

## ВОПРОСЫ ИНТЕГРАЦИИ БИБЛИОТЕК, АРХИВОВ И МУЗЕЕВ ПО НАУЧНОМУ НАСЛЕДИЮ\*

Антопольский А.Б.

(Межведомственный суперкомпьютерный центр РАН)

При развитии ЭБ «Научное наследие России» в качестве одной из центральных поставлена задача интеграции ресурсов институтов памяти (библиотек, архивов и музеев), содержащих объекты научного наследия. Эта задача сформулирована в Соглашении Президентов государственных академий наук, посвященном сотрудничеству по созданию ЭБ «Научное наследие России» [1]

Цель настоящей публикации — рассмотреть существующий опыт в данной области, сформулировать основные проблемы и наметить пути их решения.

### Цель интеграции.

Несмотря на то, что как в России, так и за рубежом эта проблема рассматривается как весьма актуальная уже в течение 10-15 лет, до сих пор высказываются сомнения в целесообразности создания таких интегральных систем. Поэтому кратко сформулируем основные цели такой деятельности. Главная цель — **обеспечить навигацию пользователя и сквозной поиск объектов научного и, шире, культурного наследия в фондах различных институтов памяти**. Причина — существенное пересечение хранения и обработки объектов культурного наследия в институтах памяти. Сведения, или артефакты (в том числе документы), относящиеся к персоне, событию, учреждению, территории или другой сущности можно найти в ресурсах разных институтов памяти.

### Препятствия интеграции.

Главным препятствием к интеграции является различие в принципах, целях и задачах сбора и хранения объ-

---

\* Работа поддержана грантом РФФИ № 11-07-00331-а

ектов научного наследия в разных институтах памяти, что выражается в различных способах организации фондов и, как следствие, в принципах учета и каталогизации этих объектов. В свою очередь, эти принципы реализуются в виде стандартов метаданных, существенно различающихся для библиотек, архивов и музеев. Ниже эти стандарты будут рассмотрены более подробно. Другими препятствиями является ведомственная разобщенность, отсутствие систем управления и финансирования для комплексных проектов такого рода.

### **Международный опыт интеграции.**

В европейских странах активно ведутся работы по интеграции информационных ресурсов институтов памяти на основании информационных технологий. Международным центром, который объединяет усилия в данном направлении, является проект ATHENA, тесно связанный с созданием электронной библиотеки EUROPEANA. Европейский проект ATHENA2, инициированный партнёрами проектов серии MINERVA3, реализуется в рамках программы eContentplus в 2008 — 2011 гг. Проект ATHENA охватывает 20 стран-членов ЕС + 3 обзорателей из стран, не входящих в Европейский Союз (включая Россию), более 100 музеев и других учреждений культуры и 20 европейских языков. Координирует проект Министерство культурного наследия Италии. Цели проекта ATHENA сформулированы так:

- усилить, поддерживать и стимулировать участие в Европейской цифровой библиотеке Europeana тех музеев и других организаций и учреждений, хранящих культурное наследие, которые ещё не полностью вовлечены в эту деятельность;
- разработать набор инструментов, рекомендаций и методических материалов по вопросам многоязычия и семантики, метаданных и тезаурусов, структур данных и прав на интеллектуальную собственность, которые могут быть использованы музеями как для внутримузеевской деятельности по оцифровке, так и для интеграции музейных коллекций в Европейскую цифровую библиотеку Europeana;

- идентифицировать цифровые информационные ресурсы европейских музеев;
- содействовать интеграции информационных ресурсов различных секторов культурного наследия, сотрудничая с другими проектами, непосредственно занимающимися библиотеками и архивами, с целью включения разнообразного контента в Европейскую цифровую библиотеку Europeana;
- разработать техническую инфраструктуру, которая обеспечит семантическую совместимость с Европейской цифровой библиотекой Europeana.

Конечной целью проекта ATHENA является объединение всех заинтересованных сторон и держателей информационных ресурсов Европы, оценка и интеграция стандартов и инструментов, необходимых для обогащения и расширения контента Европейской цифровой библиотеки Europeana, и в результате предоставление пользователю доступа к богатому и разнообразному культурному наследию всей Европы.

ATHENA собирает и объединяет (агрегирует) информационные ресурсы европейских музеев, библиотек, архивов и передает их в EUROPEANA для публикации. К началу 2011 года EUROPEANA уже предоставила открытый доступ к 14 миллионам информационных ресурсов по культуре. 1 757 029 информационных ресурсов поступило в EUROPEANA через проект ATHENA, который стал крупнейшим агрегатором информационных ресурсов для EUROPEANA.

Отчеты о результатах исследований по этому проекту опубликованы и доступны в Интернете [2].

