

Сравнение систем электронных библиотек EPrints 3.0 и DSpace 1.4.1

© Кудим К.А.

Проскудина Г.Ю

Резниченко В.А.

Институт программных систем НАН Украины
kuzma@isofts.kiev.ua

gupros@isofts.kiev.ua

reznich@isofts.kiev.ua

Аннотация

Рассматриваются основные возможности и особенности DSpace и EPrints как наиболее популярных свободно распространяемых систем построения электронных библиотек. Описан опыт построения многоязычных электронных библиотек на их основе. Приводится сравнительный анализ систем DSpace 1.4.1 и EPrints 3.1. Особое внимание уделено проблемам локализации, совместимости внешних форматов и удобству использования данных систем.

1 Введение. Почему EPrints и DSpace?

Данная работа продолжает цикл публикаций [1,2], посвященных созданию электронных библиотек (ЭБ) на основе свободно распространяемого программного обеспечения с открытым исходным кодом. Итогом данных работ стала разработка двух типовых институтских библиотечных веб-сайтов – <http://greenstone.isofts.kiev.ua>, созданного на основе программного обеспечения (ПО) Greenstone и <http://eprints.isofts.kiev.ua> на основе ПО EPrints.

За прошедший период в Институте программных систем НАН Украины изучена и опробована еще одна популярная система для построения научных библиотек – DSpace. Известно, что в России созданы и функционируют на основе DSpace репозитории Уральского <http://dspace.lib.usu.ru> (общее количество записей – 1265) и Красноярского государственных университетов <https://elib.krasu.ru> (общее количество записей – 422). В Украине – «Открытый электронный архив гражданского общества» www.e-archive.org.ua, электронные репозитории Украинского католического университета (г. Львов) <http://dspace.ucu.edu.ua>, Института биологии южных морей НАН Украины (г. Севастополь) <http://repository.ibss.org.ua/dspace/>, а также Центра гуманитарных исследований Львовского Национального университета им. И. Франко <http://www.dspace.humanities.org.ua/dspace/>.

В настоящее время электронные коллекции,

созданные с помощью DSpace, доступны в локальной сети нашего института, которые в дальнейшем предполагается открыть для всеобщего обозрения и доступа. Результатом проведенной работы и явилась настоящая статья.

В настоящее время все больше и больше научных и образовательных материалов создаются в электронной форме. Как следствие, все большему количеству учебных заведений и научных организаций требуется надежное место для хранения и легкого доступа к таким материалам. Множество статей, отчетов, экспериментальных данных, медиаданных, созданных в каком-либо подразделении организации, сохраняется на локальном жестком диске или веб-сервере подразделения. Такие данные часто теряются навсегда после реструктуризации организации. Более того, множество публикуемых работ становится доступным только после длительного периода времени с момента их реального завершения. Все это объясняет важность создания систем электронных библиотек.

Существует немало систем такого рода как коммерческих, так и бесплатных. Лидирующие позиции в этой отрасли занимают два быстро развивающихся проекта DSpace [3] и EPrints [4]. По данным [5] на июль месяц 2007 года в мире функционирует 245 электронных репозиториях (архивов, библиотек), созданных с использованием ПО DSpace и 230 – EPrints. Наиболее крупный из них – DSpace at Cambridge, насчитывающий 188395 записей.

Версии последних релизов на момент написания статьи были EPrints 3.0 и DSpace 1.4.1. Разработка систем проходит при активном участии их пользователей.

Данные системы являются представителями одного класса и имеют много общего. Обе являются системами с открытыми исходными кодами, OAI-совместимыми [6,7], интероперабельными, эквивалентными по функциональности самоархивирования, и даже написаны они изначально были одним и тем же программистом из Саутгемптона Робом Тэнсли (Rob Tansley).

Представленный в работе сравнительный анализ EPrints 3.0 и DSpace 1.4.1. может быть полезен как для специалистов, выбирающих систему электронных библиотек для своего учреждения, так и разработчикам подобных систем, чтобы увидеть

возможные пути усовершенствования программного обеспечения.

Работа структурирована следующим образом. Раздел 2 содержит историю создания систем и описание подобных проектов. Далее следует описание некоторых аспектов функциональности систем (раздел 3); архитектурных отличий (раздел 4); отличий внешнего использования (раздел 5); технических аспектов систем (раздел 6).

2 История создания и подобные проекты

DSpace. Платформа электронных библиотек DSpace разрабатывалась совместно компанией Hewlett-Packard и библиотеками MIT (Massachusetts Institute of Technology). 4-го ноября 2002 года система была запущена как действующая служба, поддерживаемая библиотеками MIT. Также на основании лицензии BSD [8] был открыт исходный код с намерением поощрить формирование сообщества открытых кодов вокруг DSpace.

