

УДК 004.415

В.А. Резниченко, Г.Ю. Проскудина, К.А. Кудим, О.М. Овдий

О ПОСТРОЕНИИ МОДЕЛЕЙ ЭЛЕКТРОННЫХ БИБЛИОТЕК

Работа посвящена задаче создания моделей электронных библиотек. Обсуждаются некоторые известные связанные проекты – CIDOC CRM, FRBR, FRBRoo, DELOS DLRM и 5S. Исследование может быть положено в основу создания собственной оригинальной модели ЭБ, которая должна по-возможности использовать их сильные стороны и избежать имеющиеся недостатки.

Введение

Появление новых электронных библиотек (ЭБ), увеличение числа хранимых в них документов и повышение качества предоставляемых ими услуг способствует развитию науки, облегчая, а иногда и просто открывая единственно возможный доступ к источникам информации для ученого, предоставляя ему замечательное средство донести плоды своей деятельности до широчайшей аудитории. В последние несколько лет при нашем непосредственном участии научное сообщество Украины продвинулось в этом направлении. В частности, в прошлом году создан портал периодических изданий НАН Украины¹. Два года назад создана ЭБ Института программных систем НАН Украины². В первом случае использовалось программное обеспечение DSpace, во втором – EPrints. Обе системы были полностью украинизированы. Были отработаны основные сценарии использования, создан ряд методик и рекомендаций по созданию и использованию электронных библиотек на основе данных программных систем. Были также изучены программные продукты Greenstone³ и Fedora⁴. Этот опыт оказался очень ценным для понимания современного состояния дел в мире программных систем ЭБ.

В настоящее время нет какой-либо универсальной ЭБ, которая отвечала бы всем требованиям и ожиданиям пользователей. Анализ существующих систем ЭБ [1–3] показывает их разнородность на нескольких уровнях:

- на уровне информационной модели, которую они обеспечивают;
- на уровне поддержки пользователей и групп пользователей;
- на уровне функциональных возможностей.

Из-за этой гетерогенности ЭБ и игнорирования нужд их пользователей возникает ряд проблем:

- интеграция информации из различных ЭБ;
- сравнение ЭБ по предоставляемой функциональности;
- оценка и сравнение производительности различных систем ЭБ;
- добавление новых типов хранимых объектов;
- добавление новых функциональных возможностей;
- резервное копирование.

Решить эти и другие возникающие проблемы на первом этапе поможет аккуратное и полное рассмотрение области ЭБ. Именно для этого создаются различные модели, обобщающие накопленный опыт в сфере создания и использования ЭБ.

В последнее время в мире предпринимаются усилия по полному и всестороннему описанию сферы ЭБ. В данной работе мы обсуждаем и анализируем следующие известные модели и стандарты, которые могут применяться для описания ЭБ в целом и ее частей: CIDOC CRM [4], FRBR [5], DELOS DLRM [6] и 5S [7]. Это исследование может быть положено в основу создания собственной оригинальной модели ЭБ, которая должна по-возможности использовать их сильные стороны и избежать имеющиеся недостатки.

¹ <http://dspace.nbuv.gov.ua:8080/dspace>

² <http://eprints.isofts.kiev.ua/>

³ <http://www.greenstone.org/>

⁴ <http://www.fedora-commons.org/>

1. CIDOC CRM

Концептуальная эталонная модель (Conceptual Reference Model, CRM) CIDOC, разработанная Международным комитетом по документации Международного совета музеев (The International Committee for Documentation of the International Council of Museums, ICOM-CIDOC), предназначена для интеграции, посредничества и обмена информацией в области мирового культурного наследия и связанных областей.

Четко обозначенные цели определили основные конструкции модели и уровень ее детализации. CIDOC CRM в терминах формальной онтологии определяет семантику схем баз данных и структур документов, используемых в культурном наследии и музейной документации. Модель не определяет терминологию, появляющуюся в конкретных структурах данных, но имеет характерные отношения для ее использования. Она не стремится предлагать то, что должны документировать учреждения культуры. Скорее она объясняет логику того, что они фактически документируют и таким образом предоставляет семантическую интероперабельность между музеями, библиотеками, архивами. Представляя определения и формальную структуру описания неявных (implicit) и явных (explicit) сущностей и отношений, модель CIDOC CRM претендует на общий язык для экспертов предметной области и специалистов по информационным технологиям. Она предназначена для покрытия контекстной информации исторического, географического и теоретического характера об отдельных экспонатах и музейных коллекциях в целом [8, 9].

До 1994 года разрабатывалась ER-модель для музейной информации, начиная с 1996 года подход разработки модели сместился к методологиям объектно-ориентированного моделирования и привел в 1999 году к появлению первой Концептуальной эталонной модели CIDOC. С 2000 года начался процесс стандартизации, который успешно завершился принятием стандарта ISO 21127:2006 – "Эталонная онтология для обмена информацией культурного наследия" (A reference ontology for

the interchange of cultural heritage information). Разработчики CIDOC CRM поставили своей целью написать стандарт модели, пригодной как для машинной обработки, так и для легкого понимания человеком. Модель совместима с формализмом RDF.

Версия 5.0.1 модели CIDOC CRM [4] состоит из 90 классов и 148 свойств (бинарных отношений), связывающих классы между собой и описывающих предметы, понятия, людей, события, место, время и их отношения. Все концепты (классы и свойства) модели можно разделить на три группы. Первая группа включает классы и отношения, охватывающие наиболее общие понятия окружающего мира: постоянные и временные сущности, отношения участия, зависимости, совпадения во времени. Вторая группа содержит понятия, частично поддерживающие функции управления: приобретение и учет единиц хранения, передача прав собственности на объекты культуры. К третьей группе относятся классы и свойства, используемые для внутренней организации самой онтологии: средства, необходимые для подключения внешних источников терминов, например, тезаурусов по отраслям культуры.

На рис. 1 представлена часть иерархии классов CIDOC CRM. Все классы, за исключением класса *простое значение* (Primitive Value) и его подклассов (эти классы являются вспомогательными), прямо или опосредовано являются подклассами класса *сущность CRM*, охватывающего все сущности, которые могут быть описаны в CIDOC CRM. Далее иерархия классов делится на 2 ветви: *постоянные сущности* и *временные сущности*.

