирования и обработки данных практико-ориентированной направленности

ры) и военно-политическую сферы деятельности (реформа правительства, госаппарата, вооруженных сил, потенциальные или реально одержанные военные победы и т. д.);

– данные об исследователях в той или иной области деятельности (например, исследователях произведений искусств, экспертах в области промышленности, экспериментаторах в области новых технологий и т. д.);

– информация о пользователях социальных сетей и Интернет-ресурсов конкретной тематики, доступная через официальные веб-сервисы.

Отметим также, что для добавления информации в хранилище данных при необходимости требуемая информация может быть извлечена из OLTP-систем.

С учетом того, что данные о персоналиях могут дополняться постоянно, включая современный и исторический аспект, их род деятельности и результаты деятельности могут охватывать несколько смежных и несмежных направлений, а геолокация и перемещения охватывают, практически, всю территорию земного шара, обобщенную схему предметных областей можно представить в виде ключевых объектных множеств со связями «один-ко-многим» по отношению к объекту «РЕЗУЛЬТАТЫ ДЕЯТЕЛЬНОСТИ».

Предлагаемая обобщенная схема является расширяемой, легко модифицируемой и дает представление о взаимосвязях между предметными областями благодаря деятельности конкретных личностей, в частности, и расширению направлений деятельности общества, в целом.

Отметим, что предлагаемая схема интеграции данных будет являться основой для соответствующей модели данных разрабатываемой универсальной системы, базирующейся на информационной технологии складирования данных и позволяющей накапливать данные о разных отраслях деятельности общества из различных источников, проводить быстрый поиск и обработку информации, применять средства Data Mining с целью обнаружения новых закономерностей в указанных массивах хранимых данных. Для системы такого плана характерны следующие аспекты [1, 4, 6, 10–12]: расширяемая комплексная предметная область, интегрированность хранимых данных,

которые поступают из различных источников, инвариантность хранимых данных во времени с обязательными темпоральными метками, относительно высокая стабильность данных, поиск необходимых компромиссов в избыточности данных, модульность отдельных блоков системы, гибкость и расширяемость архитектуры, высокие требования к безопасности хранимых данных. Основная группа пользователей системы и общие варианты использования представлены на рис. 2.

#### **Основные модели данных для некоторых предметных областей практико-ориентированной направленности**

Приведем основные требования предметной области, которая охватывает на текущий момент времени следующие данные: информация о людях, их деятельности; данные о произведениях художественной ценности и их исследовании [4–7]; некоторые данные, связанные с демографическими и миграционными показателями, а также некоторыми данными Интернет-платформ [8, 9].

Отметим, что для приводимых ниже предметных областей разрабатываются модели данных, которые в дальнейшем будут интегрироваться в единую схему базы данных, на основании чего будет формироваться обобщенная схема хранилища данных типа «созвездие фактов». Для получения моделей данных применяется структурная методология и учитываются общие принципы концептуального моделирования. Выделяются необходимые объекты хранения данных, связи, определяются ограничения на данные, ограничения целостности, пользовательские ограничения и учитываются требования по производительности разрабатываемой системы.

Анализируя предметную область, связанную с информацией о людях, следует выделить те сведения, которые будут вызывать интерес пользователей в рассматриваемом контексте. В первую очередь к ним можно отнести персональные данные личности, такие как фамилия, имя, отчество, даты рождения и смерти, род занятий, принадлежность к определенным социальным группам по различным признакам (профессия, этническая, конфессиональная принадлежность), достижения. Био-