### **Российский опыт.**

Главным российским полигоном для отработки моделей и технологий интеграции должна была стать Президентская библиотека им. Б.Н. Ельцина. Задача создания на базе этой структуры единого информационного пространства по российской государственности и русскому языку на базе библиотек, архивов и музеев была сформулирована в основополагающих документах. К этому стоит добавить уникальные финансовые и административные возможности этой структуры и её особое надве-

домственное положение в иерархии власти. Однако пока Президентская библиотека не предложила научно-информационной общественности внятной идеологии создания единого информационного пространства. Правда, существует определенная тенденция развития Президентской библиотеки в нужном направлении. С одной стороны, это создание региональных филиалов, которые поставлены перед практической необходимостью формирования единого культурно-информационного пространства региона, с обязательным участием ресурсов библиотек, архивов и музеев. С другой стороны, руководители Президентской библиотеки как будто бы осознали ошибки первого этапа и готовы искать пути решения этих проблем.

Очень полезной деятельностью в данном направлении заняты сотрудники центра ПИК Министерства культуры, которые аффилированы в европейский проект по оцифровке культурного наследия MINERVA и MINERVA PLUS (MInisterial NETwoRk for Valorising Activities in digitisation — Сетевое взаимодействие Министерств культуры стран-членов ЕС для координации, усиления и оценки деятельности по оцифровке культурного и научного наследия). С 2004 г. Российская Федерация полноправно участвует в проекте MINERVA PLUS. По поручению Минкультуры, непосредственным участником и координатором проекта MINERVA PLUS от России является Центр ПИК. На сайте проекта MINERVA PLUS [3] публикуются отчеты, рекомендации, стандарты и другие документы европейского и российского происхождения, связанные с оцифровкой культурного наследия, а также осуществляется опыт российского участия в данных проектах.

Можно упомянуть также несколько конкретных российских проектов, связанных с интеграцией библиотек, архивов и музеев.

В этом направлении активно работает известная петербургская фирма «Альтсофт», чьи проекты электронных энциклопедий региона достаточно широко распространены. Последний по времени проект этой фирмы «Виртуальный музей ГУЛАГа» [4] достаточно успешно организует навигацию по информационному пространст-

ву, в которое включены объекты архивного, музейного и библиотечного хранения.

Интерес представляет также проект, реализованный в Ярославском государственном университете им. Демидова под руководством В. Смирнова [5]. В этом проекте построена уникальная система навигации, использующая реальные электронные каталоги ряда библиотек, музеев и архивов региона.

Начат выпуск электронного журнала «Цифровое наследие» и организована поддержка соответствующего сайта [6], инициированные известным специалистом в данной области С.Т. Петровым.

Можно упомянуть также совместную с Т.С. Маркаровой работу автора [7], в которой была поставлена задача формирования единого каталога для библиотеки и архива педагогической информации на базе системы ИРБИС.

### **Архитектура интегрированных систем**

Известны следующие основные типы архитектуры интегрированных систем:

1) централизованное хранение и документов и метаданных. Примеры — Всемирная цифровая библиотека, Президентская библиотека им. Б.Н. Ельцина;

2) распределенное хранение документов и централизованное хранение метаданных. Примеры — EUROPEANA, открытые архивы, использующие протокол OAI-PMH;

3) распределенное хранение и документов и метаданных. Примеры — АРБИКОН и другие корпоративные системы, использующие протокол Z.39.50.

Можно, конечно, дискутировать о преимуществах каждого из типов архитектуры, но лично я убежден, что по соотношению затраты/эффективность значительное преимущество имеет второй тип архитектуры. Пример достижений системы EUROPEANA является достаточно убедительным. Именно этот тип я предлагаю считать перспективным для развития библиотеки «Научное наследие России», тем более, если ставится задача ее расширения в направлении культурного и духовного наследия России. Чем более широкий круг организаций и типов ре-

сурсов охватывает интегрированная система, тем более эффективным является второй тип архитектуры.

### **Функциональность распределенной системы**

Одной из наиболее обсуждаемых проблем является вопрос о составе и распределении функций участников интегрированной системы. Очевидно, что это распределение будет различно для разных типов архитектуры

Приведем примерный список возможных функций, о которых можно дискутировать — имеет ли смысл централизовать выполнение этих функций в интегрированной системе. Этот список представляет слегка модифицированный перечень функций, приведенный в основополагающих документах по ЭБ «Научное наследие»