На самых нижних уровнях иерархии классов появляются понятия, характерные для сферы культуры: *хранение*, *перемещение* (ценностей), *проект* или *процедура* (в том числе техника производства), *период* (в том числе художественный стиль). Иерархия классов может быть гибко расширена с применением встроенного класса *тип*. Наибольший интерес представляют свойства. Классы на нижних уровнях иерархии имеют около 10–15 свойств, причем большая часть свойств наследуется от

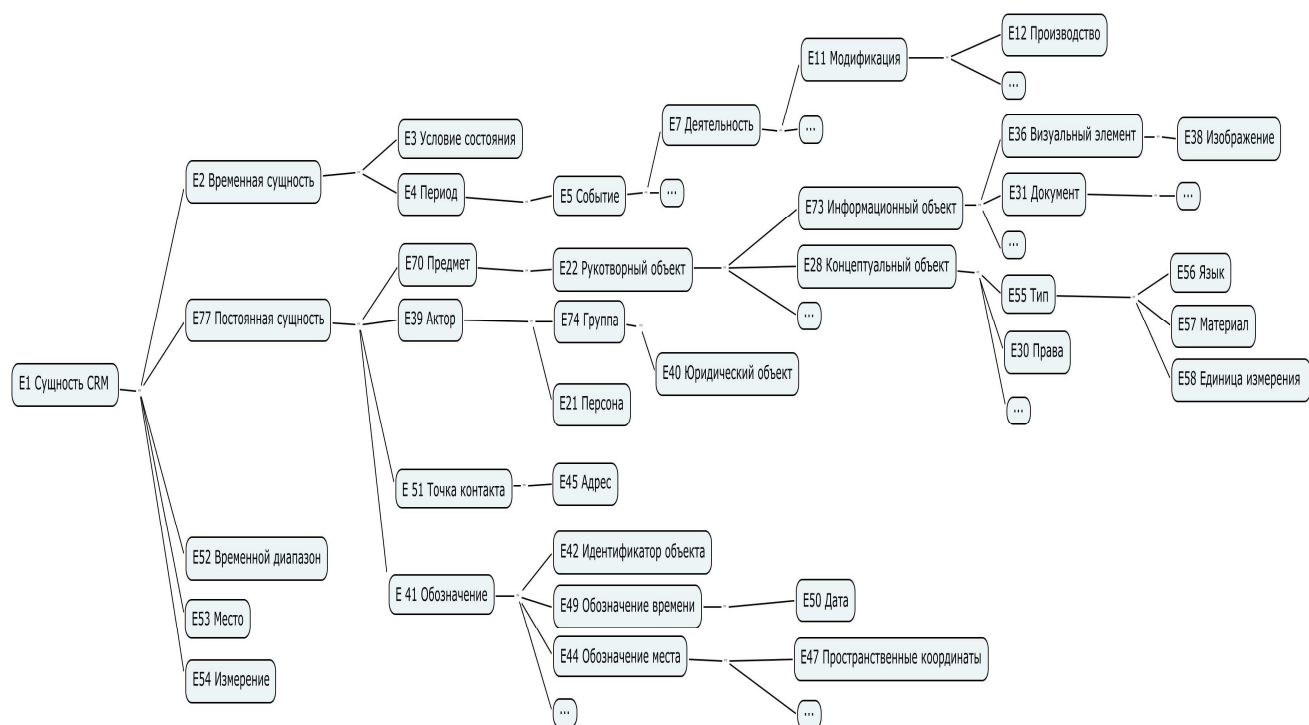


Рис. 1. Часть иерархии классов в модели CIDOC CRM

классов-предков. Названия свойств представляют собой глагольные фразы, выбранные так, что при последовательном связывании двух классов свойством получается осмысленная фраза с субъектом (первый, если считать слева направо, класс), предикатом (свойством) и объектом (второй класс).

1.1. Принципы моделирования CIDOC CRM

Разработка CIDOC CRM проводится в соответствии со следующими принципами.

Открытый мир – термин и принцип систем баз знаний подразумевает, что хранящая в этих системах информация является неполной относительно универсума той предметной области, которую они намереваются описать. Такие проблемы характерны для информационных систем в области культуры. Наши записи о прошлом обязательно неполны.

Кроме того, существуют свойства, которые не могут быть в точности назначены данному классу. В частности, отсутствие определенного свойства для описанного в системе элемента, не означает, что у этого элемента нет этого свойства. Например,

если один элемент описан как *биологический объект*, а второй как *физический объект*, то это не означает, что, второй возможно, не *биологический объект* также. Поэтому дополнения класса относительно суперкласса не могут быть в общем виде выведены из информационной системы, используя предположение Открытого мира (аналогично отсутствию понятия *универсума* в системах БД). Например, нельзя перечислить “все *физические объекты*, известные системе, которые в реальном мире не являются *биологическими объектами*”, но можно конечно перечислить “все элементы, известные системе как *физические объекты*, но не известные системе как *биологические объекты*”.

Монотонность. Поскольку основная роль CRM – интеграция информации в Открытом мире, то она стремится быть монотонной. Принцип состоит в том, что существующие конструкции CRM и выведенные из них дедукции (логические выводы) всегда остаются достоверными и правильно построенными, после добавления новых конструкций, расширяющих CRM. Например, если было уже описано, что какой-то объект является экземпляром *физического объекта*, а потом его охарактере-

ризвали как экземпляр *биологического объекта*, то в дальнейшем система не прекращает его рассматривать как экземпляр *физического объекта*. Для того чтобы формально сохранять монотонность для частых случаев альтернативных мнений, все формально определенные свойства должны быть реализованы как неограниченные (типа многие-ко-многим) и тогда, противоречивые экземпляры просто будут накапливаться. Такое интегрированное знание на основе CRM, представляющее совокупность релевантно альтернативных вариантов для четко определенных объектов, является основой исследования, а заключение об истине – задача науки.

Например, Эль Греко и Король Артур должны всегда оставаться экземплярами сущности *персона*. И как только мы ввели эти экземпляры в нашу базу знаний, эта сущность (*персона*) всегда подразумевается. Альтернативные мнения о их свойствах, например, место их рождения и жизни должны накапливаться без решения о том верны они или нет.

Минимальность. Хотя возможности CRM очень широки, сама модель построена экономно. Класс не объявлен до тех пор, пока область определения или область значений его свойств не соответствуют его суперклассу или пока он не является ключевым понятием в рассматриваемой области. По умолчанию классы и свойства CRM, которые имеют общий суперкласс, не являются взаимно исключаящими. Например, объект может быть экземпляром как *биологического объекта*, так и *искусственного объекта*.

Классы и свойства CRM являются либо примитивами, либо они – ключевые понятия в рассматриваемой области. Дополнения классов CRM не декларируются (поскольку нет универсума), т. е. модель фиксирует только то, что реально имеется.

Сокращения. Некоторые свойства декларируются как сокращения более длинных свойств. Например, свойство сущности E19 *физический объект* P52 *имеет владельца* E39 *актер* является сокращением от более полного пути: *физический объект* через сущность *приобретение* до сущности *актер*. Экземпляр

полного пути всегда подразумевает экземпляр сокращенного свойства. Однако, инверсия, возможно, не верна; экземпляр полного пути может не всегда быть выведен из экземпляра сокращенного свойства.