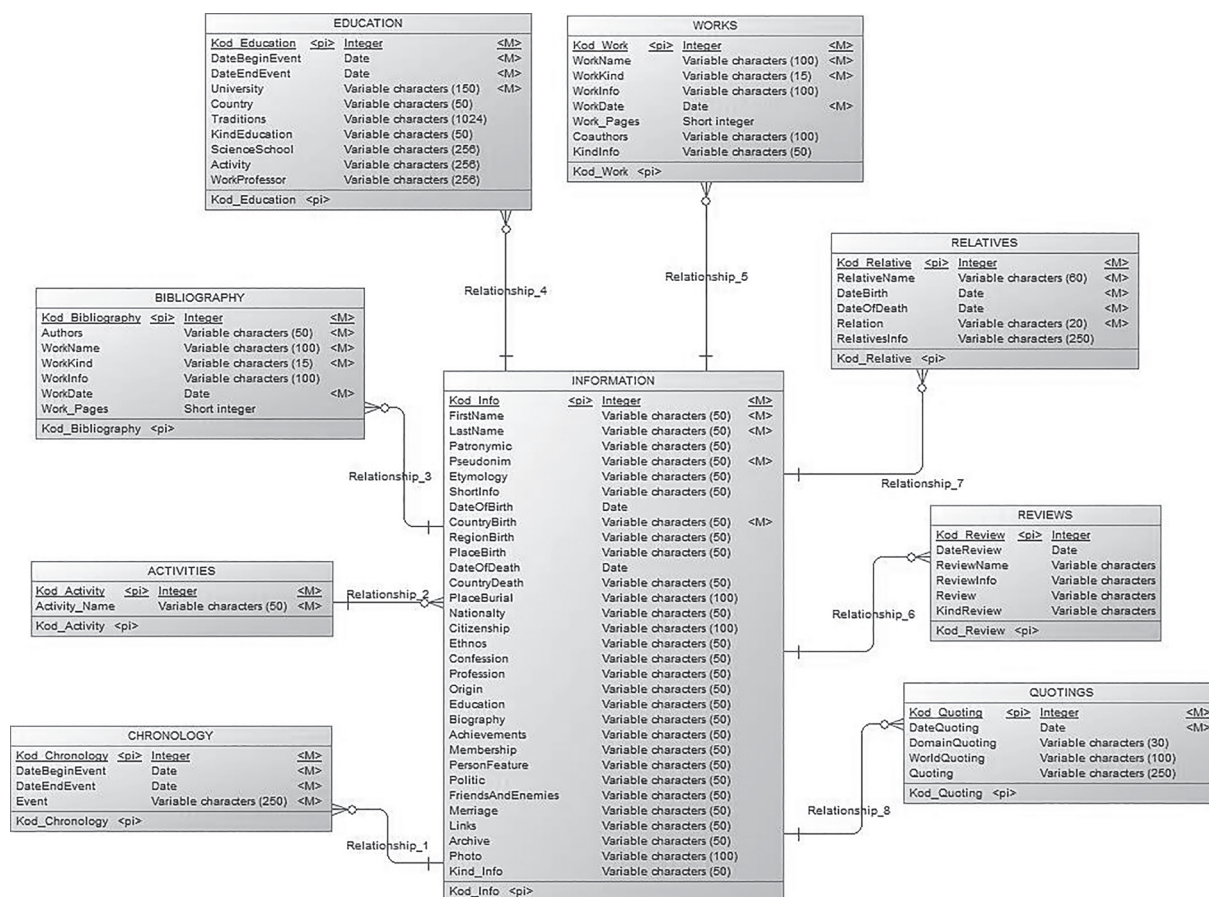


Рис. 3. Модели данных для предметной области, связанной с деятельностью известных личностей

графию человека можно рассматривать как целостную информацию, но не исключается возможность выделения из неё отдельных событий, которые упорядочены в хронологическом порядке.

Большой интерес представляют и суждения, взгляды известного человека по различным вопросам. Поэтому следует хранить информацию о работах, созданных личностью, о его высказываниях (цитатах). С другой стороны, имеет смысл хранить в базе данных и отношения других людей к данной персоне, её деятельности и вкладу в культурное (научное) наследие. Эта информация может быть представлена в виде списка книг, написанных об интересующей нас личности (т. е. библиографии), отзывов известных людей и т. д.

С учетом требований к хранимым данным, а также с возможностью расширения системы при использовании, модель данных для предметной области, связанной с деятельностью известных личностей, представлена на рис. 3. Основными объектами для хранения данных

являются следующие сущности: «Основные сведения» (BASIC) – содержит общую информацию о личности; «Хронология» (CHRONOLOGY) – фиксация событий в жизни личности, «Образование» (EDUCATION) – включает сведения о полученном образовании; «Библиография» (BIBLIOGRAPHY) – основные сведения о работах личности; «Цитаты» (QUOTES) – сущность включает информации об известных изречениях личности; «Отзывы о личности» (REVIEWS OF THE INDIVIDUAL) – отзывы друзей, современников и т. д. о личности в целом; «Работы личности» (WORK PERSONALITY) – основные сведения о работах личности, «Родственники» (RELATIVES) – информация о родственниках; «Область деятельности» (AREA OF ACTIVITY) – область (области) деятельности личности.

