

Вопросы разработки профилей электронных библиотек

Е.Н.Филинов, А.В. Бойченко
Институт системного программирования РАН
(boytchen@ispras.ru)

Введение

В докладе излагается краткое содержание подхода, предлагаемого в раздел 1 межведомственной программы "Российские электронные библиотеки".

Подход направлен на методологическое обеспечение проектирования, сопровождения и развития конкретных электронных библиотек (ЭБ), создаваемых в России. Актуальность предлагаемого подхода обусловлена ожидаемым расширением фронта работ в России в области создания и развития электронных библиотек в условиях, когда не все библиотеки располагают кадрами профессионалов по информационным технологиям и финансовыми средствами для заказов проектов специально для каждой библиотеки.

При реализации данного подхода предполагается:

- выбрать и конкретизировать применительно к распределенной архитектуре "клиент-сервер" эталонную модель среды открытых систем OSE/RM, позволяющую идентифицировать совокупность интерфейсов и протоколов взаимодействия между приложениями и средой, в которой функционируют приложения ЭБ;

- определить рекомендуемый набор стандартов на интерфейсы и протоколы взаимодействия, аппаратно-программные платформы ЭБ;

- выбрать стандарты на форматы электронного обмена данными, необходимые для доступа к электронным каталогам библиотек и обмена библиографическими данными, принятие которых сообществом библиотек России позволит сформировать национальную информационную инфраструктуру в части доступа к электронным библиотекам и доступа к мировым информационным ресурсам.

ИСП РАН в рамках подпрограммы "Информатизация России" в 1996-1998 гг. разработал методику формирования и применения профилей открытых ин-

формационных систем и основные принципы построения профилей информационной инфраструктуры региона России. В качестве одной из прикладных областей были рассмотрены ЭБ. Проработаны варианты концептуальных моделей ИС применительно к разным типам архитектуры (среды распределенных вычислений с удаленным вызовом процедур - DCE и среды с механизмом брокера объектных запросов - CORBA). Стандартизованные форматы электронного обмена данными (типа UNIMARC и др.) рассматривались вместе с поддерживаемыми их услугами телекоммуникационной среды.

В докладе предлагаются основные положения построения профиля информационной инфраструктуры науки, культуры и образования, в которой интегрируются информационные ресурсы, предоставляемые электронными библиотеками.

Работа поддержана грантом РФФИ 98-01-00978.

Рекомендуемые стандарты для профиля электронной библиотеки.

Стандарты, связанные с информационными ресурсами электронных библиотек (ЭБ):

- построение электронных каталогов библиотек и доступ к ним, а также доступ к библиографическим базам данных;

- электронные представления полных текстов изданий и аудиовизуальных материалов, хранящихся в библиотеках.

1. Международные стандарты идентификации объектов в библиотечном деле.

Для идентификации книг, периодических изданий, аудиовизуальных материалов и электронных записей, а также издательств и библиотек, приняты международные стандарты, которые ведут технические комитеты ISO TC 46 / SC9 и ISO / IEC JTC1 / WG4:

- ISO 2108:1992. Information and documentation. International Standard Book Numbering (ISBN);

© Вторая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
26-28 октября 2000 г. Протвино

- ISO 3297:1986. Documentation - International Standard Serial Numbering (ISSN);
- ISO 3901:1986. Documentation - International Standard Recording Code (ISRC);
- ISO 10444:1994. Information and documentation. International Standard Technical Report Number (ISRN);
- ISO 10957:1993. Information and documentation. International Standard Music Number (ISMN);
- ISO 15511 International Standards Library and organization Identifier (ISLOI);
- ISO/IEC 9070:1991. Information technology - SGML support facilities - Registration procedures for public text owner identifiers.

Для идентификации периодических изданий и отдельных публикаций (статей) в этих изданиях также применяется американский стандарт ANSI / NISO Z39.56-1991. Serial Issue and Contribution Identifier (SICI). Стандарт SICI в первоначальной версии широко используется на уровне идентификации изданий во многих библиотечных системах мира как важный элемент сообщений электронного обмена данными.