Непересекаемость. Классы не пересекаются, когда у них нет общих экземпляров. Например, не пересекаются пары классов *временной объект* и *постоянный объект*; *физический предмет* и *концептуальный объект*, тем самым различаются материальные и нематериальные объекты. Экземпляры *физического предмета* и *концептуального объекта* имеют принципиальное отличие; например, изготовление экземпляров *физического предмета* подразумевает использование физического материала, тогда как изготовление экземпляров *концептуального объекта* нет. Точно так же экземпляры *физического предмета* прекращают существовать когда они разрушаются, тогда как экземпляр *концептуального объекта* погибает, когда о нем забывают, или когда разрушается его последний физический носитель.

Тип. Фактически все структурированные описания музейных объектов начинаются с уникального идентификатора объекта и информации о типе объекта, который часто заносится в поля, названные как *тип объекта*, *название объекта*, *категория*, *классификация* и т. д. Все эти поля используются для терминов, которые объявляют, что объект – член специфического класса или категории объектов и описан в CRM экземплярами класса *тип*. Поскольку экземпляры этого класса в свою очередь являются самостоятельными классами, класс *тип* – фактически метакласс.

Класс *сущность CRM* является доменом свойства P2 *имеет тип* (тип чего-то), который имеет область значения класс *тип*. Следовательно, каждый класс в CRM, за исключением класса *простое значение*, наследует свойство P2 *имеет тип* (тип чего-то). Это обеспечивает общий механизм уточнения классификации экземпляров CRM до любого уровня детализации, связываясь с внешними словарями-источниками, тезаурусами, схемой классификации или онтологиями, которые функционируют как расширение к иерар-

хиям классов и свойств CRM. Внешние словари не входят в CRM.

Расширения. Поскольку предполагаемые возможности CRM – подмножество реального мира, а значит, потенциально бесконечны, модель разрабатывалась так, чтобы быть расширяемой через соединения с совместимыми внешними иерархиями типов.

Охват. Некоторые понятия CRM разработаны неполно, например: *актор* и *право*. Это – естественное следствие пребывания в пределах ясно сформулированной практической сфере CRM.

Резюмируя рассмотрение данной модели, отметим, что для наших целей это – нужный и полезный стандарт. Важным преимуществом стандарта является его формальный подход. Обобщая все многообразие музейных коллекций и задач, модель содержит широкий набор универсальных понятий. Важную роль здесь играют временные сущности, так как они связывают объекты (концептуальные или физические) с временным диапазоном, местом и субъектами. Этот стандарт вполне может служить основой для информационной составляющей концептуальной модели ЭБ. Конечно, он нуждается в расширении более конкретными сущностями, которые часто используются во многих ЭБ. Кроме того, CIDOC CRM не охватывает пользовательского и функционального аспектов ЭБ.

2. FRBR и FRBRoo

Независимо от CIDOC CRM в 1991–1997 годах Международной федерацией библиотечных ассоциаций и учреждений (International Federation of Library Associations and Institutions, IFLA) была разработана ER-модель "Функциональные требования к библиографическим записям" (Functional Requirements for Bibliographic Records, FRBR) как обобщенное представление библиографического универсума, независимого от какого-либо кода каталогизации или реализации.

В 1998 году модель была опубликована [5]. В настоящее время IFLA продолжает контролировать приложения модели

FRBR и поддерживает ее использование и развитие.

FRBR включает описание концептуальной модели (сущности, их отношения и атрибуты), предлагает универсальные библиографические записи для всех типов материалов и пользовательских задач, связанных с библиографическими ресурсами, описанными в каталогах, библиографиях и других библиографических инструментах [10, 11].

Модель FRBR различает три группы сущностей (рис. 2):

- для описываемых объектов: *произведение* (work), *выражение* (expression), *воплощение* (manifestation), *экземпляр* (item);
- для описателей-субъектов: *человек* (person) и *организация* (corporate body);
- для описателей-объектов: *концепт*, *объект*, *событие* и *место* (concept, object, event, place).

Далее приведен пример [12] экземпляров сущностей *произведения* (w1) и его *выражений* (e1-e2):

w1 Tennis-bis zum Turnierspieler Эльвангера:

- e1 оригинальный текст на немецком языке;
- e2 перевод на английский язык Венди Джилл.

Большое внимание в модели уделено отношениям между сущностями.

Отношения могут быть отражены в библиографических записях многими способами. Те, что изображены на ER-диаграмме FRBR (рис. 2), описывают логические связи между сущностями и часто реализуются простой конкатенацией одной сущности с атрибутами связанной сущности в одной записи.

Помимо логических связей в модели выделена группа так называемых контентных связей (для первой группы сущностей). Они идентифицируют основные типы отношений, которые существуют между экземплярами сущности одного типа (например, сущности *произведения*) или между экземплярами разных типов сущностей (например, сущностей *произведение* и *воплощение*). Например, в группе отношений *произведение-произведение* выделены

