

**Проблемы интеграции распределенных информационных ресурсов:
на примере виртуального музея СО РАН**

**Integration of Distributed Information Resources:
by the Example of Virtual Museum of SB RAS**

**Проблеми інтеграції розподілених інформаційних ресурсів:
на прикладі віртуального музею СВ РАН**

Мазов Н. А., Жижимов О. Л.

*Объединенный Институт геологии, геофизики и минералогии
Сибирского Отделения РАН, Новосибирск, Россия*

Федотов А. М.

*Институт вычислительных технологий
Сибирского Отделения РАН, Новосибирск, Россия*

Nikolay A. Mazov and Oleg L. Zhizhimov

*United Institute of geology, geophysics and mineralogy, Siberian Branch
of the Russian Academy of Science, Novosibirsk, Russia*

Anatoly M. Fedotov

*Institute of Computational Technologies, Siberian Branch
of the Russian Academy of Science, Novosibirsk, Russia*

Мазов М. О., Жижимов О. Л.

*Об'єднаний Інститут геофізики та мінералогії
Сибірського Відділення РАН, Новосибірськ, Росія*

Федотов А. М.

*Інститут обчислювальних технологій
Сибірського Відділення РАН, Новосибірськ, Росія*

Рассматриваются проблемы интеграции информационных ресурсов имеющихся в Сибирском отделении РАН в рамках интеграционного проекта «Виртуальный музей СО РАН» на основе применения протокола Z39.50 (ISO-23950) для построения распределенной информационной системы. Утверждается, что технологии Z39.50 являются наиболее актуальными для организации доступа к базам данных в распределенных информационных системах. Рассмотрены цели, задачи, принципы и технологии построения Виртуального музея.

The report considers the integration of information resources in Siberian branch of the RAS in the framework of the integration project «Virtual museum of SB RAS» with the application of protocol Z39.50 (ISO-23950) for the construction of a distributed information system. It is stated that Z39.50 is the most appropriate for the organization of access to databases in the distributed information systems. Objectives, tasks, principles and construction technologies of the Virtual museum are considered.

Розглянуто проблеми інтеграції інформаційних ресурсів, які є в Сибірському відділенні РАН у межах інтеграційного проекту «Віртуальний музей СВ РАН», на основі застосування протоколу Z39.50 (ISO-23950) для створення розподіленої інформаційної системи. Автори доповіді стверджують, що технології Z39.50 є найбільш актуальними для організації доступу до баз даних у розподілених інформаційних системах. Також обговорено завдання, цілі, принципи та технології створення Віртуального музею.

Внедрение новых сетевых информационных технологий в настоящее время кардинально меняет работу информационных органов, библиотек, архивов, музеев. Предоставление возможности удаленного доступа к информационным ресурсам и использования локальных сетей при работе с библиографическими базами данных, с музейными электронными коллекциями и другими информацион-

ными ресурсами, возможность интеграции ресурсов на основе использования протокола Z39.50 выводят работы, связанные с информационным обслуживанием на другой, более высокий уровень.

Опыт работы в Сибирском отделении по данному направлению в интеграционном проекте «Виртуальный музей СО РАН» позволил сделать некоторое обобщение и попытку систематизировать используемые информационные ресурсы, предоставляемые пользователям или используемые в СО РАН для удовлетворения информационных потребностей пользователей, основу которых составляют научные сотрудники и СО РАН, студенты Новосибирского государственного университета.

За все годы существования Сибирского отделения РАН в нем создана обширная научно-экспериментальная база, разработано множество оригинальных методик и приборов для научного исследования. Во многих организациях, в институтах, в конструкторских бюро, на опытных заводах и у частных лиц имеются уникальные вещественные памятники нашей технической культуры, олицетворяющие важные этапы развития отечественной науки. Сделать эти памятники достоянием специалистов и широкой общественности является нашим моральным долгом перед историей и перед молодым поколением ученых, что явилось обоснованием для необходимости проведения таких работ.

Несмотря на то, что в СО РАН ведутся широким фронтом работы в области интеграции информационных ресурсов и интерфейсов, в настоящее время все же сохраняется сложность при получении в короткое время полной и систематизированной информации в соответствии со своими информационными потребностями и в едином интерфейсе, затрудняет широкое использование информационных всех имеющихся ресурсов СО РАН самими пользователями из-за разнообразия наполнения и предоставления информации. Предполагается, что в рамках работ по интеграции ресурсов произойдет дальнейшее развитие имеющихся и становление новых тематических информационных центров в Сибирском отделении РАН, что позволит использовать весь спектр распределенных БД для еще более полного удовлетворения информационных потребностей пользователей.