Для предметной области, связанной с хранением информации о произведениях художественной ценности модель данных представлена на рис. 4. С учетом модели, полученной ранее [1, 6], для нее уточнены отдельные огра-

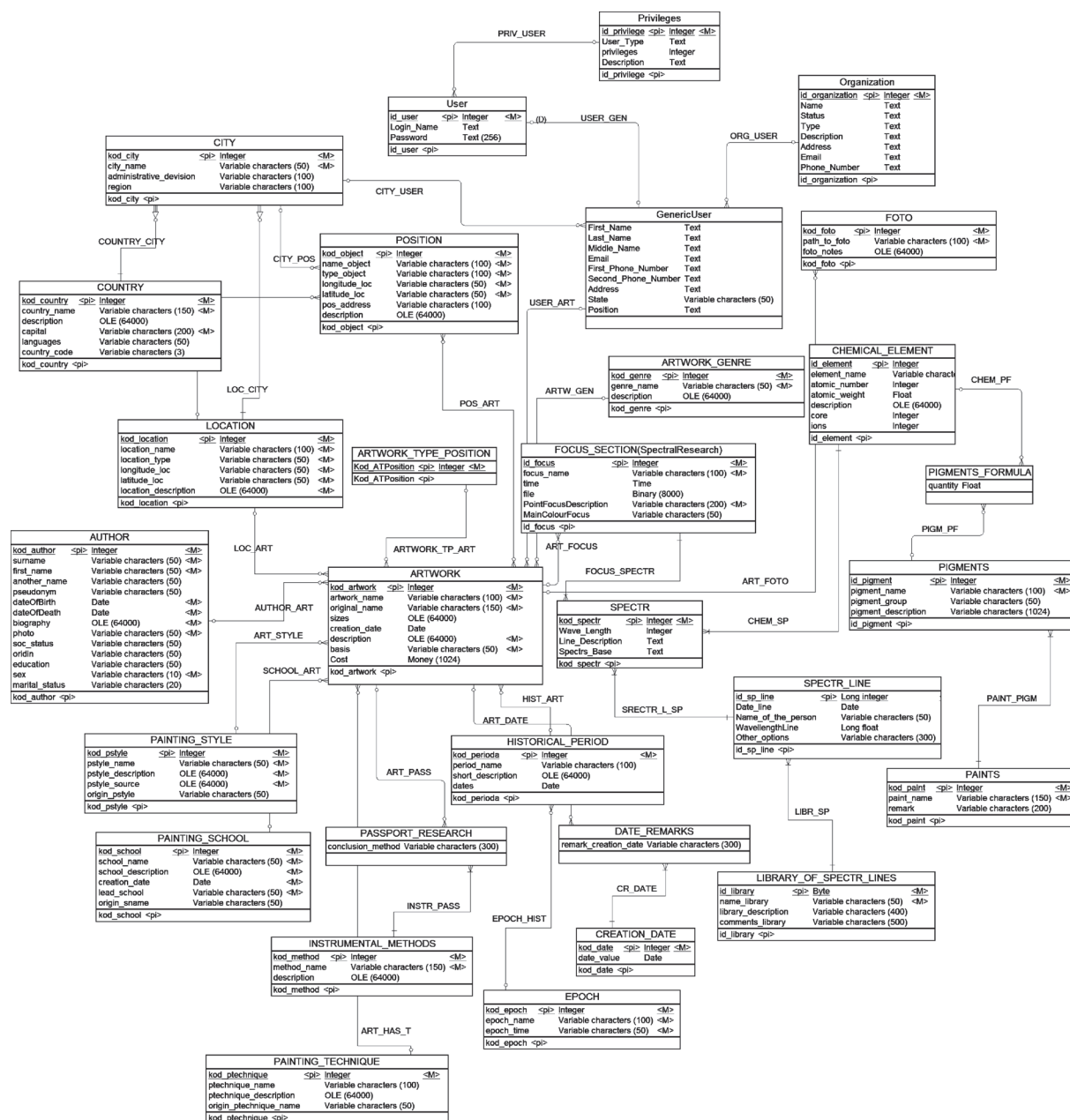


Рис. 4. Фрагмент концептуальной модели данных, связанной с хранением информации о произведениях художественной ценности

нижения предметной области, добавлены объекты (сущности) для хранения информации и соответствующие атрибуты.