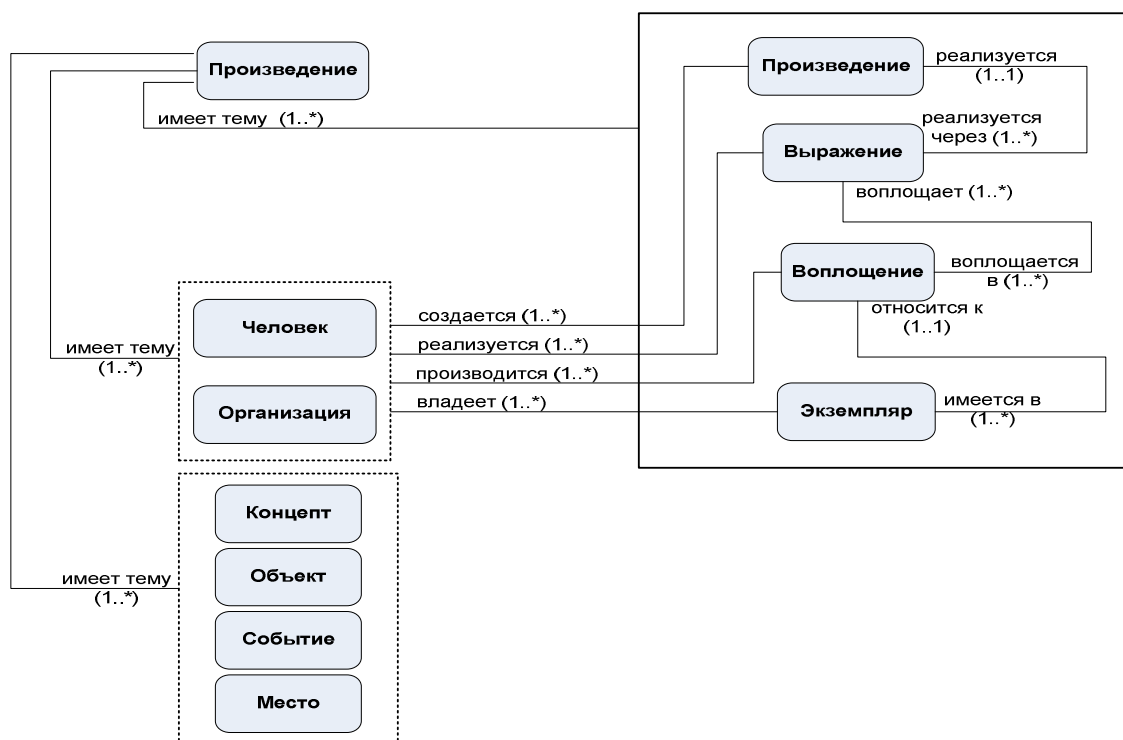


Рис. 2. Модель FRBR

такие типы отношений: *имеет адаптацию* (свободный перевод); *имеет приложение* (сходство, соответствие), *имеет продолжение*; *имеет резюме* (обзор, аннотацию); *имеет преобразование* (стихотворную форму); *имеет имитацию* (пародию). В группе отношений *выражение-выражение* перечислены следующие типы отношений: *имеет сокращение* (корректировку, уплотнение); *имеет пересмотр* (исправленную редакцию, расширенную редакцию); *имеет перевод* (буквальный перевод) и некоторые другие типы отношений, касающиеся музыкальных произведений. И наконец, отношения *часть/целое* и *часть* в части также представлены в модели FRBR.

Следует отметить, что в FRBR границы между различными типами основных сущностей (произведение и выражение) размыты и окончательное решение по тому, к какому типу отнести тот или иной объект, отдается на откуп каталогизатору. Кроме того, сущностей этих совсем немного и явно недостаточно для большинства конкретных библиотечных приложений. Для нас основной интерес представляет очень богатый набор атрибутов и отношений в этой модели. Как и в случае CIDOC CRM, в соответствии с решаемыми зада-

чами, FRBR применима только для описания информационной составляющей концептуальной модели.

Мысль о том, что и библиотечное, и музейное сообщество могут выиграть от гармонизации вышерассмотренных двух моделей, впервые прозвучала в Париже на 24-м Семинаре библиотечных систем ELAG (Европейской группы автоматизации библиотек). Однако реальная работа началась после 2003 года, когда была образована группа по пересмотру FRBR специально для гармонизации этих двух концептуальных моделей. В 2008 году рабочая группа по гармонизации FRBR и CIDOC CRM завершила важный этап работы: полная версия 0.9 draft объектно-ориентированной формулировки FRBR (FRBRoo) была представлена для обсуждения.

FRBR моделирует результаты (произведение, выражение...) процессов (таких как создание, реализация, планирование), но не сами процессы. FRBRoo, используя подход CRM, фокусируется на процессах. Такой подход позволяет принимать во внимание обстоятельства, при которых, например, конкретные произведения были задуманы или реализованы. Подобные об-

стоятельства могут быть предметом исследования (например, в теории литературы), но существующие библиографические инструменты не могут достаточно хорошо обеспечить данное исследование. Хотя можно возразить, что большинству библиотек не требуется выполнять подобные специализированные исследования, но это важно для общей модели, чтобы поддерживать как можно больше запросов. В каждом конкретном случае уровень сложности должен определяться с учетом всей имеющейся информации.

В результате в FRBRoo объекты *произведение*, *выражение* и *воплощение* были разбиты на несколько классов со специфическими свойствами. Так, в FRBRoo присутствует класс *произведение*, но также декларированы подклассы *индивидуальное произведение* (Individual Work), *составное произведение* (Complex Work), *сопроводительное произведение* (Container Work), *совокупность произведений* (Aggregation Work), *серийное произведение* (Serial Work), *издательское произведение* (Publication Work), *исполнительское произведение* (Performance Work), *записанное произведение* (Recording Work). Класс *произведение* является суперклассом, объединяющим подклассы как частные случаи, каждый из которых имеет свою специфику создания или составления. Данный анализ – это шаг к пониманию вопросов, связанных с совокупностями произведений (агрегатами).

FRBRoo следует рассматривать как интерпретацию FRBR, а не ее новую версию или замену. Главное новшество FRBRoo – реалистичная, явная модель процесса интеллектуального творчества, которая еще должна получить свое дальнейшее развитие для библиотекарей и ученых [13].

3. DELOS DLRM

Группа специалистов ассоциации в сфере ЭБ DELOS в 2006–2007 гг., основываясь на анализе имеющихся библиотечных систем [3], где большое внимание было уделено функциональным возможностям современных ЭБ, начали разработку эталонной модели ЭБ (Digital Library Reference Model, DLRM) [6]. Цель проекта

– разобраться с фундаментальными понятиями, существенными объектами и их отношениями, стандартными функциональными и структурными блоками и процессами, из которых состоит универсум ЭБ. Эталонная модель предназначена для разработки более узких моделей с конкретной архитектурой для последующей реализации программных систем.

Прежде всего, в модели было выделено три понятия для разграничения того, что обычно называется ЭБ:

– *ЭБ* – конкретная ЭБ с ее пользователями, правилами, содержимым, интернет-сайтом и ведущей организацией. Например, библиотека Института программных систем <http://eprints.isofts.kiev.ua>;

– *система ЭБ* – программное обеспечение, на основе которого создаются ЭБ. Например, EPrints 3.0.

– *система управления ЭБ* – программное обеспечение для создания и управления системами ЭБ. Например, система OpenDLib⁵.

Далее модель DELOS DLRM рассматривается в ролевом аспекте, т. е. с точки зрения разных категорий пользователей:

- конечный пользователь ЭБ;
- разработчик ЭБ;
- системный администратор ЭБ;
- разработчик приложений для ЭБ.

Соответственно DELOS DLRM имеет четыре уровня пользовательских представлений.

Весь универсум ЭБ разбит на шесть высокоуровневых ключевых областей (рис. 3):

- контент;
- пользователь;
- функциональные возможности;
- качество;
- политики;
- архитектура

и несколько дополнительных. Эти шесть областей объединены в одну область ресурса. В каждой из них вводятся и определяются свои сущности и их свойства.

Теперь вкратце рассмотрим наиболее важные области ЭБ и их структуру.