Для решения указанной проблемы, несколько лет назад по инициативе Научного совета по музеям СО РАН была начата собирательская работа в научных подразделениях Новосибирского научного центра СО РАН с целью создания музея историко-технической и научной тематики. Для того чтобы экспонаты музея стали доступны максимально широкому кругу лиц, интересующихся историей Сибирской науки, было решено отобразить музейные коллекции в электронном виде и разместить их в сети Интернет. На выездном заседании Координационного научного Совета СО РАН по целевой программе «Информационно-телекоммуникационные ресурсы СО РАН», состоявшемся в Иркутске 28-30 июля 2002 года, была поддержана идея создания информационной системы «Виртуальный музей науки и техники СО РАН».

Предполагается, что в электронных коллекциях, составляющих основу Музея, будут представлены 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 и т. д.

Создание подобного музея особенно актуально еще и потому, что во всем мире научно-технические музеи играют роль символов национальных достижений и торжества национальной технической мощи, подчеркивая особое значение музеев для развития общества. В настоящее время в сети Интернет существует ряд виртуальных музеев истории науки, созданных на основе WWW-технологий.

Однако, задача построения распределенного виртуального музея истории науки и техники, экспозиция которых охватывала бы столь широкий спектр неоднородных коллекций, ставится впервые.

Основной целью при решении этой задачи ставится создание открытой распределенной информационной системы, интегрирующей информационные ресурсы, связанные с электронными коллекциями и базами данных, и предоставление к ним публичного доступа из Интернет по стандартным протоколам через удобные пользовательские интерфейсы.

Разработка технологии создания распределенных информационных систем, предназначенных для хранения и отображения информации различного характера и создание специализированной распределенной информационной системы для представления электронных коллекций научно-технической тематики является фундаментальной научной проблемой. Это обусловлено тем, что электронные музейные экспонаты носят самый разнородный характер: это различные базы данных

и идеографическая (описательная) информация; изображения (фотографии, рисунки, чертежи и пр.) и видео- (аудио-) фрагменты; наблюдения и 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.

При правильном построении распределенной системы, все эти ресурсы могут быть объединены в единое виртуальное информационное пространство, 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 отличается 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 идет путем объединения распределенных ресурсов для создания виртуальных организаций — единой динамической среды метаданных распределенных ресурсов виртуального музея.

На сегодняшний день самым распространенным способом для реализации подобной системы является WEB, как наиболее простой и эффективный способ организации сетевого доступа к информационным ресурсам различного характера. Однако следует отметить, что технология WEB не всегда позволяет построить распределенную информационную систему со сквозным поиском и едиными для всех систем интерфейсами. В WEB отсутствует важная компонента — глобальная стандартизация на уровне организации данных и форматов их представления. Насколько известно авторам, единственной технологией, содержащей подобную компоненту и апробированной в режиме промышленной эксплуатации, является технология, основанная на международном стандарте ISO-23950 (Z39.50) [1].

При создании Музея предполагается использовать последние достижения Институты СО РАН как в области музейного дела, так и в области разработки распределенных информационных систем, в основе которых лежат разнородные базы данных удаленного доступа [2-5].

В качестве методологической основы «Виртуального музея» используется методология создания музейных коллекций научно-технической тематики, отработанная в институте Истории СО РАН. Как уже отмечалось ранее, по инициативе Научного совета по музеям СО РАН в ННЦ была начата работа по созданию Музея приборов и техники научного эксперимента научно-исследовательских Институты 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 Институты 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).

К настоящему моменту, наряду с технологией, разработанной в ИВТ, в СО РАН отработана технология создания электронных коллекций научно-технической библиографической информации с использованием метаданных и схем данных, основанных на открытых международных стандартах и протоколах [4]. В качестве основы при использовании этой технологии для построения распределенных информационных систем 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, к глобальным базам метаданных, модификации (CIP) этого протокола используются при мониторинге земной поверхности и др. Существует также возможность предоставления доступа по протоколу Z39.50 к музейной информации и цифровым коллекциям [5-6]. Технология построения распределенных

систем на основе этого протокола действительно 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 [7]. 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.

Полное описание профиля СІМІ достаточно объемно и здесь приводиться не будет. Ниже будут отмечены только некоторые аспекты, которые на наш взгляд характерны для этого профиля. Во-первых, все серверы СІМІ обязаны поддерживать запросы типа 1 (RPN) (запросы в обратной польской нотации) с набором поисковых атрибутов СІМІ-1, который содержит как атрибуты общего плана (Title, Author, Abstract и т. п.), так и атрибуты специфические для музейных коллекций (repositoryPlace, dateOfOrigin, placeOfOrigin, dateCollected и т. п.). Кроме этого набор атрибутов СІМІ содержит в себе поисковые атрибуты Dublin Core. Во-вторых, схема данных и абстрактная структура записи СІМІ основана на иерархической модели с использованием схем Digital Collections и Dublin Core. Элементы последней явно присутствуют в СІМІ на первом уровне вложения. Собственно структура записи СІМІ вложена в элемент actualDO записи Digital Collection (см. табл.1). Эта схема данных позволяет представить в стандартном виде описательную информацию о музейных ресурсах, т. е. является схемой метаданных. Связь с изображениями и полными текстами осуществляется через URL в элементе mrObject/rendition/resource.