Итак, основными сущностями модели данных являются следующие: «Произведение искусства» (ARTWORK) – хранит информацию обо всех художественных объектах, имеющих в системе; «Автор произведения» (AUTHOR) – информация об авторе; «Исторический период» (HISTORICAL\_PERIOD) – исторический период в который было создано произведение; «Фотографии объекта» (FOTO) – фотографии

произведения; «Краски» (PAINTS) – информация о красках (для картин); «Техника» (PAINTING\_TECHNIQUE) – техника, в которой создано произведение; «Паспорт исследования» (PASSPORT\_RESEARCH); «Страна» (COUNTRY) – страна, где находится художественный объект; «Инструментальный метод» (INSTRUMENTAL\_METHODS); «Замечания по дате» (DATE\_REMARKS) – замечания по периоду (или дате) создания произведения художественной ценности (актуально для произведений, которые создавались на протяжении опре-

деленного периода времени или же дата создания которых известна лишь приблизительно); «Дата создания» (CREATION\_DATE) – дата, которая указывается как время создания художественного объекта, однако, это сущность связана с общими датами в аспекте времени; «Спектр» (SPEKTR) – расшифровка файла спектральных линий для объекта исследований; «Спектральная линия» (SPEKTR\_LINE) – набор спектральных линий, которые входят в конкретную библиотеку спектральных линий; «Библиотека спектральных линий» (LIBRARY\_OF\_SPEKTR\_LINES) – информация о библиотеках спектральных линий; «Участки фокусировки» (FOCUS\_SECTION (SpectralResearch)) – фиксируются участки спектрального исследования художественного произведения; «Химический элемент» (CHEMICAL\_ELEMENT) – информация о химическом элементе; «Пигменты» (PIGMENTS) – информация о пигментах; «Жанр произведения искусства» (ARTWORK\_GENRE) – содержит информацию о жанре, к которому относится произведение искусства; «Город» (CITY) – информация городе, где находится (находился) художественный объект; «Положение» (POSITION) – положение карти-

ны; «Местонахождение» (LOCATION) – местонахождение произведения живописи; «Тип местоположения картины» (ARTWORK\_TYPE\_POSITION) – тип местоположения картины (государственный музей, государственная галерея, частный музей, частная галерея и т. д.); «Эпоха» (EPOCH) – информация об эпохе живописи; «Школа живописи» (PAINTING\_SCHOOL) – информация о школе живописи; «Стиль живописи» (PAINTING\_STYLE) – информация о стиле написания картины; «Общий профиль пользователя» (GENERIC\_USER) – общая информация о пользователях системы; «Пользователь» (USER) – авторизационные данные пользователей системы; «Организация» (ORGANIZATION) – информация об организации, в которой работает пользователь; «Привилегии» (PRIVILEGES) – описаны привилегии, которые доступны тому или иному пользователю в системе.

Следующее направление, на наш взгляд, логичным образом связанное с персоналиями и результатами их деятельности, представляет собой миграционные и демографические процессы.

В современном мире движение населения, связанное как с миграцией, так и с обычным