Существенная ревизия этого стандарта, выполненная в последние годы, обеспечила возможность широкого его применения в электронном обмене данными с использованием среды Интернет (унифицированные имена ресурсов).

В 1996-97 гг. ассоциация американских издателей (AAP- Association of American Publishers) представила проект, названный DOI (Digital Object Identifier), который обеспечивает управление защитой авторских прав при обмене электронной информацией. Идентификатор, установленный для некоторого информационного ресурса, позволяет пользователю направить запрос к центральной базе данных DOI. Запрос передается издателю, который возвращает пользователю полную запись ресурса, включающую как библиографическую информацию, так и сведения о статусе копирайта этого ресурса.

Вопрос полного или частичного применения указанных выше международных стандартов в российских библиотеках выходит за пределы настоящего документа. Отметим только, что решения по этому вопросу, принятые соответствующими федеральными органами РФ, должны быть приняты как исходные при формировании профилей ЭБ, касается правил идентификации объектов.

2. Стандартный формат обмена библиографическими данными MARC.

Стандартный формат обмена библиографическими данными MARC (Machine Readable Cataloguing) является основой для формирования электронных каталогов библиотек и библиографических баз данных.

Стандарты на формат обмена MARC поддерживаются техническим комитетом ISO TC46 / SC4, международной ассоциацией библиотек IFLA, Библиотекой Конгресса США, Британской библиотекой, Национальной библиотекой Канады, Постоянным комитетом UNIMARC. Формат MARC широко принят для обмена библиографическими данными в разных странах. В частности, его используют 69% библиотечных систем в Европе.

Исторически сложились разные версии формата MARC: американская USMARC, разработанная для Библиотеки конгресса США, и английская UKMARC, разработанная для Британской библиотеки. В настоящее время проводится работа, направленная на то, чтобы принять USMARC в расширенной форме и для Великобритании.

Разработан международный формат UNIMARC, который поддерживает Permanent UNIMARC Committee. Недавно UNIMARC был выбран многими европейскими библиотеками для обмена каталожными данными, а также Консорциумом Европейских исследовательских библиотек (CERL) для формирования базы данных Hand Printed Books, объединяющей данные из разных библиотек Европы.

Форматы MARC для обмена библиографическими данными определены следующими документами:

- The Anglo-American Cataloging Rules (1994). Ottawa: Canadian Library Association; London: Library Association Publishing; Chicago: American Library Association;
- ISBD(G) General International Standard Bibliographic Description (1992) prepared by IFLA Committee on cataloging;
- The UKMARC Manual: a Cataloguer's guide to the bibliographic format (1996);
- UNIMARC Manual (1994) IFLA Universal Bibliographic Control and International MARC Programme, British Library Bibliographic Services;
- USMARC Format for Bibliographic Data. Including Guidelines for Content Destination (1994), ISBN 0-8444-0809-3.

База данных OCLC (Online Computer Library Center), в которой применен формат USMARC, используется более чем 20000 библиотеками в 61 стране. Она содержит в настоящее время более 30 млн. библиографических записей на 400 языках. Ежедневно в нее добавляются более 34 тысяч записей по монографиям, периодическим изданиям, аудиовидео материалам, картам, архивам, рукописям, компьютерным файлам. Эти записи оформляются в соответствии с англо-американскими правилами каталогизации (AACR2), принятыми в 1978 г. и дополненными в 1993 г.

OCLC является результатом кооперированной деятельности по каталогизации Библиотеки Конгресса США, Британской библиотеки, Национальной библиотеки США по медицине, Национальной библиотеки Канады и Национальной библиотеки Австралии.

При формировании библиотечного сектора в регионах России следует учитывать обеспечение возможности доступа к OCLC, как к мировой базе библиографических данных.

3. Форматы электронных сообщений, связанных с межбиблиотечным обменом.

Форматы электронных сообщений, связанных с межбиблиотечным обменом (по межбиблиотечным абонентам) установлены международными стандартами, поддерживаемыми Техническим комитетом ISO TC46 / SC4:

- ISO 10160:1997. Information and documentation - Open Systems Interconnection. Interlibrary Loan Application Service Definition;

- ISO 10161:1997. Information and documentation- Open Systems Interconnection. Interlibrary Loan Application Protocol Specification. Part1: Protocol Specification. Part2: Protocol implementation conformance Statement proforma.