⁵ <http://www.opendlib.com>

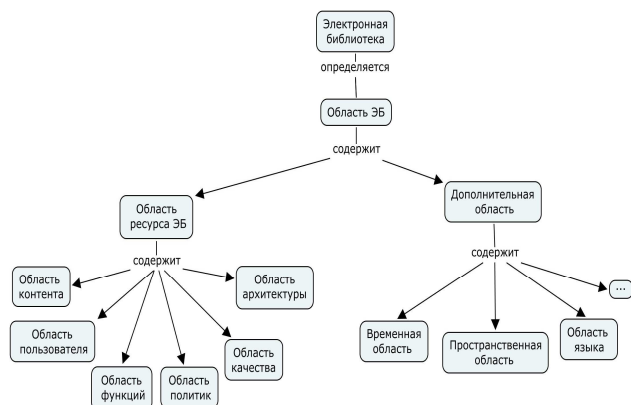


Рис. 3. Иерархия областей ЭБ в модели DELOS DLRM

Область ресурса ЭБ – наиболее общая область в данной модели, представляет все сущности и связи, населяющие универсум ЭБ. *Ресурс* – наиболее общее понятие, включающее любую сущность ЭБ. По аналогии с ресурсом в Веб, ресурс – это все то, что может быть идентифицировано, названо или адресовано. Представленная здесь модель ресурса исходит из веб-архитектуры, но дополнена некоторыми аспектами, специфичными для предметной области ЭБ.

Ресурс – абстрактное понятие, в том смысле, что непосредственно не имеет экземпляров, он только выражен экземплярами одной из своих специализаций. В частности, экземплярами понятия ресурс в универсуме ЭБ являются экземпляры *информационного объекта* любого типа (например, документы, изображения, видео, мультимедийные объекты, наборы метаданных и аннотаций, потоки, базы данных, коллекции, запросы и результаты запросов), *акторы* (как одушевленные так и неодушевленные сущности), *функции*, *политики*, *параметры качества ЭБ* и *архитектурные компоненты*. Каждый из этих экземпляров представляет главное понятие в своей области, таким образом в представленной модели ЭБ каждая область состоит из ресурсов, а ресурсы – строительные блоки всех областей ЭБ. Каждый ресурс:

- имеет идентификатор;
- организован в соответствии с *форматом ресурса*. Формат здесь выражен

онтологией. Ресурс может быть сложным и структурированным, поскольку он, в свою очередь, может состоять из меньших ресурсов и иметь связи с другими ресурсами;

- может характеризоваться параметрами качества;
- может регулироваться политиками, управляющими его жизненным циклом;
- выражается через информационный объект;
- может быть описан или дополнен информационным объектом, обычно – метаданными и аннотациями.

С организационной точки зрения, ресурсы могут группироваться в наборы ресурсов, которые рассматриваются как единая сущность. Например, *коллекции* в области контента или *группы* в области пользователя.

Область контента (рис. 4) представляет все объекты, связанные с информацией, которой управляет ЭБ. *Информационный объект* – наиболее общее понятие в этой области, представляет произвольную единицу информации, управляемую в универсуме ЭБ. В DELOS DLRM различают *информационный объект по уровню абстракции*, где заимствуются типы объектов из модели FRBR (произведение, выражение, воплощение) и *информационный объект по связи* – "абстрактный концептуальный контейнер для классов, которые порождают эти объекты", а именно:

- *первичный информационный объект* – информационный объект, который используется самостоятельно, например, текстовые документы, изображения;
- *метаданные* – информационный объект, главная цель которого состоит в том, чтобы дать информацию о целевом ресурсе (как правило о первичном информационном объекте);
- *аннотация* – информационный объект, главная цель которого состоит в том, чтобы аннотировать целевой ресурс или его часть (на рис. 4 части ресурса соответствует сущность *регион*). Примеры таких аннотаций включают примечания, структурированные комментарии и связи. Объекты аннотации помогают интерпретировать целевой ресурс, содержат либо

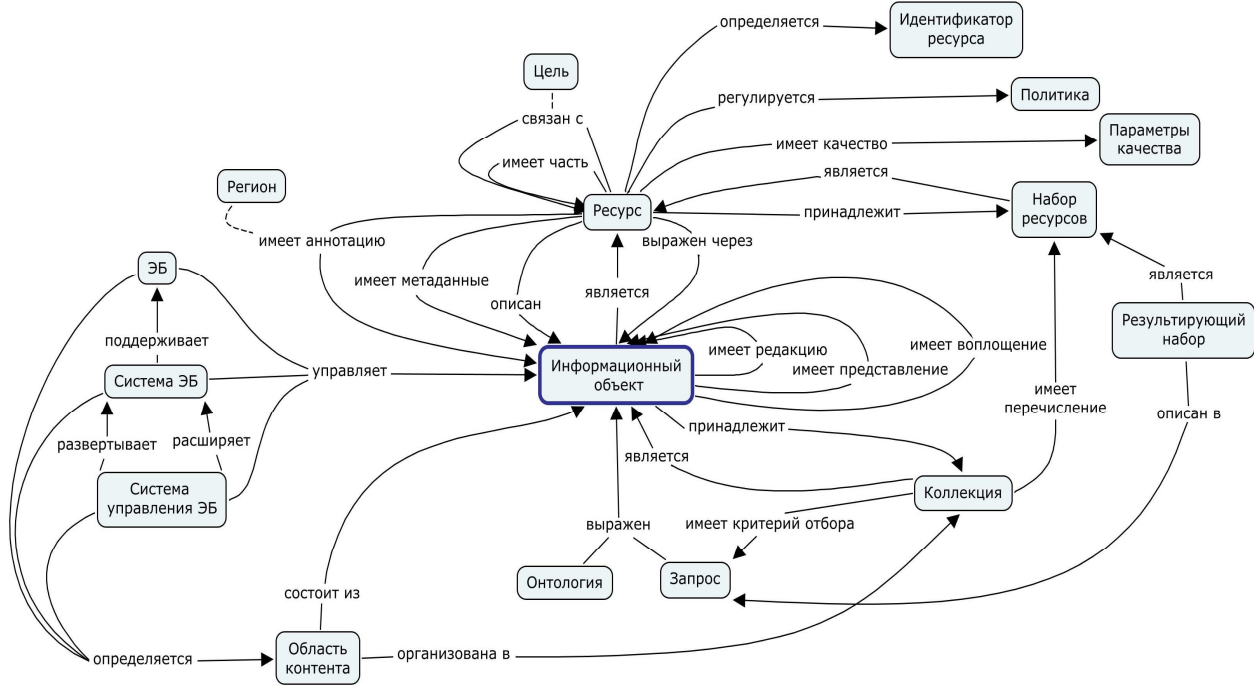


Рис. 4. Область контента ЭБ в модели DELOS DLRM

поддержку, либо возражения, либо более детальные объяснения.

Поскольку информационный объект является ресурсом, то он наследует все вышеперечисленные свойства ресурса.

