

Концепция создания виртуального музея СО РАН

Федотов А.М., Барахнин В.Б., Бычков И.В., Жижимов О.Л., Мазов Н.А.,
Москвичев В.В.

Институт вычислительных технологий СО РАН (Новосибирск)

Иркутский научный центр СО РАН

ОИГГМ СО РАН (Новосибирск)

Красноярский научный центр СО РАН

1 Введение

За 45 лет существования в Сибирском отделении Российской академии наук создана обширная научно-экспериментальная база, разработано множество оригинальных приборов и методик для научного исследования. Во многих организациях и архивах Отделения, в институтах, в конструкторских бюро, на опытных заводах, у частных лиц имеются уникальные вещественные памятники научно-технической культуры, олицетворяющие важные этапы развития отечественной науки. В качестве примера достаточно привести такие уникальные разработки, как “Виращ-83”, “Камак”, “РАТАН-600”, система автоматизации Сибирского солнечного радиотелескопа — “ССРТ” и многие другие. Сделать информацию об этих памятниках достоянием специалистов и широкой общественности, а также раскрыть реальную картину истории Отделения является нашим моральным долгом перед будущими поколениями.

Несколько лет назад по инициативе Научного совета по музеям СО РАН была начата собирательская работа в научных подразделениях Отделения с целью создания исторического музея технической и научной тематики: приборов и техники научного эксперимента научно-исследовательских Институт академической науки в Сибири. Президиумом СО РАН создана комиссия по инвентаризации предметов музейного значения для детального изучения вопроса возможности создания такого музея. К настоящему времени выявлено большое количество объектов, представляющих значительный познавательный и исторический интерес. В ряде Институт еще не утрачены уникальные приборы и научные установки, способные войти в золотой фонд раритетов науки и техники. Кроме того, необходимой для наполнения музейных коллекций информацией об уникальных вещественных памятниках технической культуры, олицетворяющих важные этапы развития науки в СО РАН, обладают Президиум СО РАН и Президиумы Научных центров (Красноярск, Иркутск, Томск).

Для того чтобы экспонаты и коллекции данного музея, а также другие документы отражающие историю СО РАН, стали доступны максимально широкому кругу лиц, интересующихся историей Сибирской науки, было решено создать электронный (виртуальный) музей истории Отделения с доступом через сеть Интернет.

На выездном заседании Координационного научного Совета СО РАН по целевой программе “Информационно-телекоммуникационные ресурсы СО РАН”, состоявшемся в Иркутске в июле 2002 года, была поддержана идея создания информационной системы “Виртуальный музей науки и техники СО РАН”. Позднее этот проект был поддержан как на конкурсе интеграционных проектов Отделения (проект N 132), так и РФФИ (проект № 03-07-90423).

В электронных коллекциях, составляющих основу Музея, будут представлены описа-

ния основных событий из истории Отделения: основные научно-технические достижения институтов Сибирского отделения, а именно, описания важнейших разработок и результатов, полученных в СО РАН, ключевые события, научные биографии выдающихся ученых и специалистов, внесших наиболее весомый вклад в развитие науки и техники СО РАН. В качестве методологической основы “Виртуального музея” используется методология создания музейных коллекций научно-технической тематики, отработанная в институте Истории СО РАН.

2 Основные принципы создания “Виртуального музея”

Во всем мире научно-технические музеи играют роль символов национальных достижений и торжества национальной технической мощи, подчеркивая особое значение музеев для развития общественного сознания и общества в целом. Среди таких музеев можно выделить Российский политехнический музей в Москве, Немецкий музей в Мюнхене, Чикагский музей науки и индустрии, Лондонский музей науки, Парк науки в Бомбее, а также музеи Японии, Франции, Израиля и других стран. В сети Интернет в настоящее время существует ряд виртуальных музеев истории науки, из разработок стран СНГ следует отметить созданный в Киеве Европейский Виртуальный музей истории компьютерной науки и техники. Авторы проекта надеются, что создаваемый музей займет достойное место в ряду мировых информационных ресурсов.

В процессе создания музея необходимо решить несколько принципиально новых задач:

1. Сформулировать концепцию отбора информационных документов (музейных экспонатов и материалов), а также принципы формирования виртуального музейного фонда.
2. Определить набор интерфейсов доступа к документам музея.
3. Сформулировать требования к лингвистическому процессору системы, определить порядок формирования онтологии музея, словарей, тезаурусов, классификаторов, каталогов и других информационных структур.
4. Сформировать базовые информационные структуры (БИС) для представления документов и музейных материалов (экспонатов).
5. Разработать архитектуру распределенной информационной системы, способной устойчиво и непрерывно функционировать в глобальной сети.
6. Разработать компоненты программного обеспечения для поддержки функционирования распределенной системы.
7. Провести наполнение баз данных документами (музейными материалами).

