

УДК 004

В. А. Серебряков

*Вычислительный центр им. А. А. Дородницына РАН
ул. Вавилова, 40, Москва, 119333, Россия*

serebr@ultimeta.ru

РАБОТЫ ВЫЧИСЛИТЕЛЬНОГО ЦЕНТРА РАН В ОБЛАСТИ РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ*

Рассматриваются реализованные в ВЦ РАН проекты распределенных информационных систем в области поддержки научных исследований. Среди них портал Российской академии наук, проект создания Единого информационного пространства РАН, цифровая библиотека «Научное наследие России», геопортал «ГеоМета», персональная семантическая цифровая библиотека, система интеграции данных по теплофизике. Рассматриваются проблемы, возникающие при внедрении научных информационных систем.

Ключевые слова: распределенные информационные системы, научные исследования, единое информационное пространство, цифровая библиотека, Научное наследие России, ГеоМета, интеграция данных.

Введение

Работы в направлении создания систем, интегрирующих информационные ресурсы РАН, были начаты в 1998 г. Благодаря поддержке Межведомственной программы «Национальная сеть компьютерных телекоммуникаций для науки и высшей школы» был реализован пилотный проект «Интегрированная система информационных ресурсов (ИСИР) РАН». Успешное завершение этого проекта позволило развернуть работы по интеграции разнородных научных информационных ресурсов в общеакадемическую научную информационную систему. В 2001 г. по инициативе Отделения математики РАН была принята новая программа целевых расходов Президиума РАН «Информатизация научных учреждений и Президиума РАН». Главной задачей этой программы стала поэтапная интеграция информационных ресурсов организаций РАН в объединенное информационное пространство – Единую информационную систему (ЕИС) РАН. Координация этих работ осуществлялась Советом РАН «Научные телекоммуникации и информационная инфраструктура». Основная часть работ по собственно разработке системы была выполнена в отделе систем математического обеспечения ВЦ РАН и отделе информационных технологий ЦНТК РАН.

Первоочередной задачей проекта ЕИС РАН стала разработка концептуальной основы и инфраструктуры для интеграции разнородных информационных и вычислительных ресурсов организаций РАН в единое информационное пространство. Единое информационное

* Статья написана по материалам заказного доклада, представленного на XV Российскую конференцию с международным участием «Распределенные информационно-вычислительные ресурсы» (DICR-2014), Новосибирск, 2–5 декабря 2014 г. (<http://conf.nsc.ru/dicr2014/>).

пространство (информационную инфраструктуру фундаментальных и прикладных исследований РАН) должны составлять всевозможные цифровые библиотеки, информационные и вычислительные системы организаций РАН, использующие как собственные принципы организации, так и, по возможности, технологию открытой архитектуры проекта ЕИС или непосредственно ее релизы. В результате был подготовлен системный проект, который определил структуру системы как таковой, типы информационных ресурсов, участвующих в системе, общую функциональность компонентов системы. В проекте также были отражены принципы организации распределенности системы и интеграции в систему уже существующих ресурсов.

Важным этапом был проект создания Единого научного информационного пространства (ЕНИП) РАН. Этот проект основывался на подсистеме «Научный институт РАН», созданной в ВЦ РАН и ЦНТК РАН. Эта подсистема обеспечивает возможность интеграции информационных ресурсов отдельных организаций в ЕИС. На базе этой системы были реализованы Web-информационные системы ряда институтов и отделений РАН, а также такие информационные системы, как Научное наследие России, портал интеграции пространственных данных Геомета и ряд других. Система была запущена в опытную эксплуатацию.

Интегрированная система информационных ресурсов РАН

С 2001 г. выполнялась целевая программа Президиума РАН «Информатизация научных учреждений и Президиума РАН» (с 2004 г. – «Информатизация»). С самого начала официальной деятельности по программе значительные усилия были приложены к выработке согласованного системного взгляда на стоящие проблемы и пути их решения, к формированию целей и задач, подходов к решению, базовых требований к используемым методам (технологиям, стандартам и т. п.). С участием всех заинтересованных сторон был разработан ряд документов, положенных в основу большинства проектов, выполняющихся в рамках Программы.

В целом надо сказать, что до определенного момента все процессы, связанные с применением информационных технологий в РАН, двигались полностью бессистемно, не управлялись и не контролировались. Тем более не было никакого анализа полученных результатов, эффективности вложений и т. п. С момента деятельности рабочих групп по программе информатизации эта бессистемность постепенно начала исчезать. Кроме бессистемности, были и такими проблемы, как:

- отсутствие полного понимания, согласованного со всеми заинтересованными сторонами в РАН, необходимости развития работ в направлении интеграции;
- как следствие, задержки при окончательной формулировке и принятии общей концепции и программы работ по информатизации РАН;
- отсутствие юридической базы, которая могла бы создать условия для защиты авторских прав и прав интеллектуальной собственности на разработки, выполняемые в РАН;
- различные уровни подготовленности организаций РАН к внедрению и использованию современных ИТ;
- отсутствие или недостаточная подготовленность к интеграции базовых информационных блоков, которыми должны быть информационные системы институтов, центральных библиотек, отделений и Президиума РАН;
- отсутствие адаптированных к требованиям РАН разработок в области стандартизации объектов и механизмов единой системы;
- как следствие, отсутствие возможности полноценного обмена информацией в электронном виде.

Общая задача Интегрированной системы информационных ресурсов (ИСИР) РАН [1; 2] состоит в организации единого информационного пространства. Это требует решения задач по извлечению и структуризации метаданных, по обеспечению их ввода в структурированном виде. Второй класс задач состоит в предоставлении средств интеграции информации разнообразных информационных систем (репозиторий), тем или иным способом накопивших структурированную информацию. Модель данных представлена на рис. 1.

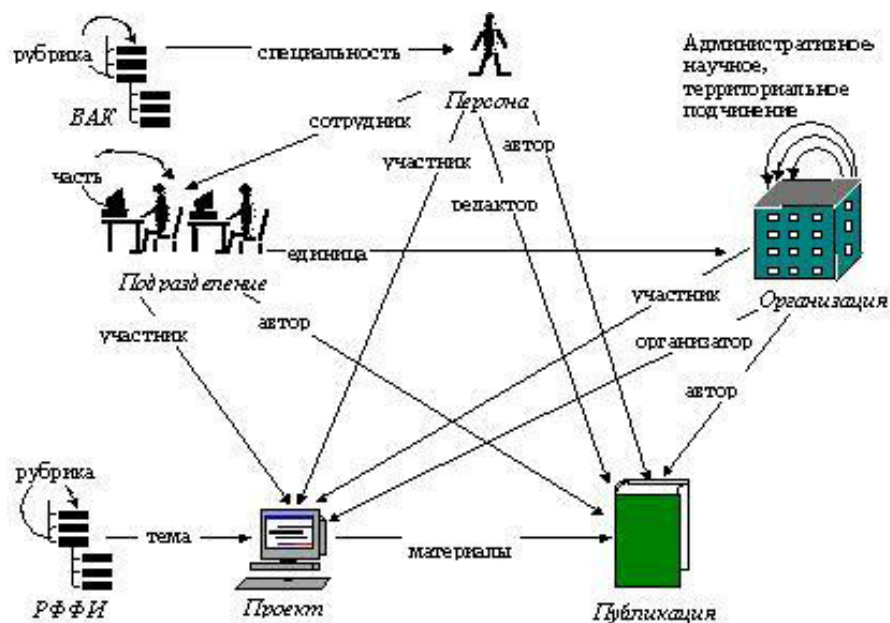


Рис. 1. Архитектура ИСИР РАН



Рис. 2. Портал РАН

С точки зрения пользователя ИСИР представлена как Портал РАН. Портал реализован (совместно с ИПИ РАН) как Информационно-поисковый справочник РАН, ориентированный на накопление и предоставление оперативной научно-административной информации. В настоящий момент основными типами ресурсов справочника (Портала РАН) являются следующие:

- организации РАН в соответствии со структурным делением РАН (президиум, отделения, секции, научные центры, филиалы РАН) и сведения о них;
- сотрудники РАН (аппарат РАН, аппарат отделений, руководство организаций и учреждений, научные сотрудники) и сведения о них (адреса, телефоны и т. д.);
- публикации;
- проекты.

Справочник РАН отражает организационно-структурное деление РАН, позволяет получить информацию о структурных подразделениях РАН и обеспечивает доступ к информационным ресурсам этих подразделений, данным о сотрудниках учреждения, их научной деятельности.

Исполнительная система справочника обеспечивает следующие возможности:

- просмотр информации и средства навигации по структуре информации;
- поиск информации по различным видам запросов и просмотр выданной по запросам информации;
- средства ввода, редактирования и сопровождения информации;
- средства администрирования непосредственно в подразделениях РАН (рис. 2).

При выборе платформы были проанализированы существующие инструментальные средства для создания подобной системы. Выбор был сделан в пользу платформы ASP.NET, как обеспечивающей максимальную производительность, удобные средства разработки, компонентную ориентированность, открытость и расширяемость архитектуры, позволяющую вмешиваться практически во все этапы обработки поступающих Web-запросов.

Текущее состояние портала представлено на рис. 2.

Информационная система «Научный институт РАН»

Рассмотрим типовой научный институт, входящий в состав РАН. Он представляет собой полноценную организацию со сложной административной структурой, основным направлением деятельности являются научные исследования. Задачи, решаемые каждой такой структурной единицей РАН, можно разделить на административные, научные, публичные и т. д.

Административные задачи. В любой организации для нормального функционирования требуется постоянное решение управленческих задач, влияющих прямым образом на деятельность организации в целом и на выполнение конкретных задач на всех уровнях. Это такие задачи, как управление организационной структурой и кадрами, управление проектами, обеспечение документооборота и проч.

Научные задачи. Основным направлением деятельности любого научного института РАН являются научные исследования, а основной задачей организации – обеспечение научной деятельности сотрудников.

Публичные задачи. Взаимодействие с другими научными учреждениями, организация и проведение конференции и научных семинаров, публикация научных трудов сотрудников, предоставление доступа к результатам научных экспериментов, научным данным – все это неотъемлемая часть деятельности научного института.

Информационная система Института РАН должна, с одной стороны, стать центром научно-информационного сервиса сотрудников Института, а с другой – обеспечивать полное представление информации о научной деятельности Института для мирового сообщества. Информационная система Института РАН должна представлять собой узел в распределенной архитектуре множества узлов – информационных систем институтов РАН.

На основе описанных выше задач научных организаций в составе РАН можно сформулировать набор требований к программному комплексу ИС «НИ РАН».