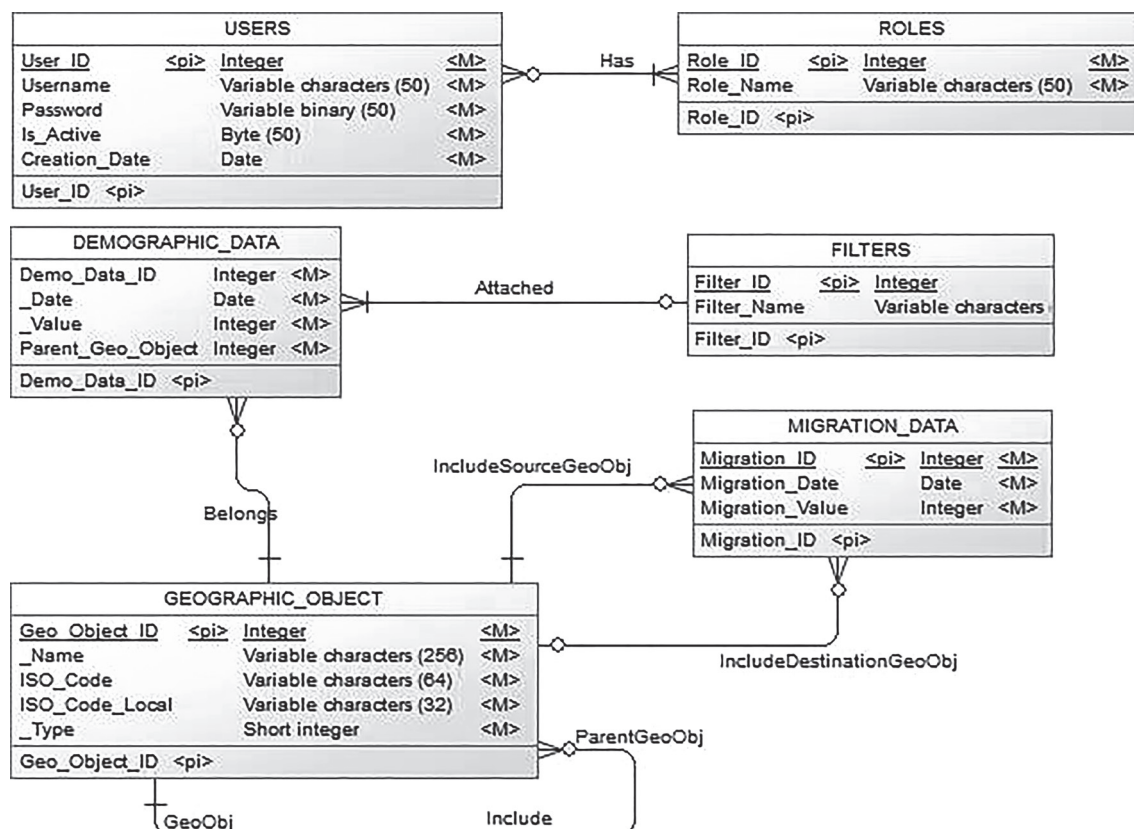


Рис. 5. Фрагмент концептуальной модели данных, связанной с демографическими и миграционными показателями



туризмом, стремительно растет. Этому способствуют высокая интегрированность современного мира, упрощение визового режима между странами, лояльное отношение стран Европы к политической миграции и другие факторы. Сбором статистики такого рода как правило занимаются некоммерческие организации или государственные учреждения.

Сбор таких статистических сведений чаще всего носит изолированный характер и представляет лишь численные показатели выбранных метрик. Полученные данные принимаются как отображение текущего положения. При анализе таких данных интересна, в первую очередь, их связь с внешними факторами, такими, как демографические, политические, социальные или иные побочные факторы, которые могут являться причиной полученных значений или их изменения с течением времени.

Исходя из предметной области, выделены сущности для системы обработки данных, связанных с миграциями населения «Пользователи» (USERS) – пользователи разрабатываемого приложения; «Роли пользователей» (ROLES) – доступные пользователю роли; «Географические объекты» (GEOGRAPHICOBJECTS) – сущность для хранения данных о географическом регионе; «Миграционные данные» (MIGRATIONDATA) – набор всех загруженных в базу данных миграционных данных; «Демографические данные» (DEMOGRAPHICDATA) – набор всех загруженных в базу данных демографических данных; «Демографические показатели» (FILTERS) – доступные демографические показатели. Соответствующая концептуальная модель представлена на рис. 5.

На заключительном этапе произведен анализ и интеграция полученных моделей в еди-

ную схему базы данных. Реализован соответствующий прототип базы данных на основе сервера MS SQL Server. В настоящее время проводятся работы по проектированию и разработке единого хранилища данных на основе открытой архитектуры и проработка вопроса о возможных методах обработки данных, включая построения систем на основе технологии OLAP.

### Заключение

В работе изложен общий подход к представлению данных предметных областей практико-ориентированной направленности, связанных с различными видами деятельности людей и являющихся основой общей модели данных для создания единого хранилища данных. Приводится краткая характеристика предметных областей, которые взяты за основу на текущий момент времени, предложены соответствующие модели данных, а также рассмотрена их взаимосвязь между собой в аспекте обработки данных и построения единой системы.