Основной задачей, поставленной перед авторами проекта, является разработка технологии создания распределенных информационных систем, предназначенных для хранения и отображения разнородной информации, предназначенной для системы представления электронных коллекций научно-технической тематики. Отметим, что электронные документы (экспонаты) Музея носят разноплановый характер: это базы (таблицы) данных и идеографическая информация (описания, книги и статьи), изображения (фотографии)

и пр.) и видео- (аудио-) информация, библиографическая информация, списки данных, модели и модельные данные и др. Эти ресурсы, как правило, принадлежат различным институтам и организациям, которые проводят самостоятельную политику в отношении их использования и публичного доступа к ним. В этом отношении задача построения распределенного виртуального музея, экспозиция которого охватывала бы столь широкий спектр неоднородных коллекций, ставится впервые.

Виртуальный музей как распределенная информационная система должен обладать рядом свойств, обеспечивающих ее устойчивое функционирование.

1. Узлы системы должны обеспечивать автономное функционирование. Выход из строя некоторых узлов системы не должен влиять на работу узла, на котором нет отказов. Все узлы должны рассматриваться как равные. Например, пользователи узла в Томске при обращении к музейным материалам из Томска не должны заметить “ремонтных работ” на узлах Новосибирска или Иркутска.
2. Система должна обладать свойством “прозрачности расположения”. Пользователь не должен знать о физическом месте размещения необходимой ему информации.
3. Архитектура системы должна обеспечивать независимость системы от типов используемых СУБД, операционных систем, аппаратных платформ и сетевых протоколов

При правильном построении распределенной системы информационные ресурсы, включаемые в музейную экспозицию, всех институтов и организаций СО РАН могут быть объединены в единое виртуальное хранилище, что позволит для пользователей открыть всю информацию, накопленную за долгое время и в разных местах, позволит быстро получать исчерпывающие ответы на сложные запросы. При этом фрагменты данных, выбранные по запросу, подвергаются определенным процедурам, выполняемым специализированным программным обеспечением, что позволяет представлять их пользователю в адекватной форме.

Неоднородность хранения информации по разным научным центрам СО РАН качественно отличаются от известных моделей распределенного взаимодействия и требуют создания новых инструментов. Так, данные составляют ценность лишь тогда, когда они соответствующим образом организованы и доступны для всех пользователей информационной системы. Решение проблемы достигается путем объединения распределенных ресурсов для создания виртуальных организаций — единой динамической среды метаданных распределенных ресурсов виртуального Музея.

3 Информационное наполнение “Виртуального музея”

Данный проект осуществляется в рамках работ по созданию “Электронной библиотеки СО РАН” [1]. При создании Музея предполагается использовать последние достижения Институтов СО РАН как в области музейного дела, так и в области разработки распределенных информационных систем, в основе которых лежат разнородные базы данных удаленного доступа.

Структура информационного наполнения “Виртуального музея” строится на основе существующих схем данных — открытых международных стандартов (OSI) и базируется на использовании следующих основных базовых информационных структур (БИС):

- персоны — со связями с БД СО РАН (научные биографии, достижения, результаты и т.д.);
- организации — со связями с БД СО РАН (направления деятельности, проекты, достижения, основные разработки и т.д.);
- артефакты — со связями с организациями и персонами;
- достижения — со связями с организациями и персонами;
- проекты и разработки (разработки институтов и ученых);
- ключевые события;
- документы, в том числе кино, фото и аудио документы;
- каталог публикаций (ссылочный ресурс);
- электронные публикации и электронные копии публикаций;
- страницы — выставочные залы — классификационный ресурс;
- тематические коллекции;
- каталоги, словари и классификаторы, словари — привязка ресурсов к термину.

Предполагается, что при разработке указанных БИС будут использованы имеющиеся наработки: например, первые две БИС содержат ссылки на БД СО РАН, каталог публикаций — на библиографические базы данных ГПНТБ СО РАН, ОИГТМ СО РАН и других институтов, события — на базу данных газеты “Наука в Сибири”.

Важнейшими задачами при создании музея является создание онтологического описания основных объектов (метаобъектов), составляющих информационное наполнение базовых информационных структур (БИС) музея Музея, а именно:

- создание систем поиска информации, обеспечивающих поиск данных по атрибутам,
- выделение метаданных, семантическую интероперабельность,
- общие онтологии и улучшенные системы аннотирования данных,
- поддержка таких базовых технологий обработки информации, как публикация, каталогизация и хранение наборов данных.