DSpace формировалась под влиянием научных исследований в области систем электронных библиотек. Архитектура DSpace происходит из фреймворка Кана и Виленского для сервисов распределенных электронных объектов [9], а так же из работ Армса и других по архитектуре электронной библиотеки [10,11]. DSpace основывается также на работе Лагозе и других по архитектуре FEDORA [12] и прототипе реализации FEDORA в университете Вирджинии [13]. Другая существенная часть работы по созданию DSpace связана с OAI (Open Archival Information System) [14], где определены все основные понятия и термины, используемые в DSpace.

EPrints. Благодаря своему происхождению из движения «Общение в мире науки» (Scholarly Communication), конфигурация EPrints по умолчанию направлена на поддержку научных статей, но может быть адаптирована для других целей и содержимого. Система разработана в группе Интеллекта, агентов и мультимедиа в отделе Электроники и компьютерных наук Саутгемптонского университета в 2000 году. EPrints свободно распространяется под лицензией GNU General Public License [15].

EPrints имеет много сходства с DSpace, но оптимизирована так, чтобы обеспечить доступ к самостоятельному депонированию (внесению) автором материалов (статей, отчетов, книг и пр.), в то время как DSpace обеспечивает платформу, для долгосрочного хранения цифровых материалов, используемых в академических исследованиях [16]. Интерфейс пользователя для депонирования материалов в DSpace основан на опыте, полученном из разработки и использования интерфейса пользователя для депонирования системы EPrints. Интероперабельность с EPrints, в настоящее время частично может достигаться через использование протокола OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) для сбора метадан-

ных [7], обеспечивающего сервис доступа к разным архивам.

Среди систем того же класса можно также назвать ПО Greenstone Новозеландского проекта по ЭБ в Университете Waikato [17] – открытый инструмент создания ЭБ, главным образом сосредоточившийся на централизованной публикации документов (и других объектов коллекционирования), кроме того, CERN разработал ПО Сервер Документов CERN (CDSware) [18], который является сервером предварительной публикации научных документов.

3 Обзор функциональности

DSpace функционирует как централизованный сервис организации. Разные подразделения в пределах учреждения (лаборатории, центры, школы или отделы) могут иметь свои собственные отдельные области в пределах системы. Члены этих подразделений непосредственно вносят контент через веб-интерфейс пользователя, который разработан так, что внесение осуществляется максимально просто. Альтернативно система предусматривает импорт множества элементов для пакетной загрузки контента.

В каждом подразделении можно также назначить людей, которые могут просматривать и редактировать внесения перед тем, как они будут включены в основной репозиторий. Затем DSpace индексирует метаданные, поступившие вместе с электронным документом, и делает их доступными согласно привилегиям доступа, определенным в данном подразделении.

Функциональные аспекты DSpace [19]:

- для базовой организации данных зафиксирована определенная *модель данных*;
- система хранит и индексирует *метаданные* в разнообразных форматах;
- система хранит информацию о *пользователях* системы;
- несмотря на то, что большие усилия прилагаются для облегчения доступа к электронным материалам учреждения, тем не менее открывать полный доступ к содержимому репозитория не всегда целесообразно. Кроме того, такие функции как депонирование и редакторская проверка должны быть привязаны к соответствующим пользователям и ограничиваться ими. Поэтому система обладает функцией *авторизации*;
- система должна быть способна принимать приходящие материалы, такой процесс называется *поглощением* [14];
- в некоторых случаях требуется, чтобы материалы или связанные с ними метаданные, вносимые в архив, были проверены или дополнены назначенными пользователями. Эта последовательность действий называется *рабочим процессом (workflow)*;

- материалы в архиве доступны по ссылкам, приведенным в описании конкретного элемента. По этому же описанию можно делать библиографические ссылки на данный материал;
- конечные пользователи должны иметь возможность просматривать и находить содержимое репозитория. В связи с этим система должна обеспечивать функции поиска и просмотра (навигации);
- для предоставления возможности интегрированного поиска документов поддерживается протокол сбора метаданных OAI-PMH;
- должна существовать возможность уведомления конечных пользователей системы об интересующих свежих поступлениях репозитория, чтобы избежать их от необходимости регулярно посещать репозиторий для выполнения такой проверки. Система предоставляет автоматическую рассылку уведомлений по электронной почте через службу подписки;
- предоставляется возможность обрабатывать данные произвольных форматов, от простых текстовых документов до наборов данных и цифрового видео;
- доступ к перечисленным функциональным возможностям предоставляется посредством веб-интерфейса.