Информационная система «Научный институт РАН» должна:

- обеспечивать решение основных информационных задач научного института в составе Российской академии наук;
- позволять гибко изменять конфигурацию системы под нужды конкретной организации, реализацию новых модулей для решения специфических задач;
- предоставлять средства интеграции и структуризации существующих данных;

- обеспечивать поддержку распределенного взаимодействия, в том числе со сторонними системами (через специализированные адаптеры, создаваемые отдельно).

Система должна включать:

- средства интеграции существующих данных;
- автоматизированные интерактивные средства структуризации и пакетной загрузки данных;
- пользовательские и административные интерфейсы ввода новых данных и управления уже находящимися в системе данными;
- систему (возможно распределенную) хранения данных;
- систему безопасности, обеспечивающую аутентификацию пользователей и авторизацию доступа к ресурсам системы;
- спецификации по разработке дополнительных модулей, обеспечивающих решение специфических задач научного института.

ИС «НИ РАН» [3; 4] представляет собой типовой программный комплекс автоматизации информационной деятельности научного института в составе Российской академии наук, обеспечения научной деятельности его сотрудников, взаимодействующий с другими информационными системами в составе ЕНИИП.

Разработанная платформа ИС «НИ РАН» предоставляет широкие возможности по конфигурированию под нужды конкретного научного института. Ядро всей системы составляют инфраструктурные службы. Они обеспечивают хранение, индексирование и поиск ресурсов, обеспечивают безопасность и взаимодействие между другими модулями. Базовые компоненты ИС «НИ РАН» обеспечивают выполнение самых общих информационных задач научного института – управление содержанием портала, организационной структурой, ведение сведений о публикациях и проектах сотрудников. Все действия конечный пользователь производит через Web-интерфейс.

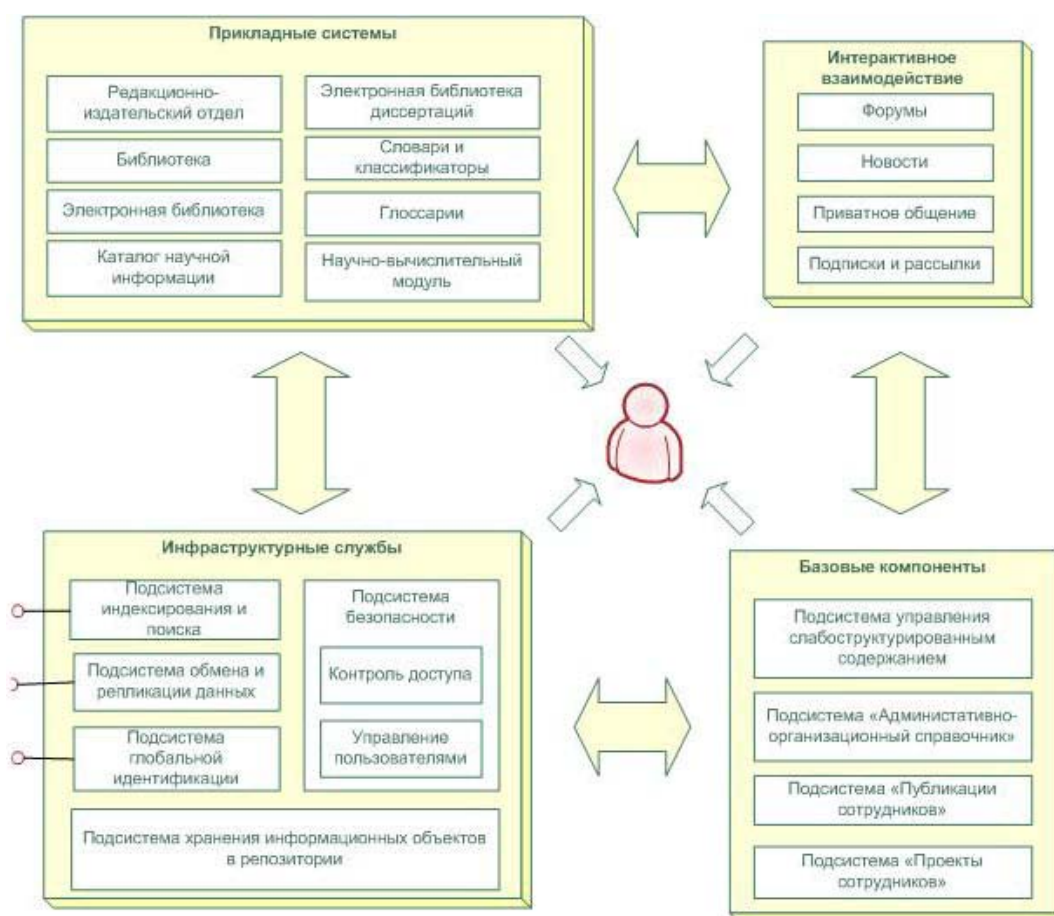


Рис. 3. Архитектура НИ РАН

ИС «НИ РАН» представляет собой модульную расширяемую систему, решающую типовые информационные задачи научного института в составе Российской академии наук. Но реальные потребности таких организаций и их сотрудников зачастую бывают очень специфичными и относятся к узкой предметной области. Для удовлетворения таких нужд разрабатываются прикладные подсистемы, расширяющие функциональность типового решения в конкретных экземплярах. Архитектура НИ РАН представлена на рис. 3.

В основном профиле метаданных можно выделить общую поддержку следующих четырех основных групп информационных сущностей.

Участники научной деятельности – центральное звено, вся информация в РАН связана с научной деятельностью ее сотрудников, «Персон», образующих разнообразные организационные объединения от формальных («Организации» и «Подразделения») до неформальных («Коллективы», «Сообщества», «Рабочие группы»).

«Научная деятельность», в частности, «Проекты», отражающие процесс научной деятельности, информация о результатах проектов, патентах и т. п., а также Научные мероприятия – как разовые, так и повторяющиеся, такие как «Конференции», «Семинары», «Симпозиумы».

«Результаты научной деятельности», в которые могут входить «Интернет-системы» – Web-сайты и пр., «Базы данных», предоставляющие автономные коллекции информации с той или иной степенью интеграции с ЕНИП и т. п., «Экспериментальные данные» и их «Математические модели», «Программные системы», в частности, «Научные вычислительные приложения», «Экспериментальные установки», «Изобретения», «Технологии», и т. п.

Документы и публикации – ресурсы этого типа представляют собой научные труды, статьи, отчеты сотрудников (научные «Публикации» и «Диссертации» сотрудников), возможно, административные «Постановления» и «Распоряжения». Примерами специализации публикации могут служить «Тезисы конференций» и т. п. Профиль метаданных НИ РАН представлен на рис. 4.

На базе информационной системы Научный институт РАН были созданы информационные системы ряда организаций (институтов и отделений) РАН:

- Отделение общественных наук (на базе системы Соционет);
- Библиотека по естественным наукам (БЕН РАН);
- Вычислительный центр (ВЦ РАН);
- Институт физики твердого тела;
- Палеонтологический институт;
- Пермский научный центр и институт механики сплошных сред УРО;
- Институт проблем химической физики и научный центр Черноголовка;
- Тихоокеанский океанологический институт им. В. И. Ильичева;
- Отделение математических наук (ОМН);
- Санкт-Петербургский научный центр;
- Дальневосточное отделение (ДВО РАН);
- Институт научной информации по общественным наукам (ИНИОН);
- Институт США и Канады (ИСКРАН);
- Институт проблем информатики (ИПИ РАН);
- Портал пространственных метаданных ГеоМета;
- Цифровая библиотека «Научное наследие РАН»;
- Северокавказский научный центр.

Единое научное информационное пространство РАН

Российская академия наук обладает уникальными научными информационными ресурсами. Среди них – опубликованные результаты научных исследований и экспериментов, библиографические и фактографические базы данных, сведения об ученых, их научной деятельности, публикациях, проектах и т. п. Эти ресурсы представляют значительный интерес для сотрудников РАН, членов мирового научного сообщества, для представителей промыш-

ленности и предпринимателей, которые заинтересованы во внедрении результатов научных исследований.

Предполагалось, что Единое научное информационное пространство (ЕНИП) РАН должно стать интегрированным источником научной информации.

Система предусматривала объединение сведений о разнородных научных информационных ресурсах РАН, обеспечение актуальности этих сведений и широких возможностей для достаточно точного поиска научных ресурсов на основе этих сведений, поддержку средств научной коммуникации, сервисов, связанных с возможностью оперативного информирования пользователей о необходимых им ресурсах и т. п.

Такая система могла обеспечить пользователей актуальными данными о текущем состоянии и характеристиках информационно-научной базы институтов РАН и их подразделений, упростить анализ состояния и тенденций развития науки. Облегченный доступ к информации мог бы изменить способы ведения научной деятельности, способы обучения.

Для обеспечения взаимодействия существующих разнородных научных систем в рамках ЕНИП предполагалось выработать корпоративные стандарты на интерфейсы взаимодействия, а также профили метаданных, что позволило бы реализовать инструментальные средства, обеспечивающие интеграцию данных в единую среду. Результатом решения этих первоочередных проблем должны были явиться предложения ЕНИП по:

- типовым интерфейсам взаимодействия (форматы данных, протоколы обмена) отдельных информационных источников (организаций РАН, поддерживающих собственные научные информационные ресурсы);
- профилям метаинформации, предоставляемой этими источниками. В частности, производится разработка набора элементов метаданных для научной информации общего характера, предложений по формированию элементов метаданных для отдельных областей науки и согласование их с научным сообществом и международными открытыми стандартами;
- справочникам и классификаторам ресурсов;
- реализации политики информационной безопасности и требований по разграничению прав доступа к цифровым ресурсам.

ЕНИП РАН было призвано помочь научным коллективам сделать ряд шагов в направлении интеграции разнородных научных информационных и программных ресурсов отдельных научных учреждений, предоставления пользователям более эффективных средств интеграции и поиска информации, научной коммуникации, сотрудничества и совместной работы. Под единым пространством понимается не формирование централизованной системы, не навязывание всем одних и тех же решений, а стремление к последовательности практических шагов, совместными усилиями научных коллективов:

- сформулировать взаимосогласованный набор соглашений, правил и открытых стандартов;
- приготовить совокупность макетов и типовых решений для реализации адаптеров прикладных систем, инфраструктурных служб, поддерживающих разные уровни интероперабельности распределенных гетерогенных данных и приложений;
- создать ряд информационных систем общего назначения, следующих этим соглашениям, использующих эти реализации, допускающих модульную организацию, наращивание функциональных возможностей;
- применить эти результаты для решения соответствующих задач научных учреждений.