Реализация распределенной информационной системы позволит перейти к построению интеллектуальной системы обработки запросов, например в идеологии популярной сегодня технологии GRID, главную роль в которой играет программное обеспечение, реализующее функцию управления моделями данных и метаданных (диспетчера). Полученный от приложения запрос направляется в систему обработки, которая посредством системы поиска информации разыскивает необходимые данные и после выполнения удаленных процедур (например, выделения требуемого подмножества из данного множества) возвращает затребованные данные приложению. Разработанная динамическая система формирования электронных коллекций предоставляет возможности для точного и адекватного удовлетворения потребностей пользователей, формально обращающихся к одной и той же

информации (цифровому объекту в репозитории). Эта методология была использована при создании “Электронного атласа биоразнообразия животного и растительного мира Сибири” [2,3].

Система обмена метаданным будет основываться на сервере метаданных СО РАН, который поддержит следующий набор служб:

- публикация/регистрация новых наборов данных;
- база метаданных для поиска данных по атрибутам;
- доступ к гетерогенным ресурсам посредством брокера ресурсов;
- контроль аутентификации и доступа;
- мониторинг информационных ресурсов и ресурсов ввода/вывода;
- распределенное исполнение служб.

Соответствующие службы могут регистрироваться в объектно-ориентированной базе данных и вызываться для обработки любого набора данных, хранящегося в библиотеке (репозитории). Комбинация возможности доступа к данным через базу метаданных, служб каталогов, зарегистрированного набора методов обработки данных позволяет решать задачи для создания среды обработки данных. Одним из основных результатов работ по проекту является разработка промежуточного ПО, экранирующего сложность различных распределенных гетерогенных ресурсов при сохранении их возможностей и прозрачном позиционировании трех подсистем:

Именованье. Поскольку уникальные имена практически невозможны, используются метаданные. Система поиска информации, выполняя запросы по этим данным, осуществляет поиск в каталоге метаданных.

Нахождение ресурса. Описание расположения хранится в каталоге метаданных. Становится возможным разнесенное хранение данных и метаданных, а в каталоге должны храниться и данные о протоколе для доступа к конкретному набору.

Конверсия протоколов, которая должна осуществлять преобразование протоколов с целью обеспечения интероперабельности.

4 Технологии, используемые при создании “Виртуального музея”

На сегодняшний день самым распространенным способом для реализации подобной системы является WWW сервис Интернет, как наиболее простой и эффективный способ организации сетевого доступа к информационным ресурсам различного характера. В современной технологической среде возможно построение распределенной информационной системы со сквозным поиском и единым для всех подсистем интерфейсами.

В качестве технологической основы “Виртуального музея” предлагается использовать методологию создания электронных коллекций в рамках распределенных информационных систем и электронных коллекций. Данная технология обеспечивает раздельное хранение и презентацию данных, разграничение доступа к информационным ресурсам и к

полнотекстовым базам данных, создаваемых на сервере Сибирского отделения РАН [4,5], и позволяет организовать авторизованный доступ к данным.

Однако, в используемых в настоящее время технологиях, пока отсутствует такая важная компонента, как глобальная стандартизация на уровне организации данных и форматов их представления. В последнее время в мировом сообществе в достаточной степени активизировались работы в этом направлении. Были предложены соответствующие подходы и технологии, в основе которых лежат различные способы работы с метаданными, как например RDF или XML-схема, тем не менее, как известно авторам, пока единственной технологией, содержащей подобную компоненту и апробированной в режиме промышленной эксплуатации, является технология, основанная на международном стандарте ISO-23950 (Z39.50) [6].

К настоящему моменту в СО РАН отработана технология доступа к электронным коллекциям научно-технической библиографической информации с использованием метаданных и схем данных, основанных на открытых международных стандартах и протоколах [9]. В качестве основы технологии для построения распределенных информационных систем (РИС) используется открытый протокол Z39.50 (ISO 23950), который позволяет унифицировать сетевой доступ к любым базам данных, поскольку в своей основе предполагает абстрагирование от конкретных БД и СУБД. Несмотря на то, что этот протокол создан в конце восьмидесятых годов по инициативе библиотечного сообщества и для библиотек, сегодня область его применения существенно расширена. Он эффективно используется для доступа к научно-технической информации, к геоинформационным ресурсам и к глобальным базам метаданных и к другой информации. Существует также возможность предоставления доступа по протоколу Z39.50 к музейной информации и к цифровым коллекциям [7-8]. Технология построения РИС на основе этого протокола позволяет объединить разнородные ресурсы и обеспечить к ним унифицированный доступ.

Правила, которые регламентируют работу с музейной информацией по Z39.50, сформулированы в документах СІМІ (Consortium for the Computer Interchange of Museum Information) и составляют профиль СІМІ (схему данных). Элементы этого профиля имеют глобальные идентификаторы и являются частью международного стандарта ISO-23950. На основе этих объектов определена модель поиска и извлечения данных с музейной информацией.