В России на сегодняшний день существует лишь один программный продукт с поддержкой профиля СІМІ и позволяющий организовывать распределенные информационные системы — серверное программное обеспечение Z39.50 ZooPARK, разработанный в ОИГГМ СО РАН и успешно функционирующий 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 (БД), 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 конкретными СУБД скрыты в провайдерах данных, а сервер ZooPARK взаимодействует с ними при помощи унифицированного открытого интерфейса.

Структура информационного наполнения «Виртуального музея» строится на основе существующих схем данных — открытых международных стандартов и базируется на использовании следующих основных базовых информационных структур:

- персоны — со связями с различными БД СО РАН (научные биографии, достижения, результаты и т. д.);
- организации — со связями с различными БД СО РАН (направления деятельности, достижения и т. д.);
- артефакты — со связями с организациями и персонами;
- достижения — со связями с организациями и персонами;
- разработки (разработки Институты и ученых);
- события;
- каталог публикаций (ссылочный ресурс);
- электронные копии публикаций;
- каталоги, словари, тезаурусы и классификаторы.

ВЛОЖЕНИЕ СХЕМ DUBLIN CORE И CIMI В СХЕМУ DIGITAL COLLECTIONS

localControlNumber, schemaldentifier,
title, creator, contributor, date, description, identifier, type, language, subject, publisher, format, source, relation, coverage, rights,
typeOfDescriptiveRecord, objectInfo typeOfObject, categoryOfObject, digitalObject actualDO
schemaldentifier, objectName, objectTitle, bibliographicTitle, creatorGeneral, creatorInfo name, dateOfBirth, dateOfDeath, nationalityCultureRace, role fieldCollector, repositoryName, repositoryPlace, owner, creditLine, subject, publisher, objectID, materialMedium, processTechnique, dimensions, placeOfOrigin, dateOfOrigin, dateCollected, agePeriod, typeSpecimen, stylePeriod, periodName, provenance, quantity, award, collection, inscriptionMark, objectLanguage, condition, physicalDescription, protectionStatus, protectionDate, spatialReferencingSystem, x-coordinateInSpatialReferencingSystem, y-coordinateInSpatialReferencingSystem, address, relatedObjects, relatedTextualReferences, associationGeneral, association name, place, event, activity, description contentGeneral, content name, place, event, activity, description contextHistorical, contextArchaeological, copyrightRestriction, wallTextLabel, displayObject, mrObject title, creator, contributor, date, description, type, language, subject, publisher, format, source, relation, coverage, rights, rendition resource, title, creator, contributor, date, description, type, language, subject, publisher, format, source, relation, coverage, rights administrativeEventGeneral, administrator, administrativeEvent administrativeEventType, date, creator, role

Предполагается, что при разработке указанных структур будут использованы имеющиеся наработки — первые две содержат ссылки на БД СО РАН, каталог публикаций — на базы данных ОИГТМ СО РАН, ГПНТБ СО РАН, ИК СО РАН и БД газеты «Наука в Сибири».

Важнейшими задачами при создании Музея является создание онтологического описания основных объектов (метаобъектов), составляющих основное информационное наполнение базовых информационных структур Музея. 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. 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.

Предполагается, что информационные ресурсы, создаваемые в рамках Проекта, будут доступны для пользователей по стандартным протоколам Интернет (HTTP, Z39.50, LDAP и др.). Правила, регламентирующие ограничение доступа к каждому информационному ресурсу, формулируются администраторами соответствующих ресурсов и являются неотъемлемой частью общего описания информационного ресурса. Общее описание конкретного информационного ресурса создается администратором соответствующего ресурса и должно быть доступно по протоколам HTTP и Z39.50 (Explain).

Литература

1. 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.
2. Федотов А. М., Шокин Ю. И. Электронная библиотека Сибирского отделения РАН // Информационное общество, N 2, 2000.
3. 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.
4. 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.
5. 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.
6. Жижимов О. Л., Мазов Н. А. Тезаурусы и классификационные схемы в распределенных информационных системах: проблемы и решения // 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, 08-9-1 — 08-9-3.
7. The CIMI Profile Release 1.0H A Z39.50 Profile for Cultural Heritage Information <http://www.cimi.org/documents/HarmonizedProfile/HarmonProfile1.htm>