Основу ЕНИП РАН составляют, прежде всего, стандарты на метаданные информации, циркулирующей в ЕНИП. Эти стандарты должны отвечать следующим требованиям:

- включать основные типы информации, требующейся для поддержки работы научного сотрудника;
- быть открытыми, т. е. обеспечивать доступ к информации по этим описаниям;
- быть расширяемыми, т. е. обеспечивать возможность детализации описаний;
- обеспечивать возможности интеграции информации;
- обеспечивать возможности уникальной идентификации информации;
- обеспечивать возможности размещения и поиска информации в распределенной среде;

- быть ориентированными на современные и перспективные технологии описания и использования информации (в нашем понимании – ориентироваться на семантический Веб (Semantic Web));

- обеспечивать возможности интероперабельности с внешней средой.

Основу единого информационного пространства РАН составляет Информационная магистраль ЕНИП РАН (рис. 5), представляющая собой комплекс аппаратных, программных и организационных мер, обеспечивающих:

- формирование состава цифровых ресурсов и служб ЕНИП РАН;
- предоставление доступа к цифровым ресурсам и службам ЕНИП РАН;
- обеспечение защиты цифровых ресурсов и служб ЕНИП РАН;
- ведение и поддержку в актуальном состоянии метаданных системы;
- поиск по хранимой метаинформации и идентификацию ресурсов;
- интеграцию ресурсов различных областей и отраслей знаний.

На рис. 6 представлена схема взаимодействия узлов ЕНИП. Изображены четыре типа узлов ЕНИП:

- центральный узел;
- узлы организаций;
- независимые источники данных;
- независимые системы, включенные в ЕНИП.

Центральный узел осуществляет интеграцию данных с других узлов с помощью механизма репликации (копирования) метаинформации. На основе реплицируемой на центральный узел метаинформации строятся поисковые индексы и на их базе осуществляется единый поиск по этим узлам. Загрузка данных в узлы системы может осуществляться из других источников, например из сайтов организаций. Независимые информационные системы могут быть включены в ЕНИП самостоятельно, если обеспечены протоколы взаимодействия.

Система ЕНИП оперирует такими ресурсами, как персоны, публикации, организации, подразделения и проекты. Данные по этим ресурсам обновляются каждую неделю. Центральный узел предоставляет пользователям две возможности поиска: поиск по локальной базе данных и полнотекстовый поиск. Поиск по локальной базе осуществляется по стандартным ресурсам: персона, организация, публикация, проекты. Актуальность информации может составлять разницу в 6 дней от информации на сервере-источнике данных.

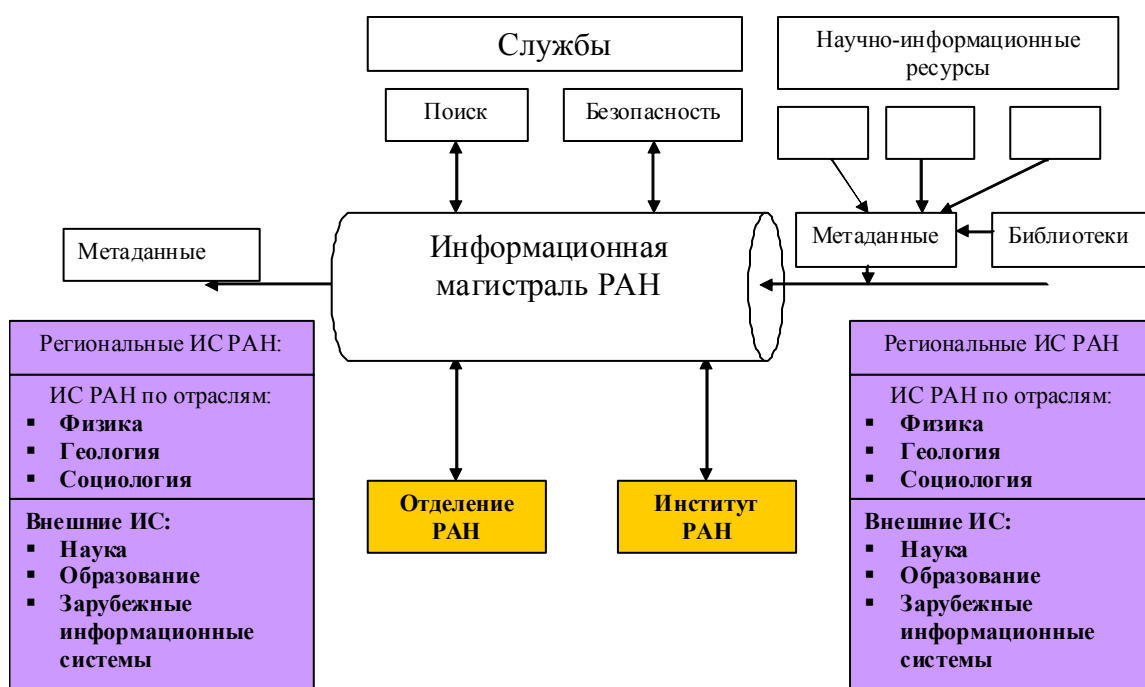


Рис. 5. Информационная магистраль ЕНИП РАН

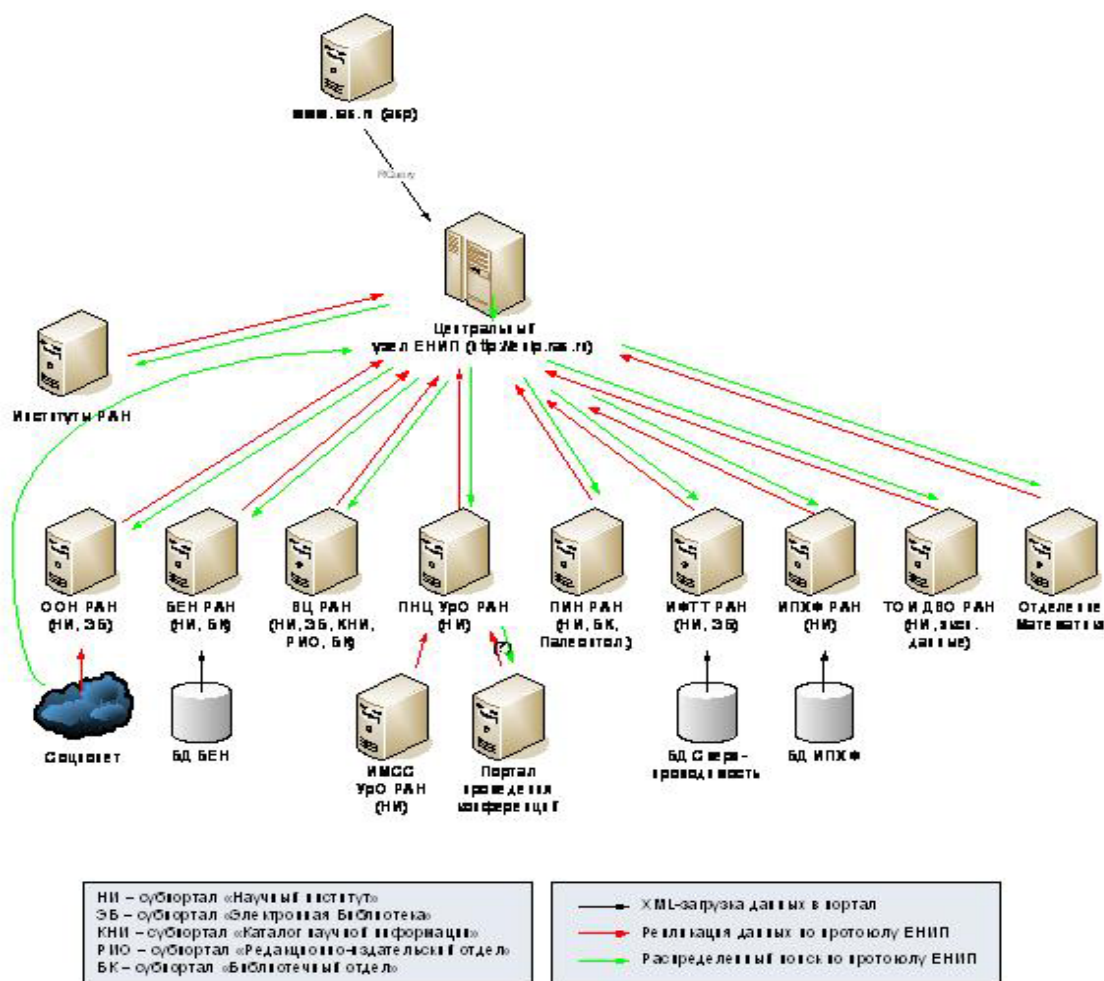


Рис. 6. Схема взаимодействия узлов ЕНИП

Полнотекстовый поиск позволяет получить полную и актуальную информацию, но выполняется дольше.

ЦБ «Научное наследие России»

Учитывая важность формирования цифровых библиотек, Российская академия наук приняла в 2006 г. целевую научную программу «Создание ЦБ “Научное наследие России”». ЦБ призвана аккумулировать цифровые копии книг, статей, документов, хранящихся в библиотеках, архивах и музеях РАН. В первую очередь акцент сделан на перевод в цифровую форму редких и уникальных изданий, важнейших документов по истории РАН, материалов экспозиционного характера, включая аудио- и видеоматериалы [5].

Основная цель создания ЦБ – предоставление через Интернет всем желающим информации о выдающихся российских ученых, внесших вклад в развитие фундаментальных естественных и гуманитарных наук, с возможностью ознакомления с полными текстами опубликованных ими наиболее значительных работ. Исходя из этой цели в ЦБ было решено включать не только электронные версии книг, но и развернутые сведения о российских ученых – биографические данные, основные этапы их научной деятельности, разнородную архивную и музейную информацию, отсканированные фотографии, аудио- и видеозаписи, относящиеся к теме научного наследия.

Другая цель создания ЦБ – обеспечение сохранности оригиналов изданий, являющихся исторической ценностью, поскольку возможность работы с цифровыми копиями существенно снижает потребности в работе с печатными материалами, а каждая «книговыдача» на руки раритетных изданий сокращает срок их «жизни».

Третья цель создания ЦБ – включение сведений об ученых и основных результатах их научной деятельности в ЕНИП РАН. На начальном этапе реализации основными задачами Программы были разработка основных принципов формирования ЦБ, технологии сканирования, обработки и предоставления пользователям материалов, включаемых в Библиотеку, а также создание программного обеспечения, сопровождающего все этапы создания ЦБ.