EPrints также обеспечивают подобную функциональность, необходимую для управления электронным репозиторием, решая задачу длительного хранения и доступа, и предоставляя основу для дальнейшего развития системы [2]. Рассмотрим более подробно некоторые из приведенных аспектов, где возможно показывая отличительные особенности DSpace и EPrints.

4 Архитектурные отличия

В этой части рассмотрим некоторые стороны архитектурных отличий: модель данных, формат файлов, формат и экспорт/импорт метаданных.

4.1 Модель данных

DSpace. Способ организации данных в DSpace выбран таким образом, чтобы отразить структуру организации, использующей данную систему (рис.1). Каждый сайт DSpace делится на разделы, соответствующие подразделениям организации. Раздел является самым высоким уровнем иерархии DSpace. Разделы могут содержать подразделы, т.е. могут образовывать иерархию. Разделы содержат коллекции логически связанных материалов. Каждая коллекция может принадлежать только одному разделу. Коллекция состоит из *элементов*, которые являются основной единицей или «атомом» архивирования. Элемент принадлежит одной и только одной коллекции (связь *принадлежит*), но

дополнительно может быть отображен еще и в других коллекциях (связь *включает*). Элемент состоит из сгруппированного, связанного между собой содержимого и соответствующих описаний (метаданных). Метаданные, описывающие элемент, индексируются для навигации и поиска. Далее, элементы, в свою очередь, представляют собой *наборы* (связки) *битовых потоков* (файлов). Цель подобных наборов – хранить тесно связанные файлы вместе. В табл. 1 приведены примеры для каждого типа объекта модели данных [3,19].

Таблица 1. Примеры объектов DSpace

Тип объекта	Пример
Раздел (Community)	Институт программных систем
Подраздел (Sub-community)	Отдел компьютерных вычислений
Коллекция (Collection)	Отчеты Публикации конференций
Элемент (Item)	Статья «Сравнение систем электронных библиотек EPrints 3.0 и DSpace 1.4.1»
Набор (Bundle)	HTML файл и связанная картинка, представляющие единый HTML-документ
Файл (Bitstream)	Отдельный HTML файл; отдельный файл изображения
Формат файла (Bitstream Format)	Версия Microsoft Word; изображение в формате JPEG

Каждый элемент имеет запись метаданных стандарта квалифицированного Дублинского ядра (Qualified Dublin Core – QDC). Другие метаданные могут сохраняться в элементе в виде текстового файла, но обязательным набором метаданных для каждого элемента является QDC, что обеспечивает интероперабельность и простоту нахождения элементов. Метаданные QDC могут быть введены конечными пользователями вместе с самими документами или они могут быть получены из других метаданных как часть процесса поглощения (ingest process).

EPrints. Здесь нет такого строгого структурного деления на разделы и коллекции, которые все же играют важную роль, например, для сужения области поиска по репозиторию. Идея модели данных EPrints заключается в том, что все записи эквивалентны и не составляют иерархию. Тем не менее, иерархия необходима для навигации по репозиторию, поскольку пользователь может точно не знать цель своих поисков, имея лишь приблизительное представление о ней. В EPrints эта задача решена с помощью обзоров или представлений (views) – способ генерировать навигацию любого необходимого типа, используя связанные с элементами поля метаданных, т.е. обзор может осуществляться по Подразделениям организации, либо по Автору, либо, более сложный вариант, по Году публикации, а затем по Типу и т.п.