Интересным в дальнейшем, несомненно, представляется разработка универсальной системы, построенной на основе технологии складирования данных, которая способна работать сразу со многими источниками данных, представлять хранимые данные в удобном виде, осуществлять их быстрый поиск, а также проводить как интеллектуальный анализ данных конкретной предметной области, так и отслеживать возможные межпредметные взаимодействия. Например, проследить биографический и творческий путь той или иной личности в аспекте его привязке к местности, аспекте миграционных процессов или же, например, в аспекте простых перемещений.

### Литература

1. Рудикова, Л. В. О разработке системы для поддержки лазерной экспрессной экспертизы. Монография / Л. В. Рудикова. – LAP LAMBERT Academic Publishing, 2014. – 134 с.
2. Барсегян, А. А. Анализ данных и процессов / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. – 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.: ил. + CD-ROM.
3. Паклин, Н. Б. Бизнес-аналитика: от данных к знаниям (+CD): учеб. пособие. / Н. Б. Паклин. – 2-е изд., перераб. и доп. – СПб.: Питер, 2010. – 704 с.: ил.
4. Рудикова, Л. В. Универсальная комплексная система, поддерживающая организацию лазерной экспрессной экспертизы / Л. В. Рудикова // Доклады БГУИР. – 2013. – № 3 (73) – С. 26–32.
5. Рудикова, Л. В. Разработка программного визуализатора спектров для поддержки лазерной экспрессной экспертизы / Л. В. Рудикова // Доклады БГУИР. – 2014. – № 1 (79) – С. 46–52.
6. Рудикова, Л. В. О концепции универсальной системы хранения и обработки данных произведений художественной ценности / Л. В. Рудикова // Фундаментальные проблемы естествознания и техники. Серия: Проблемы исследования Вселенной. – Санкт-Петербург, 2016. – Т. 37, № 2. – С. 209–227.

7. Рудикова, Л. В. О проектировании системы для поддержки экспертизы объектов художественной ценности / Л. В. Рудикова // Информационные системы и технологии: управление и безопасность: сб. ст. IV междунар. заоч. науч.-практ. конф. / Поволж. гос. ун-т сервиса. – Тольятти, 2016. – С. 154–167.
8. Belyi, A. Global multi-layer network of human mobility // Alexander Belyi, Iva Bojic, Stanislav Sobolevsky, Izabela Sitko, Bartosz Hawelka, Lada Rudikova, Alexander Kurbatski, Carlo Ratti / International Journal of Geographical Information Science. – Mode of access: [http://www.tandfonline.com/doi/full/10.1080/13658816.2017.1301455]. – Date of access: [20.04.2017].
9. Community structure in time-dependent, multiscale, and multiplex networks / P. J. Mucha [et al.]. – Science, 2010. – Vol. 328, № 5980. – P. 876–878.
10. Inmon, W. H. Building the Data Warehouse / W. H. Inmon // Third Edition. – John Wiley & Sons, Inc. New York, 2002. – 428 p.
11. Kimbell, R. The Data Warehouse Toolkit: The Complete Guide to Dimensional Data Warehouses / R. Kimbell, M. Ross // Second Edition. – J. Wiley & Sons, 2002. – 447 p.
12. Примеры реализации хранилищ данных предприятия [Электронный ресурс]. – Режим доступа: http://www.internet-technologies.ru/articles/article\_994.html. – Дата доступа: 26.12.2016.