В основу технологии формирования ЦБ положен принцип распределенного наполнения и централизованной поддержки. Руководство Программой осуществляет Межведомственный суперкомпьютерный центр РАН, разрабатывающий вместе с ВЦ РАН и БЕН РАН технологии и программного обеспечения наполнения и поддержки ЦБ. Основными поставщиками информации для загрузки в ЦБ в настоящее время являются центральные академические библиотеки (БАН и БЕН РАН с их отделами в институтах и научных центрах РАН), ИНИОН, Центральный архив РАН с его санкт-петербургским филиалом, Геологический музей РАН им. В. И. Вернадского, Институт русской литературы РАН (Пушкинский дом).

В настоящее время наполнение ЦБ осуществляется копиями изданий, которые не подпадают под действие закона о защите авторских прав (в основном это издания, вышедшие из печати до 1920 г.).

Основные элементы функциональности распределенной цифровой библиотеки (рис. 7):

- *доступ к ресурсам* – запрос, определение местоположения, извлечение, трансформация и сохранение ресурса; поиск может осуществляться как по атрибутам ресурса, так и по полным текстам;
- *управление ресурсом* – создание нового ресурса, внесение его в ЦБ, удаление старого ресурса и изменение существующего;
- *управление метаданными* – их создание, обработка и преобразование; состав метаданных определяется соглашениями;
- *управление словарями* – их создание, обработка и преобразование; состав словарей определяется соглашениями;
- *управление участниками* – их регистрация, подписка, права доступа и персональная информация;
- *управление цифровой библиотекой* – управление коллекциями, группами пользователей, членством, так же, как общее управление политикой, качеством или функциональностью;
- *системное администрирование* – установка, конфигурирование, необходимые периодические мероприятия, восстановление после сбоев и мониторинг ЦБ.

Цифровая библиотека строится как распределенная информационная система с выделенным центральным узлом. Узлы системы, с одной стороны, являются точками входа в цифровые библиотеки организаций – участников проекта, с другой – поставщиками информации для всей распределенной системы. Таким образом, ключевой принцип архитектуры – независимое развитие цифровых библиотек организаций – участников с одновременной интеграцией данных в единое информационное пространство. Это достигается стандартизацией предоставления метаданных, форматов предоставления данных, интерфейсов поиска и словарей. Таким образом, каждая из цифровых библиотек организаций-участников может хранить данные в собственных форматах и предоставлять собственные сервисы, но в то же время должна обеспечить единые для всех интерфейсы, упомянутые выше.

Центральный узел системы должен обеспечить навигацию, поиск и предоставление данных по всем цифровым библиотекам в соответствии с унифицированными форматами и сервисами.

Сервера хранения оцифрованных данных обеспечивают надежное хранение и резервирование оцифрованных данных библиотеки, а также подмножества метаданных, отражающих структуру информации (например, оглавление книг). Кроме того, на них возлагается задача по предоставлению доступа к данным конечных пользователей, перенаправленных с цен-

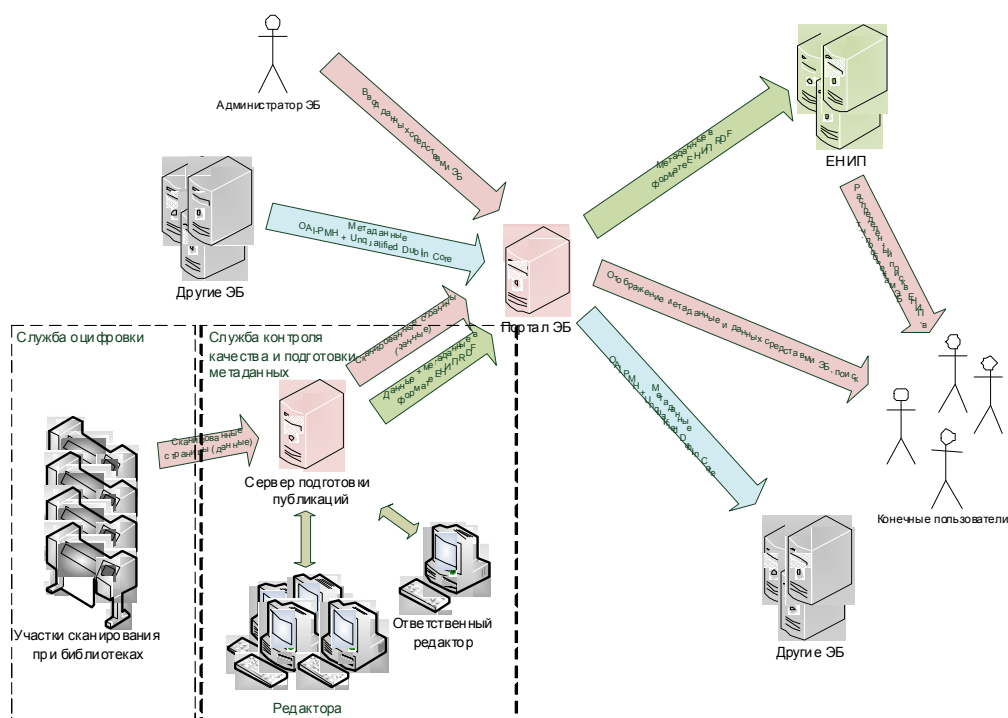


Рис. 7. Функциональная схема ЦБ «Научное наследие»

трального портала цифровой библиотеки. Серверы хранения данных для центров оцифровки предоставляют также средства автоматизации размещения и поддержания актуальности данных.

Центральный Web-портал цифровой библиотеки «Научное наследие РАН» осуществляет консолидацию метаданных, полученных из центров оцифровки, в рамках централизованного хранилища, обеспечивая, таким образом, централизованный доступ к ним пользователей. Взаимодействуя с серверами хранения оцифрованных данных, он является также единой точкой доступа к электронным версиям научных трудов. Вторая задача, решаемая центральным порталом, – обеспечение интеграции библиотеки в ЕНИП РАН путем предоставления на центральный сервер ЕНИП метаданных, по которым возможен распределенный поиск.

Функциональная схема ЦБ «Научное наследие» приведена на рис. 7, главная страница – на рис. 8.

Портал GeoMeta

К настоящему времени в учреждениях РАН накоплен большой опыт использования геоинформационных технологий, реализованы многочисленные геоинформационные проекты, созданы базы и банки пространственных данных.

Академические ресурсы пространственных данных составляют значительную часть национальных информационных ресурсов. Основным производителем пространственных данных являются учреждения геологического, геофизического, географического и экологического (природоохранного) профилей. В то же время данные рассредоточены, их использование ограничено, как правило, рамками того проекта, где они созданы, затруднены или невозможны поиск существующих данных и доступ к ним, не налажен обмен ими. Причина этого – отсутствие эффективной системы управления пространственными данными. Ее создание позволит интегрировать данные и знания о территории, строить и использовать модели природных и социально-экономических явлений и процессов, их взаимодействия в системе

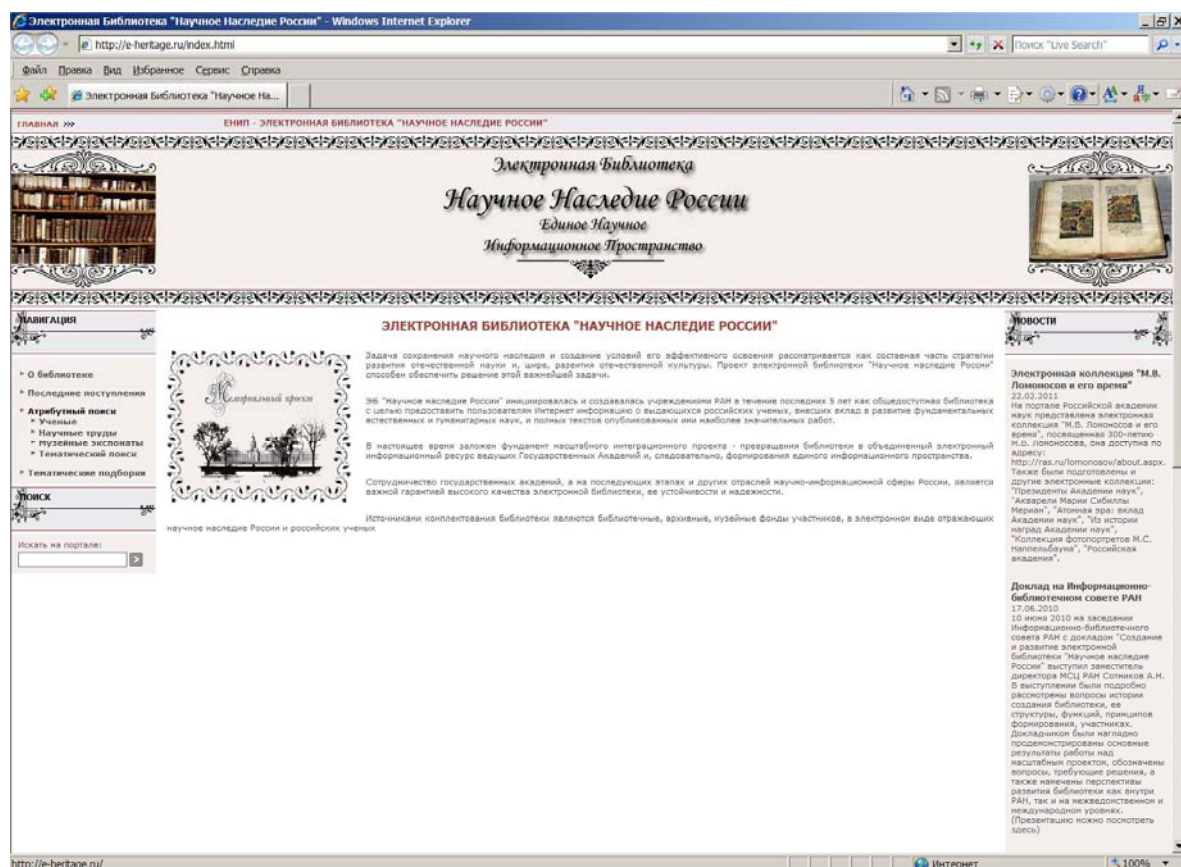


Рис. 8. Главная страница ЦБ «Научное наследие»

«общество – природная среда», использовать методы пространственного анализа, обеспечить территориальное планирование и управление.

В целом в учреждениях РАН имеется опыт выполнения разнообразных геоинформационных проектов для различных приложений, сформированы подразделения, отделы и лаборатории геоинформатики, укомплектованные высокопрофессиональными научными кадрами, располагающими необходимой технической базой, современными программными средствами геоинформационных систем (ГИС) и данными, т. е. созданы необходимые условия для разработки ГИС и их интеграции.