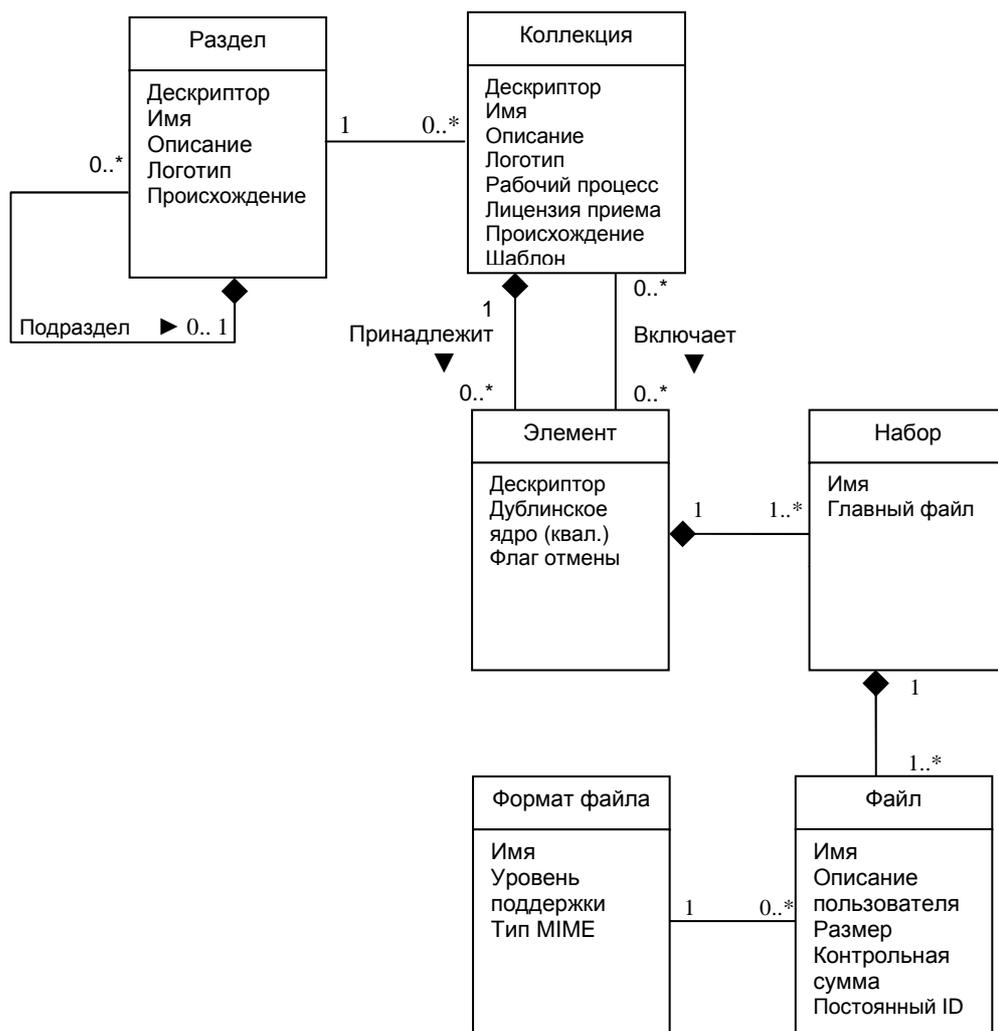


Рис. 1. Модель данных DSpace

Таким образом, в модели данных EPrints можно обеспечить гибкую поддержку иерархической тематической классификации (по умолчанию, Классификации библиотеки конгресса) и дерева подразделений организации.

Такие объекты, как элемент, набор файлов, файл, сходны с аналогичными в DSpace. Элемент тоже является основной единицей хранения и содержит все метаданные, предоставляемые для внешнего использования. Отличительной чертой EPrints является возможность динамически генерировать метаданные в различных форматах из внутреннего представления. Еще одно значительное отличие состоит в том, что все типы хранимых материалов классифицированы (книга, статья, диссертация и т.д.) и каждому типу поставлен в соответствие набор внутренних полей метаданных.

Таким образом, объекты типа элемент сходны в обеих системах. Это соответствует модели, используемой в OAI-PMH, которая вкратце может быть описана как *ресурс-элемент-запись*. Ресурс может представлять собой как традиционный для

библиотек объект (книга, статья), так и другие объекты (изображение, фильм). Элемент является такой компонентой репозитория, из которой распространяются метаданные о ресурсе. Элемент хранит или динамически генерирует метаданные об отдельном ресурсе в множестве форматов, каждый из которых может быть собран в виде записи посредством протокола OAI-PMH.

Иерархическая структура элементов существенно различна. DSpace использует более жесткую систему, хотя она и покрывает большинство нужд репозитория. EPrints позволяет создавать более сложные иерархии на основе разных внешних представлений. Можно сказать, что модель данных EPrints более универсальна.

4.2 Форматы файлов

Каждый хранимый в системе файл связан с определенным форматом. Так как служба хранения является основополагающей функцией библиотечной системы, то для таких систем является существенной возможность распознавать конкретные форматы

файлов, загружаемые пользователем. Неотъемлемой частью формата файла является явно или скрыто выраженный способ интерпретации его содержимого.

Список поддерживаемых форматов файлов для обеих систем приведен в табл. 2. Как видно из таблицы системы позволяют сохранять и предоставить доступ к большинству наиболее известных типов файлов [20].