- *Доступ к ресурсам* — обеспечение возможности навигации и поиска информации через Интернет, определение местоположения, извлечение, трансформация и визуализация ресурса включая:
- Навигацию по иерархической классификации;
- Нормированный поиск по атрибутам информационных объектов;
- Лексический (ненормированный) поиск по наименованиям и другим текстовым реквизитам метаданных;
- Лексический поиск по полным текстам документов;
- *Управление ресурсом* — диспетчеризация предложения по вводу нового ресурса, создание нового ресурса, внесение его в БД, учет, удаление старого ресурса и изменение существующего, включая контроль дублирования при отборе и поиске документов при оцифровке и обработке документов (распознавание, коррекция);
- *Каталогизация (формирование метаданных)*, включая индексирование, классификацию, фиксацию DRM;
- *Управление метаданными и словарями*, их создание, обработка преобразование;
- *Управление участниками* — их регистрация, подписка, права доступа и персональная информация;
- *Управление электронной библиотекой* — управление коллекциями, группами пользователей, член-

ством, общее управление политикой, качеством и функциональностью;

- Обеспечение информационной безопасности;
- *Системное администрирование* — установка, конфигурирование, необходимые периодические мероприятия, восстановление после сбоев и мониторинг библиотеки;
- Ведение каталога ссылок на смежные российские проекты;
- *Мониторинг использования ЭБ* — сбор статистики, ведение гостевых книг и форумов.

Достаточно очевидно, что применительно к интегрированной распределенной системе с архитектурой второго типа, включающей разнородные ресурсы, централизовать разумно лишь часть этих функций.

### **Системы метаданных [8]**

Как мы уже отмечали, различие в системах метаданных, отражающее различие в менталитете библиотечных, музейных и архивных работников, является основной проблемой интеграции упомянутых информационных систем. Кратко охарактеризуем состояние проблемы с метаданными в рассматриваемой сфере.

Наиболее универсальным и известным стандартом метаданных является, безусловно, стандарт Дублинского ядра метаданных. Имеется значительный опыт его применения в различных информационных системах. Однако в реальных институтах памяти этот стандарт выглядит пока достаточно чужеродным телом и применяются другие системы метаданных.

Наиболее просто обстоит дело в библиотечной сфере. В АБИС, как в России, так и за рубежом, доминируют системы метаданных, основанные на форматах семейства MARC (различия между версиями MARC представляются несущественными). При этом большинство специалистов признают этот формат избыточным для электронного информационного пространства. В результате возникли проекты систем метаданных METS • Metadata Encoding and Transmission Standard и MODS • Metadata Object Description Schema, в которых формат MARC адаптируется к современным требованиям информационных систем.

Эти стандарты рассматриваются как промежуточные, в том числе при переходе к Дублинскому ядру.

В архивной области ситуация следующая. За рубежом широко распространены стандарты метаданных

- **EAD** — Encoded Archival Description;
- **ISAD(G)**- General International Standard Archival Description, Second Edition;
- **ISAAR (CPF)** International standard archival authority record for corporate bodies, persons and families, Second Edition.

Однако в России пока практически в информационных архивных системах эти стандарты почти не применяются, хотя из наиболее известной архивной системы КАИСА имеется конвертор в EAD.

В музейной сфере дискуссия о преимуществах стандартов метаданных еще не окончена. Цитированная выше рекомендация Европейского союза приводит несколько музейных стандартов метаданных, наиболее популярных в мире.

- **CDWA** Categories for the of Works of Art (Категории для произведений искусства) 25, разработка Getty Research Institute. Описывает информационные ресурсы баз данных по изобразительному искусству, предлагает концептуальный подход к описанию и доступу к произведениям искусства и изображениям. Может служить основой для взаимодействия существующих информационных систем и для разработки новых систем.
- **MUSEUMDAT** Fachgruppe Dokumentation im Deutschen Museumsbund Institut für Museumsforschung SMB-PK Zuse-Institut Berlin. Формат для агрегирования (сбора) данных (XML схема), оптимизированный для поиска и публикации на музейных порталах, предназначен для автоматического сбора ключевых данных.
- **OBJECT ID** Стандарт для описания объектов культуры. Разработан в результате сотрудничества музейного сообщества, правоохранительных органов, таможни, коммерческих структур, страховых



агентств и оценщиков произведений искусства и древностей.

- **SPECTRUM** Разработчик — The UK Museum Documentation Standard, 3rd Edition (Version 3.1) (Стандарт музейной документации Соединенного Королевства, версия 3.1). Стандарт для музейной документации и управления коллекциями.

В этих рекомендациях также приводится перечень из 3 стандартов, ориентированных на культурное наследие в целом

- **CIDOC-CRM** Conceptual Reference Model<sup>41</sup> Концептуальная образцовая модель CIDOC, которая обеспечивает расширяемую онтологию для концепций и информации по культурному наследию и музейной документации.
- **VRA** Разработчик — Visual Resources Association's Data Standards Committee (Комитет стандартов данных Ассоциации визуальных ресурсов) Набор элементов метаданных (единицы информации, такие как название, местонахождение, дата и пр.). Набор элементов дает возможность организации категорий описания произведений визуальной культуры, а также изображений.
- **LIDO** Lightweight Information Describing Objects (облегченный стандарт информации для описания объектов), разработчик — Европейская Комиссия ATHENA<sup>44</sup>. Описание и агрегирование ресурсов музеев, библиотек, архивов; совмещение схем метаданных в сфере культуры. LIDO был обнародован и одобрен на ежегодном заседании CIDOC, которое состоялось на конференции международного комитета музеев ICOM в ноябре 2010 года.