Информационные объекты также могут быть сложными объектами и могут быть сгруппированы в *коллекции* информационных объектов. Коллекции, в свою очередь, тоже являются информационными объектами, они наследуют все аспекты моделирования информационных объектов и средства их обслуживания, например они могут аннотироваться. Кроме того, коллекция – специализация понятия *набора ресурсов*. Коллекции определяются *критерием отбора* (hasIntension) либо *перечислением элементов* (hasExtension). Другая специализация понятия набор ресурсов в данной модели – *результующий набор*. В традиционных ЭБ он представляет собой набор документов, которые извлекаются в ответ на *запрос*.

Область пользователя в DELOS DLRM содержит все объекты, которые являются "внешними по отношению к системе ЭБ и с ней взаимодействуют: люди и неодушевленные объекты, например, программы или физические инструменты... или даже другая ЭБ может быть среди пользователей ЭБ".

Поскольку главная сущность в этой области – *актор* является ресурсом и следовательно, наследует все его свойства, а именно:

- имеет уникальный идентификатор (идентификатор пользователя);
- организован в соответствии с форматом (модель пользователя);
- благодаря свойствам ресурса композиции и соединения может быть составлен в различные сложные и структурированные группы например, сотрудничество пользователей или соавторов;
- описан или дополнен метаданными и аннотациями.

Область функций представляет наиболее объемную и наиболее открытую часть модели DELOS DLRM, поскольку охватывает всю обработку ресурсов, а также действия пользователей в ЭБ. Здесь наиболее общим понятием является сущность *функция*. Функция – специфическая задача обработки, которая может быть реализована на наборе ресурсов или одном ресурсе в результате действий отдельного пользователя. Описание функций основано на пользовательском аспекте и ресурсе, представляющем все объекты, вовлеченные в ЭБ. Хотя функции в традиционных моделях ЭБ обычно связываются с контентом в ЭБ и выполняются людьми,

здесь, в данной модели, функции могут выполняться неодушевленными пользователями на любом типе ресурсов.

В данной модели ЭБ каждая функция также является ресурсом и потому наследует все его характеристики.

Функции разделены на пять классов:

- доступ к ресурсам;
- управление ресурсами;
- совместная работа;
- управление ЭБ;
- настройка ЭБ.

Внимательное изучение данной модели помогло не только обозреть всю сферу ЭБ, но и найти некоторые пробелы в самой модели. Вот некоторые из них:

- недостаточно формализованные определения, оставляющие размытыми границы многих сущностей (например, сущности, заимствованные из FRBR, или граница между *метаданными* и *аннотацией*);

- в некоторых местах остаются неясными критерии выделения сущностей (в частности, область качества наименее убедительна в этом отношении);

- неоднородность описания различных областей ЭБ, скрытая за внешне однообразным описанием (достаточно сравнить простую иерархию области функций со сложной структурой области контента).

К преимуществам DELOS DLRM следует отнести наибольшую полноту охвата среди существующих концептуальных моделей ЭБ.

4. Модель 5S

Goncalves и другие предложили и формализовали модель 5S – Streams, Structures, Spaces, Scenarios and Societies (Потоки, Структуры, Пространства, Сценарии и Сообщества) [7]. Целью авторов было не только разработать структуру для проектирования и разработки ЭБ, но также для того чтобы пользователям было проще понять данные технологии. В частности, 5S ориентирована на описание информационных систем как таковых. В рамках данной модели ЭБ – это комплексная система, которая помогает: удовлетворить информационные потребности пользователей (Сообществ), предоставить инфор-

мационные сервисы (Сценарии), организовать информацию в удобном виде (Структуры), предоставить полезную информацию (Пространства) и связать информацию с пользователями (Потоки) [14]. 5S позволяет определить ЭБ, используя эти 5 взаимодополняющих элементов. В табл. 1. приведены примитивы, формализмы и цели элементов [15].

С помощью формальной модели 5S, абстракции, такие как электронные объекты, метаданные, коллекции, сервисы и т.д. могут быть описаны посредством композиции базовых и высокоуровневых математических объектов.

Например,

- Сообщества + Сценарии = Пользовательская модель.

- Сообщества + Сценарии + Пространства + Потоки = Пользовательский интерфейс.

- Потоки + Структуры = Разметка.

- Потоки + Структуры + Сценарии = Объект.

- Структуры + Сценарии = СУБД.

Рассмотрим подробнее, что же из себя представляют элементы модели 5S.

Потоки (Streams) – последовательности элементов произвольных типов, используемых для описания как статического (например, текст) так и динамического (например, видео) контента. Динамический поток может представить поток информации или последовательность сообщений, таким образом, являясь важным элементом для представления любых коммуникаций в ЭБ.

Как правило, динамический поток воспринимается посредством его временной природы. Тогда динамический поток может быть интерпретирован как конечная последовательность показателей времени и связанных значений, которые могут использоваться для определения потоковой алгебры, позволяющей выполнять операции на разнообразных видах мультимедиа потоков. В статической интерпретации временная природа обычно не используется, и поток соответствует некоторому информационному наполнению, которое интерпретируется как последовательность

Таблиця 1. Примитивы, формализмы и цели элементов 5S

Измерение	Примитивы	Формализм	Цели
Потоки	Текст, видео, аудио, программа	Последовательности, типы	Описывает свойства контента ЭБ, такие как система кодировки и язык для текстовых материальных или специфических форм мультимедиа данных
Структуры	Коллекция, каталог, гипертекст, документ, метаданные, организационные инструментальные средства	Графы, узлы, ссылки, метки, иерархии	Определяет организационные аспекты контента ЭБ
Пространства	Пользовательский интерфейс, индекс; модель поиска	Наборы, операции, векторное пространство, пространство с мерой, пространство вероятности	Определяет логические и презентационные представления некоторых компонентов ЭБ
Сценарии	Сервис, событие, условие, действие	Диаграммы последовательностей, диаграммы сотрудничества	Детализирует поведение сервисов ЭБ
Сообщества	Содружество, менеджеры, актеры, классы; отношения, атрибуты, операции	Объектно-ориентированные конструкции моделирования, шаблоны	Определяет менеджеров; ответственных за выполнение сервисов ЭБ; актеров, которые используют эти службы; и отношения между ними

базовых элементов, обычно одного и того же типа. Примером статического потока является текст (последовательность символов). Тип потока определяет его семантику и область применения. Например, любое текстовое представление можно рассматривать как поток символов, таким образом, текстовые документы, такие как научные статьи и книги, можно рассматривать как структурированные потоки.

Структуры (Structures) определяют способ, которым части целого упорядочены или организованы. В ЭБ структуры представляют гипертексты, таксономии, взаимосвязи пользователей и т. д. Книги, например, можно логически структурировать в главы, разделы, подразделы или физически в страницы и строки. Электронные документы обычно структурированы с помощью языков разметки (например, HTML, XML). В реляционных базах данных данные структурированы с помощью схем, используя таблицы. В модели 5S структура определяется как помеченный ориентированный граф.