В России на сегодняшний день существует лишь один программный продукт, позволяющий организовывать РИС с поддержкой профиля СІМІ — это серверное программное обеспечение Z39.50 “ZooPARK”, разработанное в ОИГГМ СО РАН и успешно функционирующее не только в СО РАН, но и в различных регионах России. Сервер “ZooPARK” предназначен для эксплуатации в качестве базового сервера Z39.50 (ISO-23950) в распределенной информационной системе общего назначения на различных программно-аппаратных платформах. Сервер Z39.50 “ZooPARK” построен по модульному принципу и допускает работу с данными различных СУБД, что существенно отличает его от разработок подобного класса. Внутренняя структура сервера такова, что доступ к конкретным базам данных, будь то данные СУБД иерархического или реляционного типа, осуществляется через специальные динамически подгружаемые модули, называемыми провайдерами данных.

Важными аспектами такого подхода являются расширяемость и независимость ядра сервера от типа конкретной базы данных и СУБД. С каждой конкретной СУБД взаимодействует специальный модуль (провайдер данных), причем он и только он становится жестко привязанным к конкретной СУБД. Все детали взаимодействия сервера с конкрет-

ными СУБД скрыты в провайдерах данных, а сервер “ZooPARK” взаимодействует с ними при помощи унифицированного открытого интерфейса.

Наряду с разработкой сервера, разработан ряд специальных провайдеров данных для работы с конкретными СУБД, например такими, как CDS/ISIS, MySQL, MS-SQL и др. Здесь особо следует отметить специальный провайдер для организации удаленного доступа (Z-Remote). Именно при его помощи осуществляется взаимодействие с другими серверами Z39.50 и создает основу для эффективного построения РИС.

Описанный сервер удовлетворяет следующим требованиям:

- поддержка протокола Z39.50-1995 (версия 3);
- работа с различными СУБД;
- переносимость на различные аппаратные платформы.

Таким образом, установленные в различных узлах однотипные серверы “ZooPARK” могут образовывать кластеры с перераспределением поисковых запросов. Серверы Z39.50 других производителей также могут быть включены в подобные кластеры, но с ограниченными возможностями. Системы, построенные по технологии Z39.50, успешно функционируют сегодня в различных городах России. Здесь следует отметить наиболее значимые РИС г. Новосибирска (“Региональная библиотечная система”), СО РАН (“Интегрированная РИС СО РАН”), РИБС г. Москвы (“Корпоративная сеть публичных библиотек Москвы”), РИС LibWeb (“Распределенный каталог LibWeb”), а также системы в ряде других городов России, в составе которых функционирует указанное серверное программное обеспечение.

5 Литература

1. Федотов А.М., Шокин Ю.И. Электронная библиотека Сибирского отделения РАН // Информационное общество, N 2, 2000.
2. Коропачинский И.Ю., Шокин Ю.И., Шумный В.К., Ермаков Н.Б., Колчанов Н.А., Федотов А.М. Электронный атлас "Биоразнообразие животного и растительного мира Сибири". [<http://www.sbras.ru/win/elbib/bio/>].
3. Байков К.С., Коропачинский И.Ю., Шокин Ю.И., Шумный В.К., Ермаков Н.Б., Колчанов Н.А., Федотов А.М. Электронные коллекции и проблемы биоразнообразия // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Вторая Всероссийская научная конференция, Протвино, 26-28 сентября 2000 г.: Сб. докл., Протвино, ГНЦ ИФВЗ, 2000, 58-65, [<http://www.protvino.ru/dl2000/reports/pdf/40.pdf>]
4. Информационный сервер Сибирского отделения РАН. [<http://www.sbras.ru/win/>].
5. Шокин Ю.И., Федотов А.М. Информационная система Сибирского Отделения РАН // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Вторая Всероссийская научная конференция, Протвино, 26-28 сентября 2000 г.: Сб. докл., Протвино, ГНЦ ИФВЗ, 2000, С. 6-15, ISBN 5-88738-029-2.

6. ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Z39.50 Maintenance Agency Official Text for Z39.50 -1995, July 1995.
7. Жижимов О.Л., Мазов Н.А. О доступе к информационным ресурсам по культурному наследию по протоколу Z39.50 // EVA'2000. "Электронная конвергенция: новые технологии в музеях, галереях, библиотеках и архивах", 30 октября - 3 ноября 2000 г.: Матер. конф., М., Центр ПИК Минкультуры РФ, 2000, 08-2-1 - 08-2-2
8. The CIMI Profile Release 1.0H A Z39.50 Profile for Cultural Heritage Information [<http://www.cimi.org/documents/HarmonizedProfile/HarmonProfile1.htm>]
9. Жижимов О.Л., Мазов Н.А. Модель распределенной информационной системы Сибирского Отделения РАН на базе протокола Z39.50. Электронные библиотеки, 1999, т. 2, вып. 2, ISSN 1234-5678.