Основным инструментом интеграции и предоставления пространственных данных в настоящее время являются геопорталы. Понятие «геопортал» означает точку входа в Интернет с инструментами просмотра метаданных, поиска географической информации, ее визуализации, загрузки, распространения и, возможно, поиска геосервисов.

Современное требование к системам поддержки геопорталов – независимость, расширяемость и гибкость компонентов, являющаяся важной особенностью современной программной системной архитектуры.

Существует потребность в объединении этих данных, имеющих распределенный характер, в концептуально одну информационную систему, в обеспечении централизованного доступа к ним, в создании на основе Интернета технологий единого информационного пространства геоданных.

Портал «ГеоМета» [6] – это стандартизированная и децентрализованная среда управления пространственной информацией, разработанная для доступа к базам геоданных, картографическим продуктам и связанным с ними метаданным из различных источников, облегчающая обмен пространственной информацией между организациями и ее совместное использование посредством Интернета. Этот подход к управлению географической информацией имеет целью предоставить широкому сообществу пользователей средства для простого и своевре-

менного доступа к имеющимся пространственным данным и существующим тематическим картам, которые могут оказаться полезными для поддержки информированного принятия решений.

Главная цель портала – увеличить доступность разнообразных междисциплинарных данных различного масштаба вместе с сопутствующей информацией, организованных и документированных стандартным и непротиворечивым способом, улучшить кооперацию и координацию усилий при сборе данных, сохраняющих ресурсы и в то же время ограждающих данные и информацию от нежелательного доступа.

Портал «ГеоМета» представляет собой платформу для создания распределенной среды интеграции неоднородных источников геоинформационных данных и предоставления к этой среде единой точки входа (веб-портала), которая позволит ученым в сфере наук о Земле легко находить специализированные данные и приложения, производить вычислительные эксперименты, визуализировать результаты деятельности.

Благодаря тому, что портал «ГеоМета» построен на базе ИС «НИ РАН» [1], являющейся базовым инфраструктурным компонентом ЕНИП [2], он может интегрироваться в ЕНИП с предоставлением расширения схемы геопространственными метаданными и геоданными.

К функциональностям ГИС-части системы относятся:

- каталогизация, сбор, поиск геопространственных метаданных;
- размещение геоданных в собственном хранилище и предоставление к ним доступа;
- предоставление доступа к распределенным геопространственным данным по стандартизованным протоколам;
- визуализация карт, редактирование элементов.

Интерфейс системы представлен веб-порталом, поэтому для ГИС-части основным методом доступа пользователя к информации является обычный доступ к веб-страницам портала через любой распространенный браузер. Ядро системы предоставляет следующие возможности: управление статическим содержанием; хранение объектов системы (представленных RDF-тройками) в реляционных СУБД; индексирование и полнотекстовый поиск; обеспечение безопасности системы.

Система поддерживает следующие основные типы ресурсов: *Пространственные данные* (картографические данные и их метаданные) и дополнительные типы ресурсов, такие как *Организация*, *Персона*, *Публикация*, *Проект* и различные рубрикаторы и классификаторы.

Ресурс *Пространственные данные* содержит наборы пространственных данных и метаданные распределенных пространственных данных. Ресурс *Организация* включает организации РАН, научные центры и другие организации. Данные об их сотрудниках сопоставлены ресурсу *Персона*. Ресурс *Проект* поддерживает сведения о проектах, выполненных или ведущихся в РАН и других ведомствах. Ресурс *Публикация* представляет данные о публикациях и научной деятельности.

Доступ к portalу осуществляется интерактивно через Интернет посредством веб-браузера (например, Netscape Navigator или Microsoft Internet Explorer) по ссылке <http://www.geometa.ru>. Главная страница портала представлена на рис. 9.

Персональные семантические цифровые библиотеки

Под персональными семантическими цифровыми библиотеками подразумеваются такие цифровые библиотеки, наполнение которых индивидуально для каждого пользователя системы и выполняется в полуавтоматическом режиме из разнородных источников данных, интегрированных в облако LOD. Для краткости будем называть их персональными открытыми цифровыми библиотеками (ПОЦБ). Типы информационных ресурсов и их структура определяются пользователем исходя из его интересов, т. е. пользователь описывает интересующую его предметную область, определяя тематическое наполнение библиотеки.

Основная задача системы заключается в предоставлении пользователю унифицированного представления для возможности автоматизированного извлечения интересующей его информации по определенной предметной области.

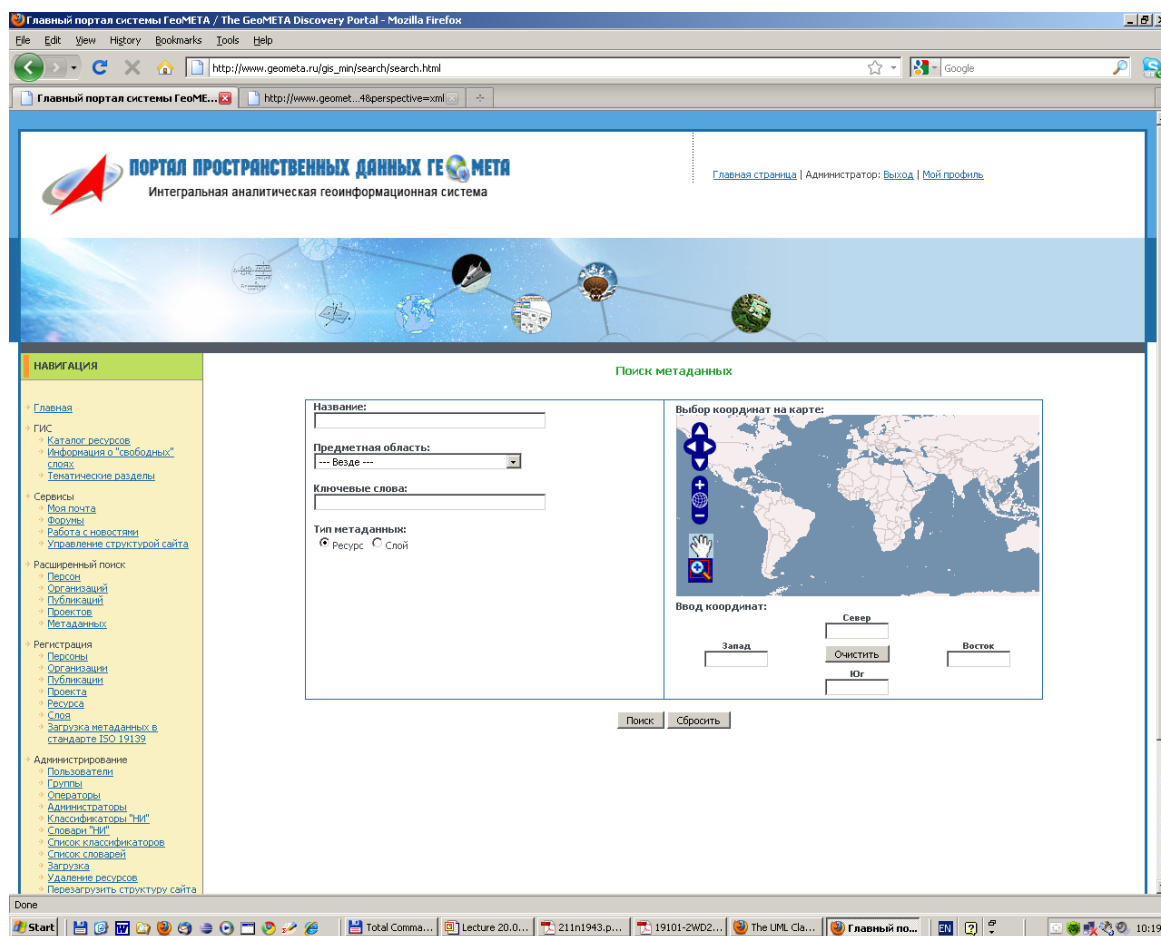


Рис. 9. Главная страница портала ГеоМета

Представление ресурсов библиотеки в виде связанных данных расширяет функциональность семантических цифровых библиотек, давая возможность:

- включать дополнительные элементы описания данных информационных ресурсов;
- полностью или частично обновлять данные из источников;
- использовать интерфейсы для создания запросов к интегрированным в LOD источникам данных на основе SPARQL;
- включать в описания ресурсов другие типы информации.

Одна из задач, которая решается в ПОЦБ, – это реализация интеграции набора данных в пространство LOD с использованием онтологии предметной области информационных ресурсов, т.е. автоматизированное обнаружение новых наборов данных и, по возможности, установка и поддержка связей с элементами данных из этих наборов данных с уже имеющимися ресурсами в репозитории библиотеки, что обеспечивает одновременно рекомендуемую проектом LOD функциональность в рамках одной системы.

Источники данных подразделяются на два типа: внешние и внутренние. Внешними мы называем те источники, которые интегрированы в LOD и данные которых представлены в RDF и доступны с использованием SPARQL. Для практических целей мы использовали такие известные источники в LOD, как DBpedia, Europeana. Внутренние источники могут представлять собой любой другой тип источника данных, который не интегрирован в LOD. На практике в качестве внутренних источников мы использовали другие библиотеки, которые предоставляли доступ к своим данным по протоколу OAI-PMH.

К основной функциональности системы, реализующей ПОЦБ, относятся:

- функции атрибутного поиска;
- функция выделения неявных связей между ресурсами по их описаниям;

- функция работы с коллекциями;
- создание / просмотр / редактирование / объединение / вложенные коллекции;
- функция отображения онтологии ИД;
- функция детализации, которая обеспечивает преобразование в подзапросы, соответствующие различным ИД;
- функция для выполнения запросов и обработки результатов и предоставления окончательного результата пользователю;
- функция автоматического мониторинга ИД на наличие новых / измененных данных;
- создание словарей, классификаторов, тезаурусов;
- редактирование элементов;
- поддержка («гибкой») классификации ресурсов;
- поддержка настройки уровней доступа к различным ветвям тезауруса.

Исходя из определения источников данных ПОЦБ и перечня функций системы, можно выделить «внутренние» функции, т. е. те, которые оперируют данными в рамках системы и интегрируют данные из «внутренних» источников и фактически определяют обычную семантическую библиотеку. «Внешние» функции обеспечивают подключение и извлечение данных из LOD и позволяют задать тематическое наполнение библиотеки и установить связи, таким образом задавая фактически определение ПОЦБ.