Пространство (Spaces) – любой набор объектов вместе с операциями над этими объектами, которые удовлетворяют определенным правилам. Объединение операций и объектов отличает пространства от потоков и структур. Они зачастую применяются, когда нельзя с помощью других измерений определить часть ЭБ, которую

могут моделировать традиционные библиотеки, используя пространства виртуального мира. Также пространства для поддержки совместной работы обеспечивают контекст, для виртуальных встреч и сотрудничества. Пространства отличаются операциями над своими объектами. ЭБ могут использовать много типов пространств, для того чтобы индексировать, визуализировать и выполнять другие операции. Самыми значимыми для ЭБ являются измеримые пространства, пространства с мерой, пространства вероятности, векторные пространства и топологические пространства.

Сценарии (Scenarios) состоят из последовательностей событий или действий, которые изменяют состояния, для достижения функциональных требований. Одним из важных типов сценария является текст, который описывает возможные способы использования системы, для достижения функции, необходимой пользователю. Сценарии полезны как часть процесса проектирования информационных систем. Они помогают визуализировать пространства, организовав потоки в соответствии со структурой. Таким образом, набор сценариев описывают сервисы, действия, задачи и операции, а они, в конечном счете, определяют функциональные возможности ЭБ.

Понятия состояния и события фундаментальны для понимания сценариев. Не-

формально, состояние определяется тем, какой информационный контент находится в указанных местах (например, в машинной памяти). Событие обозначает переход или изменение между состояниями (например, выполнение команды в программе). Сценарии определяют последовательности событий, которые вовлекают действия, которые изменяют состояния и влияют на возникновение и результаты будущих событий. В ЭБ сценарии моделируют потоки данных и рабочие процессы.

Сообщество (Societies) это ряд объектов и действий и отношений между ними. Объекты включают людей, а так же аппаратные и программные компоненты так или иначе связанные с ЭБ. Действия состоят в том, что объекты сделали, делают и будут делать друг с другом. Отношения определяют связи между объектами и действиями сообщества. Сообщества необходимы, чтобы описать контекст использования ЭБ. Сообщество - компонент высшего уровня библиотеки, поскольку ЭБ возникла, для того чтобы удовлетворить информационные потребности ее сообществ.

В электронных библиотеках определенные человеческие сообщества включают авторов, издателей, редакторов, разработчиков и библиотечный штат. В человеческом сообществе у людей есть роли, цели и отношения. Сообщества следуют определенным правилам, и их участники играют роли менеджеров, участников или пользователей. В участников есть действия и взаимосвязи. Во время их действий участники сообщества создают информацион-

ные артефакты, которыми может управлять ЭБ.

Рассмотренные измерения наряду с другими фундаментальными теоретическими определениями, используются для определения других конструкции ЭБ, таких как цифровые объекты, спецификации метаданных, коллекции, репозитории и сервисы.

На основе модели 5S было определено "ядро" ЭБ или "минимальную" ЭБ, т. е., минимальный набор компонентов, которые ее составляют и без которых приложение нельзя считать ЭБ. Каждый компонент (например, коллекции, сервисы) формально определяются с точки зрения конструкций 5S или как комбинация, или композиция двух, или более из них. На рис. 5 показаны различные уровни определений: математические основы (графы, последовательности и функции), измерения 5S и ключевые концепты (цифровой объект, коллекция) [16]. Стрелки представляют зависимости, указывая, что понятие формально определено ранее с точки зрения понятий, которые указывают на него.

Сервисы, предоставляемые ЭБ, можно классифицировать как элементарные или составные [6]. Элементарные сервисы обеспечивают базовую инфраструктуру ЭБ. К таким сервисам относятся сбор, индексация, оценка и привязки. Составные сервисы могут быть составлены из других элементарных или составных сервисов, путем их многократного использования или их расширения. Например, сервисы поиска и просмотра используют сервисы индексирования и привязки.

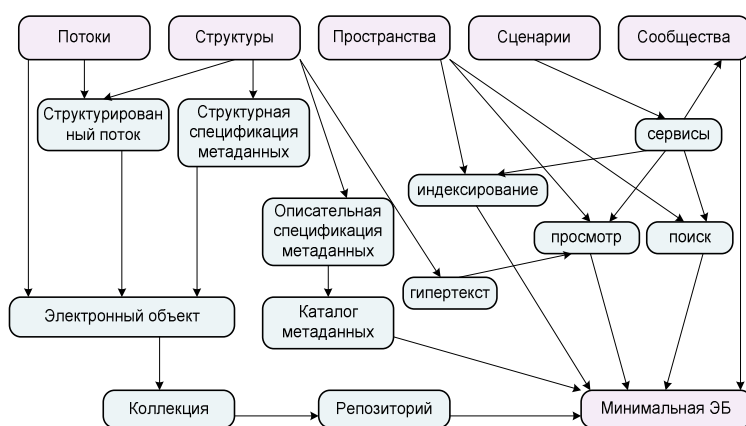


Рис. 5. Минимальная ЭБ в 5S

Таблица 2. Таксономия сервисов ЭБ онтологии 5S

Инфраструктурные сервисы		Добавление значений	Сервисы предоставления информации
Построение репозитория			
Создание	Хранение		
<u>Сбор данных</u> Авторство Каталогизация Краулинг Описание Оцифровка Харвестинг Снабжение Внесение	Сохранение Конвертация Копирование/Репликация Эмуляция Восстановление Преобразование (формата)	Аннотация <u>Классификация</u> Кластеризация <u>Настройка</u> Оценка Извлечение Индексирование Привязка <u>Протоколирование</u> <u>Измерение</u> Ранжирование Рецензирование Обзор Обучение (классификатор) Перевод (язык/формат)	<u>Присвоение</u> <u>Просмотр</u> Распространение Расширение (запрос) <u>Фильтрация</u> <u>Рекомендация</u> <u>Запрашивание</u> <u>Поиск</u> <u>Визуализация</u>

В табл. 2 приведена таксономия сервисов ЭБ [17]. В этой таксономии определяются фундаментальные сервисы (обозначенные полужирным шрифтом) такие как:

- те, которые помогают создавать элементы базовых понятий, входящих в минимальное определение ЭБ, такие как цифровые объекты, спецификации метаданных, коллекции и каталоги;

- те, которые принадлежат минимальному набору сервисов ЭБ (например, поиск и просмотр);

- те, которые поддерживают предыдущие сервисы с точки зрения расширения и многократного использования.