Стандарт ISO 10160 определяет спецификации сообщений по межбиблиотечному обмену (ILL-InterLibrary Loan), касающихся запросов на требуемые читателям издания, сведений об отсылке затребованных книг, ответов на запросы и т.д. Стандарт 10161 определяет в нотации ASN.1 спецификации протоколов телекоммуникационной среды, необходимых для обмена сообщениями в архитектуре OSI.

Для использования этих двух стандартов определены международные профили OSI-ISP ALD 2n Library and Documentation Interlibrary Loan Store-and-Forward Protocols. Часть 4 (ALD22) этого профиля специфицирует применение протокола электронной почты X-400 для передачи сообщений, связанных с межбиблиотечным обменом.

В качестве альтернативы возможно трактовать ILL-сообщения как EDIFACT-сообщения.

Следующие проекты в области электронного межбиблиотечного обмена, поддерживаются Европейской комиссией:

- DALI, обеспечивающий услуги для поставки мультимедийных документов в распределенной среде, использующей протоколы ISO-ILL;

- UNIVERSE, исследовательский проект реализации ILL;

- ION OSI, пилотный проект европейских библиотечных сетей, Великобритании, Нидерландов и Франции, базирующийся на подмножестве ILL - протоколов и протоколе поиска в библиографических базах данных ISO-SR (см. ниже), реализованных на платформах с ОС типа UNIX;

- AIDA (Alternatives for International Document Availability).

4. Форматы команд для поиска в библиографических базах данных ISO-SR.

Международные стандарты, поддерживаемые Техническим комитетом ISO TC46 / SC4, определяют команды для поиска в библиографических базах данных с использованием услуг телекоммуникационной среды по протоколам OSI:

- ISO 8777:1993. Information and documentation - Commands for interactive text searching;

- ISO 10162:1993. Information and documentation- Open Systems Interconnection. Search and Retrieve Application Service Definition;

- ISO 10163. Information and documentation - Open Systems Interconnection. Search and Retrieve Application Protocol Specification. Part1 (1993): Protocol Specification.

Стандарт ISO 8777 определяет набор из 30 поисковых команд вместе с 8 специальными символами, которые могут использоваться, чтобы квалифицировать эти команды.

Стандарт ISO 10162 описывает, как услуги поиска, отвечающие заданным правилам, специфицируются на прикладном (седьмом) уровне OSI (по ISO 7498) и могут быть представлены протоколами ACSE нижележащих уровней OSI (по ISO 8649). Стандарт определяет услуги поиска и замещения (SR-search and replace) для библиотек, информационных служб и каталожных центров, используя модель “клиент-сервер”.

Стандарт ISO 10163 определяет спецификации служб (протоколов среды), необходимых для поддержки услуг поиска, соответствующих ISO 10162.

Для использования стандартов ISO 10162, 10163 определены профили OSI-ISP ALD1n Library and Documentation Search and Retrieve Specifications.

ISO 10162 используется как подмножество протокола Z39.50 (см. ниже).

Форматы ISO-SR применены в проекте Europe Gate, выполняемом в рамках Европейской библиотечной программы. Программное обеспечение, реализующее клиентскую часть протокола ISO-SR (Z39.50 V2) разработано как часть проекта SOCKER, выполненного при поддержке Европейского сообщества. Серверная часть протокола реализует услуги, обеспечивающие доступ к онлайн-публичным каталогам. Они разрабатываются в европейском проекте ARCA (Access to Remote Catalogues).

Форматы ISO-SR применены также в проекте OPAC Network in Europe (ONE), который должен обеспечить клиентам библиотек лучший доступ к онлайн-публичным каталогам (OPAC-Online Public Access Catalogues) и другим национальным каталогам. Некоторые из участников проекта уже предлагают услуги, базирующиеся на применении протоколов Z39.50 и / или ISO-SR.

5. Определение прикладной услуги информационного поиска и спецификация протокола ANSI Z39.50.