Онтология ПОЦБ разработана в общем виде без привязки к конкретным методам и способам реализации семантических цифровых библиотек. Фактически общая онтология ПОЦБ состоит из двух онтологий:

1) онтология СЦБ, построенная на основе онтологии информационных систем, включающая в себя основные понятия, необходимые для обеспечения основной функциональности библиотеки, такие как ресурс, пользователь, коллекция, словарь, классификатор, запрос, источник и т. д.

2) онтология и тезаурус предметной области, для которой пользователь определяет ее понятия, их тип, структуру, совокупность словарей и классификаторов, которые представляют тезаурус предметной области, обеспечивающий доступ неквалифицированным пользователям, решающих задачи поиска информации, к знаниям предметной области в разных источниках. Эта онтология позволяет:

- выработать и зафиксировать общее понимание области знания;
- представить знания в удобном для обработки автоматизированными подсистемами виде, обеспечить возможность получения и накопления новых знаний, а также представить возможность многократного использования знаний.

Тезаурус же обеспечивает терминологическую поддержку и помогает пользователям сформулировать запрос к системе, в том числе подобрать правильные ключевые слова для описания искомого результата, имеющихся данных и контекстной информации.

Задача автоматизированного поиска релевантных источников данных осложняется тем, что чаще всего информация о связях между ними проставляется в основном на уровне данных с помощью связей *sameAs*, *seeAlso*. Даже простой анализ связей *sameAs*, *seeAlso* на уровне найденных данных позволит выявить эквивалентные классы, ранее не определенные связи между разными источниками или новые источники. Описание связей на уровне схем затем можно использовать при формировании запросов к источникам данных.

До недавнего времени связи между источниками на уровне схем описывались гораздо реже. В последние несколько лет эта задача решается с введением и активным распространением спецификации VOID. Для описания источников RDF-данных, в которых предоставляется информация о связанных источниках данных, VOID-описание содержит информацию об используемых словарях, статистическую информацию о том, сколько ресурсов того или иного типа или значений определенных свойств используются во множестве. При создании словаря VOID была сведена к минимуму необходимость создания новых свойств и классов путем использования существующих словарей. Например, для описания статистической информации используется словарь SCOVO. На основе этой информации можно делать вывод о релевантности источника тому или иному запросу или предметной области.

В рассматриваемой системе VOID-описание набора данных в хранилище генерируется с помощью D2R Server. В сгенерированное описание не попадает информация о подклю-

ченных источниках данных и статистика по имеющимся с ними связям. Для включения этой информации были использованы правила, по которым осуществляется поиск связанных данных. Полученное описание в рамках используемой системы позволяет формировать распределенные запросы к подключенным источникам данных в терминах онтологии, используемой в этой системе. С помощью VOID-описания, запросы из системы транслируются в термины уже источников данных. Также это описание применяется для отображения обобщенного результата поиска.

На рис. 10 представлена общая схема подключения различных источников данных с использованием технологий из стека проекта LOD.

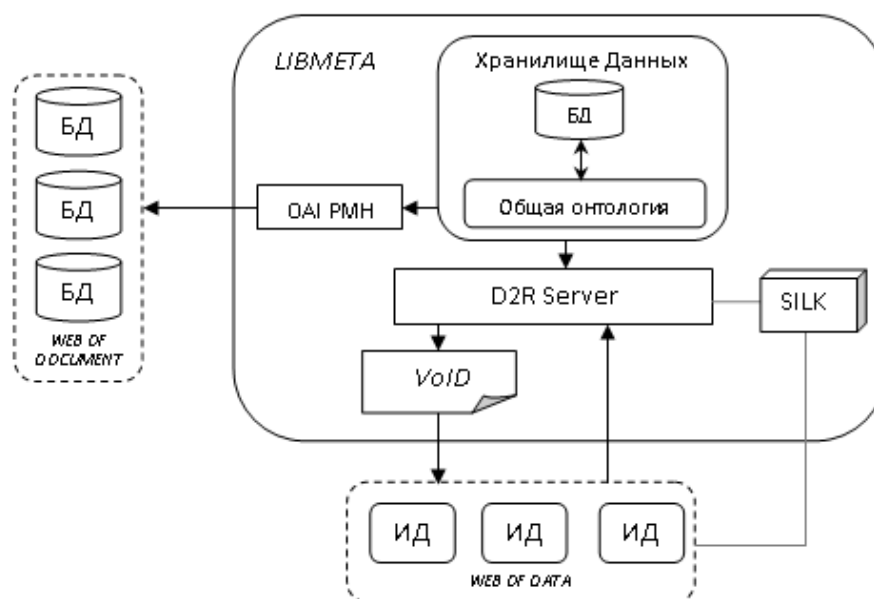


Рис. 10. Схема подключения различных источников данных

Доступ к данным Libmeta осуществляется через ее общую онтологию, которая, как было сказано, состоит из: а) онтологии семантической библиотеки; б) онтологии предметной области, которая задает тематическое направление информационных ресурсов. При этом D2R Server использует онтологию Libmeta для создания SPARQL-точки доступа к ее данным. Используются правила, которые задаются для каждого подключаемого источника (правил может быть несколько), с помощью которых осуществляется поиск и сохранение связей между данными Libmeta и источником из LOD. Для задания правил связывания используется фреймворк SILK. Правила описываются в соответствии с требованиями SILK и хранятся в определенном для каждого источника месте. После описания правила и указания его расположения все действия по запуску и анализу результатов работы SILK выполняются программно, для этого используется соответствующая задаче версия фреймворка.

При каждом подключении нового источника или обновлении набора связей уже подключенных нужно обновлять VOID-описание множества данных Libmeta, анализируя полученный набор ссылок и правила, по которым они выполнялись. Это позволит обновить статистическую и структурную части VOID, необходимых для использования при формировании запросов в терминах общей онтологии и их преобразования в запросы к релевантным источникам в соответствующих им терминах.

Libmeta также исторически поддерживает обмен данными по протоколу OAI-PMH с библиотеками, неинтегрированными в LOD, выступая агрегатором, который интегрирует их данные в LOD.

В рамках создания первой версии ПОЦБ был реализован проект по созданию стандартизированной и децентрализованной среды управления информацией электронных

фондов Libmeta. В проекте реализованы средства интеграции приложений с разными источниками / каталогами метаданных / данных, сервис директорий метаданных, унифицированный интерфейс поиска данных.

Существенное различие во внутренних моделях данных, используемых в различных музеях, библиотеках и архивах, является главной проблемой на пути решения задачи интеграции данных. Для преодоления этой проблемы в решаемой задаче интеграции данных было предложено участникам экспортировать метаданные из своего внутреннего формата в формат на базе Dublin Core с использованием синтаксиса XML, так как во внутренних используемых форматах удается выделить общую часть, которая ложится в рамки предложенного формата. В системе используется универсальный модуль загрузки метаданных в производном XML-формате в соответствии с протоколом OAI-PMH.

Особенности теплофизических данных, методы распространения и интеграции

По многим причинам в теплофизике ключевую роль играет работа с численными данными, включая их накопление, обработку и систематизацию. Повышенное внимание к первичным данным с детальным изучением их достоверности, согласованности, воспроизводимости повторяемых экспериментов связано с ограниченными возможностями теории обеспечить априорное прогнозирование свойств и закономерностей. Как следствие, развитие теплофизики сопровождается нарастающим масштабом производства новых данных, публикуемых в десятках журналов различного профиля: физического, химического, инженерного и др., как, например, *Journal of Chemistry & Engineering Data*, *Journal of Chemistry Thermodynamics*, *Fluid Phase Equilibria*.

Современный этап работ в теплофизике характеризуется повсеместным переходом от печатной формы справочников к компьютерным БД. Соответственно методы обмена неоднородными данными, различающимися форматом и структурой, возникли в теплофизике задолго до того момента, когда проблема интеграции приобрела актуальность для информационного сообщества. Так, один из первых стандартов обмена термодинамическими данными, получивший название COSTAT (Codata STANDARD Thermodynamics), был разработан Термодинамическим исследовательским центром США в течение 1985–1987 гг. под эгидой Международной комиссии по численным данным (CODATA).

Поскольку вещества и их свойства крайне многообразны, имеет смысл сузить онтологию по некоторым критериям с целью отработки методов концептуализации и программных средств. Выбор этих критериев соответствовал тем ограничениям, которые много лет назад (в 1973 г.) были приняты Теплофизическим центром ОИВТ РАН [6] при создании государственного информационного фонда: чистые (однокомпонентные) вещества; преимущественная ориентация на неорганические вещества, включая нестехиометрические соединения; сужение круга органических соединений веществами, содержащими группы не более, чем из двух атомов углерода (простейшими углеводородами, фреонами и т. п.); отказ от рассмотрения материалов, свойства которых зависят от способа получения и метода обработки. Нетрудно видеть, что эти ограничения, помимо сокращения списка веществ, упрощают правила их идентификации. При отказе от рассмотрения материалов, смесей, растворов основным дескриптором является стехиометрическая формула (H_2O , CO_2 , CH_4 ...), дополненная перечнем тривиальных или номенклатурных названий, а отказ от включения в фонды сложной «органики» позволяет обойтись без нотаций, кодирующих структуру и топологию многоатомной молекулы (например, The IUPAC International Chemical Identifier, InChI). Кроме упрощения идентификации, исключение смесей позволяет сократить число независимых переменных, не включая в рассмотрение концентрацию.

При определенных условиях можно пойти еще на одно ограничение, а именно исключить из параметров состояния одну из переменных – давление. Это связано с тем, что во множестве практических задач барическая зависимость свойств проявляется слабо. Прежде всего, это относится к твердой и жидкой фазе, если последняя удалена от критической области.