Подчеркиванием обозначены составные сервисы. В рамках данного проекта была разработана, на наш взгляд наиболее полная и проработанная, онтология ЭБ [17]. Кроме того на основе модели 5S было разработано ряд инструментов: Декларативный язык 5SL [15] на основе XML, для определения и генерации ЭБ; 5SQual [18] – инструмент для автоматической проверки качества ЭБ, ее объектов, метаданных и сервисов; 5SGen, осуществляющий генерацию кода для реализации ЭБ, используя ее определение на языке 5SL; 5SGraph – инструмент для моделирования ЭБ; 5SSuite – набор инструментов для поддержки генерации объединения ЭБ, включающий 5SGraph, 5SGen и SchemaMapper (который отображает отдельные схемы ЭБ, в общую схему объединенной ЭБ) [19].

Достоинством 5S является то, что это формальная модель с четкими математическими определениями, правилами и аксиомами. С другой стороны, на наш взгляд, модель слишком оторвана от реальности и слишком универсальна, хотя и позиционируется как модель ЭБ. Применение или "наложение" ее на реальную систему ЭБ неочевидно, хотя авторы приводят соответствующие примеры. Учитывая то, что модель развивается "снизу вверх" от минимальной модели, она охватывает меньше аспектов ЭБ по сравнению с моделью DELOS.

Заключение

В мире электронных библиотек существуют задачи, которые могут быть решены с помощью качественного концептуального описания таких систем. На сегодняшний день не существует всеохватывающей модели ЭБ, которую можно было бы с полным правом назвать эталонной. Построение хорошей модели, учитывающей мировой опыт подобных разработок, является нашей целью.

В работе представлен подробный анализ ряда известных проектов из этой области: CIDOC CRM, FRBR, FRBRoo, DELOS DLRM, 5S. На основании которого в настоящий момент проводится разработка оригинальной модели научной электронной библиотеки.

1. *Козаловский М.Р., Паринов С.И.* Информационные ресурсы, наукометрические показатели и показатели качества метаданных системы Соционет // Тр. IX Всероссийской конф. "Электронные библиотеки: перспективные методы и технологии, электронные коллекции". – RCDL'2007, г. Переславль-Залесский, Россия. – 2007. – С. 45–54.
2. *A Guide to Institutional Repository Software.* 3rd Edition. Open Society Institute. 2004. http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf
3. *Candela L., Castelli D., Fuhr N., Ioannidis Y., Klas C.-P., Pagano P., Ross S., Saidis C., Schek H.-J., Schuldt H., Springmann M.* Current Digital Library Systems: User Requirements vs Provided Functionality. IST-2002-2.3.1.12. Technology-enhanced Learning and Access to Cultural Heritage. March 2006.
4. *Crofts N., Doerr M., Gill T., Stead S., Stiff M. (editors),* Definition of the CIDOC Conceptual Reference Model, January 2008. Version 4.2.4.
5. *Functional Requirements for Bibliographic Records, Final Report / IFLA Study Group on the Functional Requirements for Bibliographic Records.* – München: K.G. Saur, 1998. (UBCIM Publications, New Series; v.19).<http://archive.ifla.org/VII/s13/frbr/frbr.htm>
6. *Candela L., Castelli D., Dobreva M., Ferro N., Ioannidis Y., Katifori H., Koutrika G., Meghini C., Pagano P., Ross S., Agosti M., Schuldt H., Soergel D.* The DELOS Digital Library Reference Model Foundations for Digital Libraries. IST-2002-2.3.1.12. Technology-enhanced Learning and Access to Cultural Heritage. Version 0.98, December 2007. http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf
7. *Goncalves M.A., Fox E.A., Watson L.T. and Kipp N.A.* Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries // ACM Trans. Inf. Syst. 22(2). – 2004. – P. 270–312.
8. The CIDOC Conceptual Reference Model. <http://cidoc.ics.forth.gr>
9. *Doerr M., Iorizzo D.* The Dream of a Global Knowledge Network – A New Approach // ACM J. on Computing and Cultural Heritage, June 2008. – Vol. 1, N. 1, Article 5.
10. *Tillett, Barbara B.* "Bibliographic Relationships." In: Relationships in the Organization of Knowledge, edited by Carol A. Bean and Rebecca Green. – Dordrecht: Kluwer Academic Publishers, 2001. – P. 19–35.
11. *Barbara Tillett* What is FRBR? A Conceptual Model for the Bibliographic Universe. <http://www.loc.gov/cds/downloads/FRBR.pdf>
12. *Функциональные требования к библиографическим записям: окончат. отчет / Рос. библио. ассоц., Рос. гос. б-ка ; пер. с англ. [В.В. Арефьев ; науч. ред. пер.: Т.А. Бахтурина, Н.Н. Каспарова, Н.Ю. Кулыгина].* – М.: ПГБ, 2006. – 150 с.
13. *Doerr M., Leboeuf P.* Modelling intellectual processes: The FRBR–CRM harmonization // In Conf. Proc. of ICOM-CIDOC Annual Meeting. Gothenburg, Sweden. – 2006. – P. 10–14.
14. *Fox E.A.* From Theory to Practice in Digital Libraries: 5S and Educational. Applications (NDLTD, CSTC) // Workshop on Digital Libraries, Albuquerque NM. – 1999.
15. <http://www.dlib.vt.edu/projects/5S-Model/>
16. *Murthy U., Gorton D., Torres R., Goncalves M.A., Fox E.A., Delcambre L.* Extending the 5S Digital Library (DL) Framework: from a minimal DL towards a DL Reference Model // ACM IEEE J. Conf. on Digital Libraries. – 2007. <http://fox.cs.vt.edu/talks/2007/20070623JCDLdlfWkshpMurthy-5S.ppt>
17. *Goncalves M.A., Fox E.A., Watson L.T.* Towards a digital library theory: a formal digital library ontology // Int. J. Conf. on Digital Libraries, 8(2). – 2008. – P. 91–114.
18. *Moreira B.L., Goncalves M.A., Laender A.H.F., Fox E.A.* 5SQual: a quality assessment tool for digital libraries // JCDL '07: Proc. of the ACM/IEEE-CS J. Conf. on Digital Libraries. – 2007. – P. 513–513.
19. *Shen R.* Applying the 5S Framework To Integrating Digital Libraries. Ph.D. Dissertation. Department of Computer Science, Virginia Tech. – 2006. http://scholar.lib.vt.edu/theses/available/etd04212006-135018/unrestricted/Rao_final_Defense_final5.pdf

Получено 17.03.2010

Об авторах:

Резниченко Валерий Анатольевич,
ведущий научный сотрудник,

Проскудина Галина Юрьевна,
научный сотрудник,

Кудим Кузьма Алексеевич,
младний научный сотрудник,

Овдий Ольга Михайловна,
младший научный сотрудник.

Место работы авторов:

Институт программных систем
НАН Украины,
03187, Киев-187,
Проспект Академика Глушкова, 40.
Тел. +38(044)526 6033

e-mail: reznich@isofts.kiev.ua
gupros@isofts.kiev.ua
kuzma@isofts.kiev.ua
olga.ovdiy@gmail.com