## References

1. Rudikova, L. About laser express expertise system implementation. Monography / Lada Rudikova. – LAP LAMBERT Academic Publishing, 2014. – 134 p.
2. Barseghyan, A. Data and processes analysis / A. Barseghyan, M. Kupriyanov, I. Kholod, M. Tess, S. Elizarov. – 3rd edition. – StP.: BHV-Petersburg, 2009. – 512 p.: il. + CD-ROM.
3. Paklin, N. Business-analytics: from data to knowledge (+CD) / N. Paklin, V. Oreshkov. – 2nd edition. – StP.: Piter, 2010. – 704 p.: il.
4. Rudikova, L. An universal complex system of laser express expertise organization // L. Rudikova / Doklady BGUIR. – 2013. – № 3 (73) – P. 26–32.
5. Rudikova, L. Software specter visualizer for laser express expertise support. // L. Rudikova / Doklady BGUIR. – 2014. – № 1 (79) – P. 46–52.
6. Rudikova, L. V. On the concept of an universal system for storing and processing artwork objects data / L. V. Rudikova // Fundamental Problems of Natural Science and Technology. Series: Problems of the study of the universe. – St. Petersburg, 2016. – P. 37, № 2. – P. 209–227.
7. Rudikova, L. V. On the design of the system for the support of the examination of artwork objects / L. V. Rudikova // Information Systems and Technologies: Management and Security: Sat. Art. IV international correspondence scientific-practical conference / Povolzhsky State University of Service. – Togliatti: Publishing house of VGUS, 2016. – P. 154–167.
8. Belyi, A. Global multi-layer network of human mobility // Alexander Belyi, Iva Bojic, Stanislav Sobolevsky, Izabela Sitko, Bartosz Hawelka, Lada Rudikova, Alexander Kurbatski, Carlo Ratti / International Journal of Geographical Information Science. – Mode of access: [http://www.tandfonline.com/doi/full/10.1080/13658816.2017.1301455]. – Date of access: [20.04.2017].
9. Community structure in time-dependent, multiscale, and multiplex networks / P. J. Mucha [et al.]. – Science, 2010. – Vol. 328, № 5980. – P. 876–878.
10. Inmon, W. H. Building the Data Warehouse / W. H. Inmon // Third Edition. – John Wiley & Sons, Inc. New York, 2002. – 428 p.
11. Kimbell, R. The Data Warehouse Toolkit: The Complete Guide to Dimensional Data Warehouses / R. Kimbell, M. Ross // Second Edition. – J. Wiley & Sons, 2002. – 447 p.
12. Implemented data ware houses of company examples [Electronic source] / Internet-technologies. – Mode of access: http://www.internet-technologies.ru/articles/article\_994.html. – Date of access: 26.12.2016.

Поступила  
02.05.2017

После доработки  
12.06.2017

Принята к печати  
10.09.2017

Rudikova, L. V. Zhavnerko E. V.

## ABOUT DATA MODELING SUBJECT DOMAINS PRACTICE-ORIENTED DIRECTION FOR UNIVERSAL SYSTEM OF STORAGE AND PROCESSING DATA

*This article describes data modeling for practice-oriented subject domains they are basis of general data model for data warehouse creation. Describes short subject domains characteristic relationship to different types of any human activities at the current time. Offered appropriate data models, considered relationship between them as data processing and data warehouse creation, which can be built on information data storage technology and which has some characteristics as extensible complex subject domain, data integration, which get from any data sources, data time invariance with required temporal marks, relatively high data stability, search necessary compromises in data redundancy, system blocks modularity, flexibility and extensibility of architecture, high requirements to data storage security. It's proposed general approach of data collection and data storage, appropriate data models, in the future, will integrate in one database scheme and create generalized scheme*



of data warehouse as type «constellation of facts». For getting of data models applies structural methodology and consider general principles of conceptual design. Using complex system, which can work with some information sources and represent data in convenient view for users will in-demand for analysis data selected subject domains and determination of possible relationships.

**Keywords:** data warehouse, subject domain, universal system, data scheme, structural methodology, practice-oriented data



**Lada Rudikova** is the Head of Modern Programming Technologies Department of Yanka Kupala State University of Grodno (YKSUG). Ph. D. degree in physical and math.

The main line of her scientific researches – management theory, information systems design, databases, CASE, data mining, business intelligence. She actively participates in international conferences. She is the author of more than 280 scientific works and books related to computer technology and data processing, a technical writer of the publishing house «BHV-St Petersburg»



**Eugene Zhavnerko** is the senior lecturer of Modern Programming Technologies Department of Yanka Kupala State University of Grodno (YKSUG). Master's degree in technical sciences.

The leading activities of his scientific researches – information systems design, database, CASE, data mining, business intelligence, modern approaches to software development.

Результаты работы получены в процессе выполнения ГПНИ «Разработка методологии и средств построения универсальных систем хранения, обработки и анализа структурированных данных большого объема практико-ориентированной направленности».