Стандарты, поддерживаемые ANSI и US National Information Standards Organization (NISO), распространяются на услуги поиска в библиографических базах данных, совместимые с протоколами OSI. К этим стандартам относятся:

– ANSI / NISO Z 39.50-1992. Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection (revision and redesignation of ANSI Z 39.50-1988);

– ANSI / NISO Z 39.50-1995. Information retrieval (Z 39.50): Application service definition and protocol Specification (revision of ANSI Z 39.50-1992). По содержанию идентичен ISO 23950;

– ISO 23950:1998. Information and documentation - Information retrieval (Z 39.50). Application service definition and protocol Specification.

Протокол Z 39.50 на прикладном (седьмом) уровне OSI специфицирует запросы поиска информации в онлайн-библиотечных каталогах и других структурированных наборах данных. ANSI Z 39.50 усиливает правила ISO, определенные стандартами ISO 10162, 10163, и представляет множество поисковых опций для использования в OSI-совместимых приложениях. После того, как ISO 10162 был реализован, для американских библиотек потребовались дополнительные функции просмотра и управления доступом. В связи с этим в ISO был представлен стандарт ISO 23950.

Для поиска информации в библиографических базах данных введен профиль ATS-1 стандарта Z 39.50, который специфицирует использование набора атрибутов BIB-1.

В США стандарт Z 39.50 составляет основу для Government Information Locator Service (GILS). Официальным органом, ведущим регистрацию реализаций Z 39.50 в мире, является Библиотека Конгресса США.

В Европе на Z 39.50 базируется проект OPAC Network in Europe (ONE), обеспечивающий клиентам библиотек доступ к онлайн-публичным каталогам. В рамках Европейской библиотечной программы протоколы Z 39.50 реализуются в проектах SOCKER и ARCA.

В связи с тем, что во многих случаях телекоммуникационную среду определяет стек протоколов TCP / IP, в Европе разрабатывается проект Paragon, финансируемый Европейским сообществом. Цель этого проекта - обеспечить клиентам библиотек простые шлюзы к электронным каталогам и архивам мультимедиа-информации через WWW. Одновременно Европейское сообщество финансирует проект EUROPA Gate, в котором обеспечивается шлюз между Z 39.50 и WWW и пользователями электронной почты через Internet.

Профили для применения Z 39.50 в Европе разрабатываются Европейским Форумом Реализаторов библиотечных приложений (EFILA) совместно с

IFOBS и международной группой реализаторов Z 39.50 (ZIG).

6. Стандарты представления электронных документов.

Технический комитет ISO TC 46 / SC9 разработал стандарт ISO 12083. Electronic manuscript preparation and markup, в который сведены требования к структуре документов и к связанным с ними библиографическим данным, которые ранее действовали как разрозненный набор стандартов ISO, выпущенных в 70-80-х годах. Стандарт ISO 12083 поддерживается Техническим Комитетом ISO TC 46 / SC4, ANSI и Ассоциацией американских издателей (AAP).

Этот стандарт содержит стандартизованный набор тегов разметки языка SGML и определение типа документа (DTD -Document Type Definition) применительно к книгам и статьям, а также к периодическим изданиям.

7. Форматы электронного обмена полными текстами документов GEDI.

Форматы и протоколы электронного обмена полными текстами документов поддерживаются Техническим Комитетом ISO TC 46 / SC4 WG4.

Они определены соглашениями GEDI (Group on Electronic Data Interchange Agreements), которые касаются двух аспектов:

- определение формата электронного документа;
- описание механизма обмена.

Формат GEDI состоит из информации дескриптивного заголовка, за которым следует образ документа. В первой версии формат GEDI содержал ограниченный набор дескрипторов для описания образа документа в формате TIFF и специфицировал механизм FTAM для обмена.

Во второй версии набор дескрипторов был расширен так, чтобы связать элементы данных между GEDI и ISO ILL, наряду с FTAM был введен как механизм обмена протокол FTP.

Третья версия GEDI, разработанная в 1997 году, предусматривает следующие форматы документов:

- TIFF;
- PDF;
- JPEG;
- Post Script.

В качестве механизма обмена введен протокол электронной почты MIME, принятый для Интернет, и удален FTAM.