LIDO— это совместная инициатива создателей форматов CDWA Lite и museumdat, в которой сделана попытка объединить эти форматы и обеспечить совместимость с CIDOC-CRM и SPECTRUM. LIDO хорошо зарекомендовал себя в проекте ATHENA и позволил агрегировать не только музейные, но и библиотечные, и архивные информационные ресурсы. Из цитированных рекомендаций следует, что Европейская комиссия поддерживает проект

LIDO, считая его наиболее актуальным и универсальным. Очевидно, что проект LIDO должен быть тщательно проанализирован, поскольку это последнее достижение европейцев по проблемам интеграции метаданных библиотек, архивов и музеев.

Конечно, нужно иметь в виду, что при проектировании реальной интегрированной системы нужно ориентироваться не только и не столько на стандарты метаданных, сколько на метаданные, реально присутствующие в интегрируемых системах.

### **Общая логическая модель интегрированных систем**

Центральным элементом логической структуры интегрированной системы является перечень сущностей, к которым могут быть отнесены информационные объекты, циркулирующие в системе. Сущности различаются составом признаков (а следовательно, метаданных), которые могут присутствовать в описаниях объектов. Именно на основе этих сущностей может быть построена система навигации и атрибутного поиска. В цитируемых рекомендациях приводится следующий перечень из 10 сущностей.

- Тексты (электронные или на бумаге);
- Физические объекты;
- Мультимедиа (изображение, звук, видео и пр.);
- Программное обеспечение;
- Персоны;
- Организации;
- Территории;
- События;
- Идеи;
- Собрания всего вышеперечисленного (коллекции, базы данных, фонды).

При этом количество обязательных признаков для каждой сущности может быть очень невелико — в пределах 3-5 признаков. Именно на обязательных признаках должна строиться навигация и атрибутный поиск. Например, тематика может быть обязательным признаком для сущностей: текст, мультимедиа, организации, события, собрания.

Примером разумной, на мой взгляд, организации навигации по интегрированному пространству может служить интерфейс упомянутой выше информационной системы «Виртуальный музей Гулага». Он включает следующие виды навигации и поиска по следующим фасетам (с точностью до редакции):

- Категории информационных объектов (музеи, экспонаты, некрополи, памятники, изображения);
- Тематика;
- Персоны (поиск имен);
- География (поиск по географическим названиям);
- Учреждения;
- Источники и библиография
- Карта;
- Хронограф;
- Словарь терминов

Кроме того система предполагает лексический поиск, как по всем текстовым полям системы, так и по выбранным полям.

Конечно, этот вариант не является образцом, но как мне кажется, может служить ориентиром при проектировании интерфейса интегрированной системы.

## **Выводы**

Таким образом, при проектировании системы, интегрирующей информационные ресурсы библиотек, архивов и музеев, необходимо:

- Выбрать архитектуру системы и протоколы взаимодействия;
- Определить функциональность системы в целом и особенно ее централизованного блока;
- Выбрать стандарты метаданных, используемые в системе;
- Определить состав сущностей, к которым могут быть отнесены информационные объекты в системе;
- Определить состав признаков для навигации и атрибутивного поиска.

Все остальные проектные решения можно строить на основе решений, заложенных в этих базовых принципах.

### Литература

1. А.Б. Антопольский, Г.А. Савин, А.Н. Сотников. Электронная библиотека «Научное наследие России»: новый этап развития // Информационные ресурсы России, 2010, № 2, с 2-6.
2. <http://www.athenaeurope.org/index.php?en/1/home>.
3. <http://www.minervaplus.ru/#more>.
4. <http://gulagmuseum.org/start.do;jsessionid=763E4A4CCA21DC38F639004769B49714>.
5. Смирнов В.Н., Грязнова Н.А., Ходнев А.С., Русаков А.И., Создание электронной библиотеки письменного наследия Ярославского края// Электронные библиотеки, 1999, т.2, вып 4.
6. <http://www.naveki.info/>.
7. Интеграция библиотечных и архивных информационных систем / А.Б. Антопольский, Т.С. Маркарова) // Ева 2007, М. декабрь 2007.
8. Публикация Рабочей группы 3. Исследование стандартов и подготовка рекомендаций» проекта АТЕНА. Пер. с англ Н.В. Бракер, Ред. Л.А.Куйбышев. Центр ПИК, 2010.