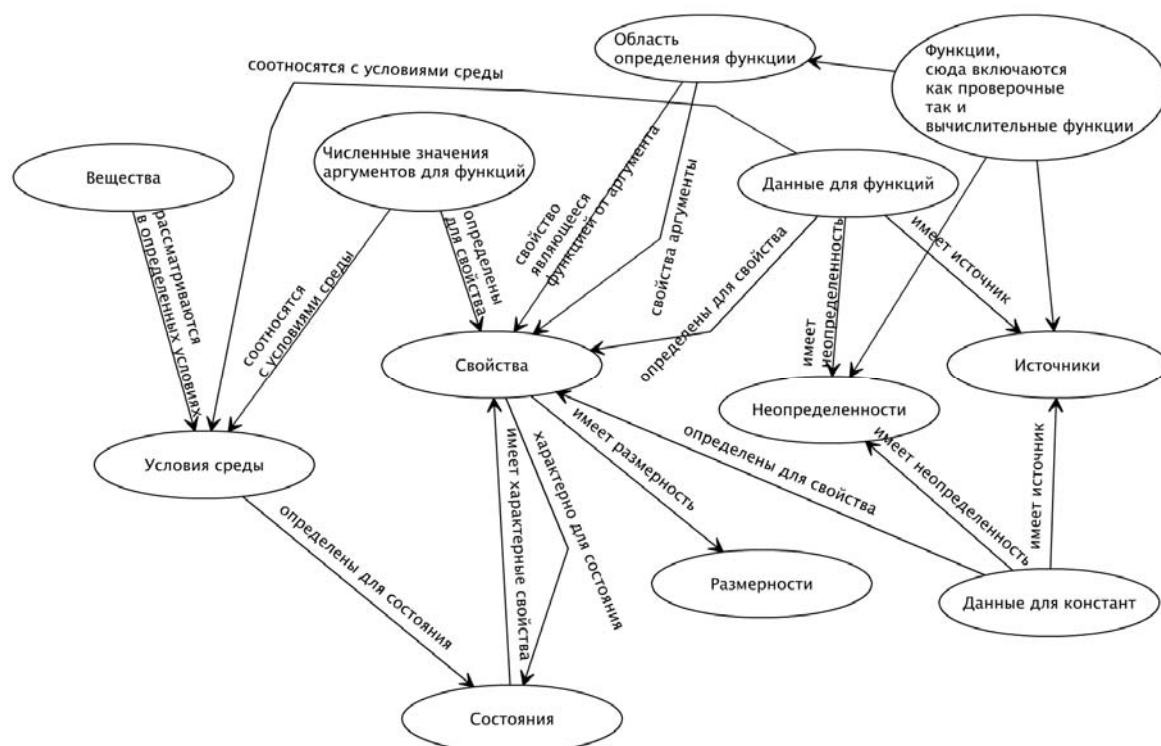


Рис. 11. Классы онтологии и их связи

Что касается газовой фазы, часто для учета неидеальности допустимо ограничиться вириальным разложением по плотности, притом, что сами вириальные коэффициенты являются функциями температуры. Точно так же и транспортные свойства, такие как вязкость или теплопроводность, слабо зависят от плотности или давления. Не предполагая универсальность этого положения, можно принять, что подавляющее большинство публикуемых или компилируемых теплофизических данных сводится к представлению некоторого числа температурных функций, как, например, теплоемкость, энтальпия, вязкость и т. п., а также численных констант, например критических постоянных вещества или энтальпий их образования.

Таким образом, если ввести указанные ограничения при осознанном сужении предметной области, мы приходим к обозримой задаче – интеграции относительно однотипных данных, представимых константами и функциями одной переменной, с доминированием табличной формы передачи данных. При таком сужении предметной области задача интеграции данных делается относительно обозримой, хотя и здесь приходится учитывать крайнее многообразие вариантов, связанное с идентификацией вещества, принимаемой номенклатурой свойств, фазовым многообразием и т. п. В настоящее время ведется разработка системы интеграции данных по теплофизике совместно с Объединенным институтом высоких температур РАН [8].

Построение онтологии должно упростить принятие некоторых дополнительных соглашений о форме представления данных, подлежащих хранению и распространению. Предполагается, что общий поток данных разбивается на порции, называемые наборами данных. Каждый набор содержит данные для одного вещества и произвольного комплекса свойств. Набор свойств включает несколько констант и несколько функций одной переменной, как правило, температуры. Все функции заданы для одних и тех же значений независимой переменной. Набор данных включает также сведения о фазовом состоянии вещества, единицах измерений свойств, неопределенности и источнике данных. Существенно при этом, что основные списки – веществ, свойств, фазовых состояний, единиц измерений и т. д. – считаются

открытыми, что позволит в рамках той же онтологии обеспечить подстройку под новые типы данных, если они удовлетворяют принятым ограничениям.

Вторым этапом построения онтологии является *спецификация* с выделением классов и записью всех связей и отношений на языке OWL. Построенная онтология включает 12 основных классов и 2 класса потомка. Диаграмма на рис. 11 показывает связи классов, после чего приведен перечень классов с указанием назначения и основными атрибутами каждого из классов. Первая четверка классов определяет ключевые понятия для представления набора данных: вещества, свойства, состояния, численные данные. Смысл большей части атрибутов можно понять из их названия. Так, атрибут *InConditoins* в классе *Substances* отсылает к перечню состояний и внешних условий, перечни которых даны в соответствующих классах. Использован также ряд атрибутов для таких понятий, как единицы измерения и неопределенность.

Опишем классы онтологии.

Класс *Substances*: определяет вещества, для которых приводятся данные.

Класс *States*: определяет агрегатные состояния вещества.

Класс *Properties*: определяет свойства вещества.

Класс *NumericalData*: определяет набор численных данных для функциональной зависимости свойств вещества в определенных условиях от аргументов; в данной онтологии свойства зависят от температуры *T* и давления *P*.

Следующая группа классов определяет базовые понятия, сопровождающие физические величины: единицы измерения (размерности), неопределенность данных и внешние условия. К условиям в данной работе относятся и константы, связанные с веществом, например, энтальпия образования при 298.15 К (*DHF(298)*). С точки зрения онтологии это не совсем корректно, но в целях упрощения схемы принято временное решение об отнесении констант к условиям. Соответственно, класс *EnvironmentConditions* содержит указание на агрегатное состояние вещества и ссылок на набор классов *NumericalData*, в которых указываются численные значения свойств с заданной погрешностью и условиями, характерными для данной среды.

Класс *ConstantsOfSubstance*: определяет набор численных значений констант в определенных условиях среды. Хотя классы *ConstantsOfSubstance* и *NumericalData* имеют одинаковую структуру, они кардинально отличаются по содержанию; соответственно с введением этих классов мы разделили содержание *Property* на константы, переменные и функции от этих переменных.

Класс *EnvironmentConditions*: задает перечень свойств, определяющих условия среды, в которых находится вещество, численные значения свойств вычисляются с помощью класса *NumericalData*. Приводимые далее два класса (*Data*, *Data_Source*) вводят данные, например, справочную информацию по молекулярным весам, а также сведения о публикациях, откуда приняты наборы данных.

Класс *Uncertainties*: определяет тип погрешности физической величины.

Класс *Dimensions*: определяет размерности физических величин.

Класс *Data*: определяет перечень данных из справочников физических величин.

Класс *Data_Source*: определяет источники данных для классов *NumericalData*, *Data* и *Functions*.

Наконец, последняя группа состоит из двух основных классов (*Functions*, *DomainOfFunctions*) и двух потомков класса *Functions*: *ControlFunc* и *ComputableFunc*. В совокупности они решают задачу вычисления свойств по формулам при контроле допустимой области изменения аргумента и функций, а также заранее установленных соотношений между различными свойствами, которые в экспериментальных данных выполняются с точностью до некоторой погрешности. Функции делятся на два типа: «вычислительные функции» и «контрольные функции». Вычислительные функции дополнительно содержат указание на вычисляемое свойство, тип и величину погрешности. Результатом вычисления функции является значение свойства, помещаемое в БД. Контрольные функции являются булевскими и отвечают на вопрос, выполняется ли заданное соотношение при допустимой погрешности или нет.

Класс *DomainOfFunctionDefinition*: определяет перечень аргументов и ограничений физических свойств для функций. Атрибут *inStates* определяет, для каких состояний характерна данная зависимость.