Соглашения GEDI широко используются в Европе, США и Австралии для поддержки национальных и международных служб доставки документов в библиотечном мире.

В Европе возможности использования соглашений GEDI с учетом многонационального аспекта европейских стран демонстрирует проект EDIL, финансируемый Европейским сообществом.

8. Рекомендации по построению профилей конкретных ЭБ.

Если принять как основу для конкретных ЭБ в России продукт CDS/ISIS, распространяемый ЮНЕСКО, (уже есть прецеденты, например ЭБ МЭСИ), то представляется важным выяснить, каким стандартам соответствует CDS/ISIS и построить ее профиль как "фотографию" существующей системы. Такая "фотография" послужит в качестве методической поддержки создания конкретных ЭБ.

Доступ пользователей Интернет к библиографическим базам данных, поддерживаемых в среде CDS/ISIS, следует рассматривать как услугу Web-технологий для читателей данной конкретной библиотеки. Одновременно следует рассмотреть использование Интернет для доступа из данной ЭБ к каталогам других библиотек.

В принципе цепочку стандартов Web-технологий в этих случаях должны составить HTML-HTTP-TCP/IP. Интерфейсы прикладного программирования (API) в профиле конкретной ЭБ должны быть специфицированы в зависимости от принятых операционных систем узлов ЛВС ЭБ: Windows NT/2000 или Unix (стандарты POSIX).

9. Источники информации.

9.1 Directorate General XIII of the European Commission. The EU Telematics for Libraries Programme.
<http://www.echo.lu/libraries/ln/libraries.html>

9.2. EU Libraries Programmer's Europe Gate project
<http://europagate.dtv.dk/reports.htm>

9.3. NISO, National Bureau of Standards. ANSI Z 39.50
<http://leweb.loc.gov/Z 3950/agency>
<http://linnea.helsinki.fi/Z 3950/Z 3950 pr.html>

9.4. Group on Electronic Data Interchange Agreement. Европейский проект EDIL.
<http://www2.echo.lu/libraries/ln/projects/edil.html>

9.5. Европейский проект AIDA (Alternatives for International Document Availability).
<http://www.cib.unibo.it/aida/>

9.6. Library of Congress Network Development and MARC standards office
<http://lcweb.loc.gov/marc/>

9.7. British Library National Bibliographic Service (NBS) UKMARC Web Page
<http://www.bl.uk/services/bsds/nbs/marc>

9.8. Elsevier Science. Publishes Item Identifier Scheme
<http://www.elsevier.nl/inca/homepage/about/pii>

9.9. Digital Object Identifiers System
<http://www.doi.org>

9.10. Европейский проект BIBLINK
<http://www.ucoln.ac.uk/metadata/BIBLINK/wp2/>

9.11. European Bureau of Library, Information and Documentation Associations
<http://www.kaapeli.fi/~eblida/ecup/>

9.12. British Library. Setting the Record Straight: A guide to the MARC format (third edition, 1998)

9.13. National Library of Canada (NLC), назначенная ISO регистрационным органом по ISO 10160 и 10161.
<http://www.nlc-bnc.ca/iso/ill/register.htm>
<http://www.nlc-bnc.ca/iso/ill/regiimp.htm>

9.14. Европейский проект SOCKER
<http://mediator.uni-c.dk/Socker>

Список литературы

- [1] ГОСТ Р ИСО/МЭК ТО 10000-1-93. Информационная технология. Основы и таксономия функциональных стандартов. Часть 1. Основы.
- [2] ISO/IEC TR 10000-3:1998. Information Technology. Framework and taxonomy of International Standardized Profiles. Part 3. Principles and taxonomy for Open System Environment Profiles.
- [3] Липаев В.В., Филинов Е.Н. Формирование и применение профилей открытых информационных систем // Информационные технологии, № 4, 1997, с. 2-11.
- [4] В.В.Липаев, Е.Н.Филинов. Мобильность программ и данных в открытых информационных системах. РФФИ. М. 1997.
- [5] Всероссийская научно-методическая конференция "ТЕЛЕМАТИКА'99", 7-10 июня 1999 года, Санкт-Петербург. Тезисы докладов. Стр. 99-105.