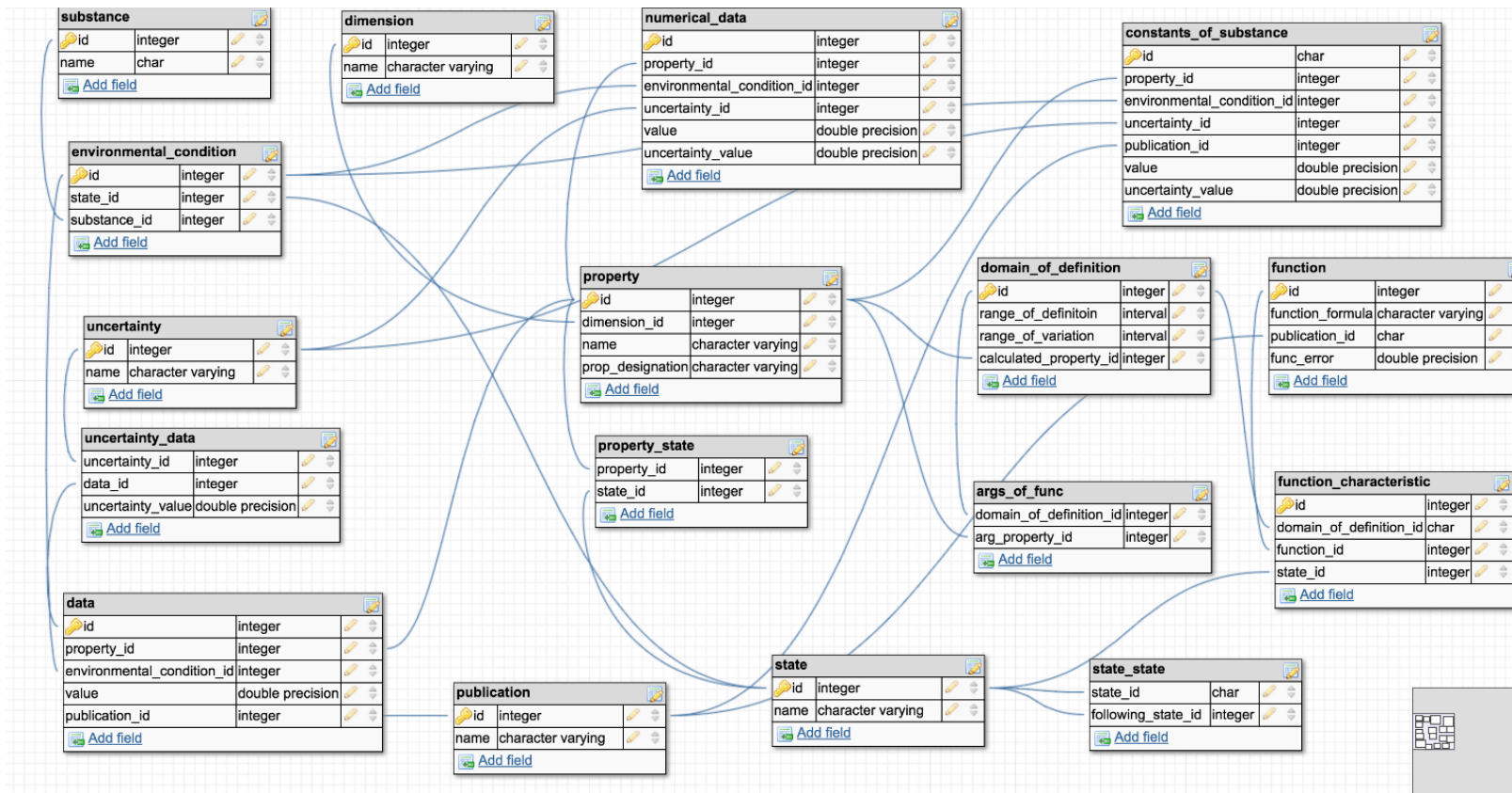


Рис. 12. Схема реляционной БД

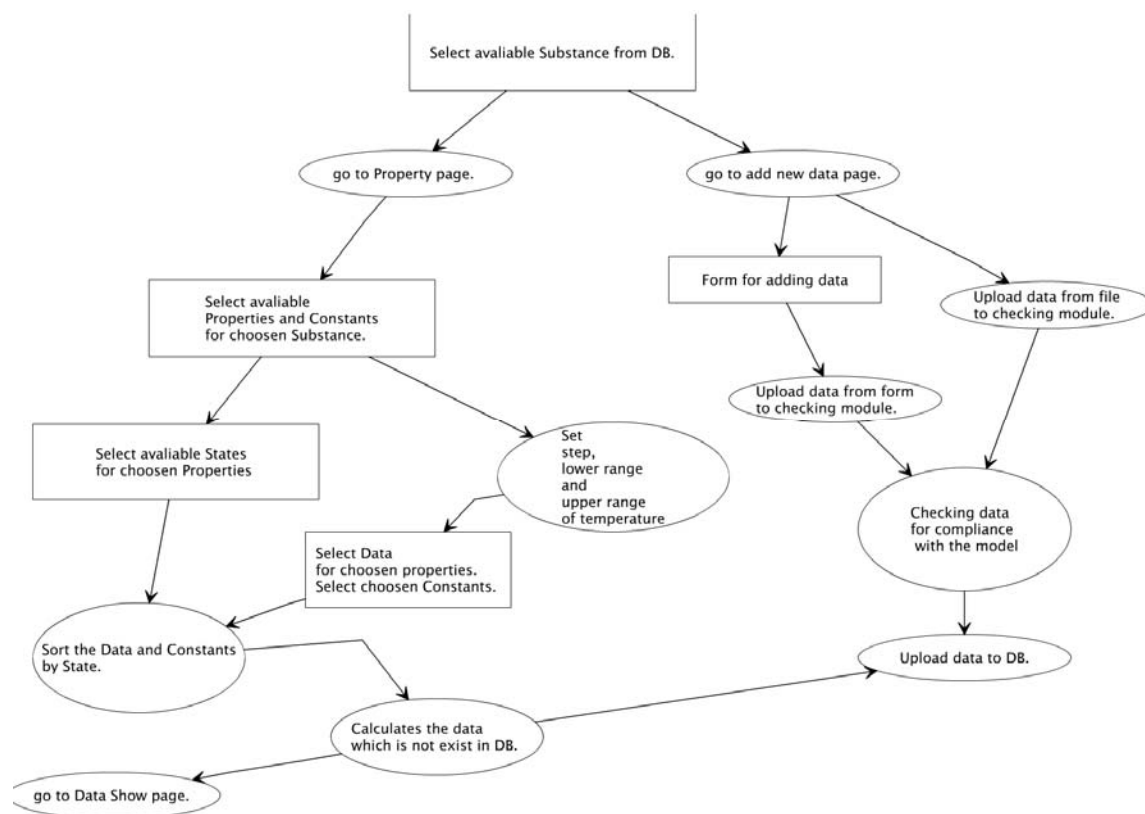


Рис. 13. Логика работы приложения

Класс *Functions*: определяет функции для вычисления и проверки корректности значений физических величин.

Класс *ControlFunc* (подкласс класса *Functions*): определяет перечень проверочных функций, которые определяют, выполняется ли заданное соотношение при допустимой погрешности.

Класс *ComputingFunc* (подкласс класса *Functions*): определяет перечень функций для вычисления значений свойств.

В итоге построенная онтология формализует предметную область до уровня, позволяющего для «суженной» предметной области охватить практически все виды представляемых в литературе или компьютерных средах данных по свойствам как в виде таблиц, так и математических выражений. При этом возможно произвольное расширение на новые виды свойств, единицы измерения, способы задания неопределенности и прочие элементы, сопровождающие набор экспериментальных или справочных данных. Окончательная задача – разработка приложения, обеспечивающего экспорт данных из БД или публикаций в форме, соответствующей разработанной онтологии.

На основании построенной онтологии посредством использования стека технологий *Hibernate ORM* и *Spring MVC* генерируется БД в реляционной СУБД *PostgreSQL*. Соответственно 12 классам онтологии создано 12 java классов, которые отображаются посредством *Hibernate* на таблицы реляционной БД данных, схема которой приведена на рис. 12.

Для системы загрузки данных разработан комплекс, позволяющий анализировать документы и загружать данные из них, причем будет выполняться проверка на соответствие содержания документа онтологической модели и выполнение ограничений, определяемых физическими законами. Логика работы приложения, связанная с вводом исходных данных и проверкой их на соответствие онтологии, показана на рис. 13.

Заключение

Отметим, что за эти годы была проделана значительная работа по созданию информационных систем для обеспечения доступа к информационным ресурсам РАН и интеграции этих ресурсов. К сожалению, так и не была решена задача создания единого информационного пространства РАН, основанного на современном подходе к интеграции данных и приложений, и эта проблема остается в том же положении, что и была в конце 1990-х и начале 2000 гг.

Список литературы

1. *Бездушный А. Н., Жижченко А. Б., Кулагин М. В., Серебряков В. А.* Интегрированная система информационных ресурсов РАН и технология разработки цифровых библиотек // Программирование. 2000. Т. 4.
2. Интегрированная система информационных ресурсов: архитектура, реализация, приложения / Под ред. В. А.Серебрякова. 2004. 240 с.
3. *Бездушный А. А., Бездушный А. Н., Серебряков В. А., Филиппов В. И.* Интеграция метаданных Единого научного информационного пространства РАН. М., 2006. 238 с.
4. *Бездушный А. А., Бездушный А. Н., Нестеренко А. К., Серебряков В. А., Сысоев Т. М., Теймуразов К. Б., Филиппов В. И.* Информационная Web-система «Научный институт» на платформе ЕНИИП. Б., 2007. 257 с.
5. *Каленов Н. Е., Савин Г. И., Серебряков В. А., Сотников А. Н.* Принципы построения и формирования электронной библиотеки «Научное наследие России» // Программные продукты и системы. 2012. № 4. С. 28–31.
6. *Атаева О. М., Кузнецов К. А., Серебряков В. А., Филиппов В. И.* Портал интеграции пространственных данных «ГеоМета». Препринт ВЦ РАН, 2010. 106 с.
7. *Атаева О. М., Серебряков В. А.* Персональная цифровая библиотека Libmeta как среда интеграции связанных открытых данных // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Тр. Всерос. науч. конф. / Сост. Л. А. Калмыкова, М. Р. Когаловский. Дубна, 2014. С. 66–71.
8. *Серебряков В. А., Теймуразов К. Б., Хайруллин Р. И., Еркимбаев А. О., Зицерман В. Ю., Кобзев Г. А., Трахтенгерц М. С.* Система интеграция данных по теплофизике // Инфраструктура научных информационных ресурсов с систем: Сб. науч. ст. / Под ред. Е. Б. Кудашева, В. А.Серебрякова. М., 2014.

Материал поступил в редколлегию 06.10.2014

V. A. Serebryakov

*Dorodnicyn Computing Centre of RAS
40 Vavilov Str. Moscow, 119333, Russian Federation*

serebr@ultimeta.ru

RESEARCH AND DEVELOPMENT IN THE COMPUTING CENTRE OF RAS IN THE FIELD OF DISTRIBUTED INFORMATION SYSTEMS

The article discusses research and development projects implemented in CCAS in the field of distributed information systems supporting research activities. Among them: the portal of the Russian Academy of Sciences, a project of unified information space of the RAS, the digital library of the scientific heritage of Russia, geoportal GeoMeta, semantic personal digital library and a system to integrate data on thermophysics. The problems are considered that arise in the process of implementation of scientific information systems.

Keywords: distributed information systems, research activities, unified information space, digital library, Scientific heritage of Russia, GeoMeta, data integration

References

1. Bezdushnyi N., Zhizhchenko A. B., Kulagin M. V., Serebryakov V. A. Integrated Information Resource System of the Russian Academy of Sciences and a Technology for Developing Digital Libraries. *Programming and Computer Software*, 2000, vol. 26, no. 4, p. 177–185. (in Russ.)
2. Serebryakov V. A. (ed.) *Integrated System of Information Resources: Architecture, Implementation, and Applications*. Moscow, 2004, 204 p. (in Russ.)
3. Bezdushnyi A. N., Bezdushnyi A. A., Serebryakov V. A., Filippov V. I. *Integration of metadata of the Unified Scientific Information Space of the Russian Academy of Sciences*. Moscow, 2006, 238 p. (in Russ.)
4. Bezdushnyi A. N., Bezdushnyi A. A., Nesterenko A. K., Serebryakov V. A., Sysoev T. M., Teimurazov K. B., Filippov V. I. *The Information System «Scientific Institution» on the platform of the Unified Scientific Information Space*. Moscow, 2007, 257 p. (in Russ.)
5. Kalenov N. E., Savin G. I., Serebryakov V. A., Sotnikov A. N. Scientific heritage of Russia digital library: Construction and sources aggregation philosophy. *Programmnye produkty i sistemy (Software & Systems)*, 2012, no. 4, p. 28–31. (in Russ.)
6. Ataeva O.M., Kuznetsov K. A. GeoMeta: Portal for Integration of geospatial data. Computing Centre of the Russian Academy of Sciences, Moscow, 2010, 106 p. (in Russ.)
7. Ataeva O. M., Serebryakov V. A. Personal Digital Library Libmeta as an Integrating Environment for Linked Open Data Digital Libraries. *Proc. of XVI All-Russian Scientific Conference RCDL-2014*. Dubna, 2014, p. 66–71. (in Russ.)
8. Serebryakov V. A., Teimurazov K. B., Khairullin R. I., Erkimbaev A. O., Zitserman V. Y., Kobzev G. A., Trachtenherz M. S. A System for integration thermophysics data. *Infrastructure of scientific data and systems*. Moscow, 2010, 106 p. (in Russ.)