Таблица 2. Поддерживаемые форматы файлов

EPrints 3.0	DSpace 1.4.1	
HTML	Adobe PDF	MPEG
PDF	AIFF	MPEG Audio
Postscript	audio/basic	Photo CD
Простой текст	BMP	Photoshop
MS PowerPoint	FMP3	Postscript
MS Word	GIF	RealAudio
Изображение (JPEG)	HTML	RTF
Изображение (PNG)	image/png	SGML
Изображение (GIF)	JPEG	TeX
Изображение (BMP)	LateX	TeX dvi
Изображение (TIFF)	MARC	Text
Видео (MPEG)	Mathematica	TIFF
Видео (QuickTime)	Microsoft Excel	Video Quicktime
Видео (AVI)	Microsoft Powerpoint	WAV
	Microsoft Project	WordPerfect
	Microsoft Visio	XML
	Microsoft Word	

В DSpace-системе каждый хранимый файл дополнительно характеризуется такими уровнем поддержки:

- поддерживаемый формат распознается, и организация предоставляющая услуги по размещению информации уверена, что файл этого формата можно будет использовать в будущем, используя любую комбинацию методов, соответствующих тем или иным потребностям;
- известный формат распознается, и организация предоставляющая услуги по размещению информации может сохранять файл как есть и извлекать его в том же формате. Организация, предоставляющая услуги по размещению информации, попытается получить дополнительную информацию, чтобы модернизировать формат до «поддерживаемого» уровня;
- неподдерживаемый формат нераспознаваем, но организация предоставляющая услуги по размещению информации сохраняет файл как есть и позволяет извлекать его из архива в том же формате.

4.3 Метаданные

DSpace поддерживает три вида метаданных об архивируемом контенте.

Описательные метаданные. Каждый элемент архива имеет одну запись метаданных в формате QDC, который не жестко основан на наборе элементов и квалификаторов профиля библиотечных приложений [21,22], обеспечиваемого в системе DSpace по умолчанию. Организации с другими требованиями могут его легко изменять, через реестр элементов и квалификаторов.

Другие описательные наборы метаданных, например UNIMARC, могут поддерживаться в обычных текстовых файлах. Разделы и коллекции имеют несколько простых описательных полей метаданных (имя и несколько описательных предложений), хранящихся в СУБД.

Административные метаданные включают метаданные сохранности (preservation), происхождения (provenance) и политики авторизации данных. Большинство из них хранится в СУБД системы DSpace. Метаданные происхождения хранятся в записях DC. Дополнительно некоторые другие административные метаданные (например, размер файла или MIME-типы) копируются в записях DC так, чтобы они были легко доступны через протокол OAI.

Структурные метаданные включают информацию о том, как представлять элемент или файлы внутри элемента пользователям и отношения между составляющими частями элемента. В качестве примера, рассмотрим диссертацию, состоящую из набора TIFF-изображений, каждое из которых представляет отдельную страницу диссертации. Структурные метаданные должны включать информацию, что каждое изображение есть одна страница и порядок изображений/страниц. Структурные метаданные в DSpace занимают довольно важное место. Внутри элемента, как было описано выше, файлы могут быть упорядочены в отдельные наборы. Набор может иметь *главный файл* (primary bitstream). Дополнительные структурные метаданные могут сохраняться в простых текстовых файлах.

EPrints использует различные типы записей, описывающих документ. С каждым типом документа связан отдельный набор полей метаданных (подмножество множества всех полей метаданных EPrints). Набор включает в себя только записи, используемые конкретным типом документа. Веб-страницы генерируются таким образом, чтобы отображались только те поля метаданных, которые соответствуют своему типу документа.

В EPrints выделены следующие типы документов:

- статья в журнале, газете, не обязательно рецензированная, возможно существующая только в электронном виде, как то в онлайн-журнале или на новостном веб-сайте;
- книга или том конференции;
- раздел книги или глава книги;
- монография, возможно технический отчет,

отчет по проекту, документация, руководство, протокол дискуссии;

- доклад конференции или семинара – документ, постер, речь, лекция или презентация с конференции, семинара или другого события;
- набор данных – связанная коллекция количественных данных;
- обучающий ресурс – записи лекций, упражнения, экзаменационные документы или программы курсов;
- другие – что-то в рамках репозитория, но не покрываемое другими категориями.

Для целостности репозитория некоторые поля метаданных являются обязательными и, следовательно, должны быть заполнены. Каждое поле ввода сопровождается детальным справочным описанием.

4.4 Экспорт/Импорт

EPrints поддерживает целый спектр наборов метаданных. Среди них есть Dublin Core, который считается обязательным в OAI-PMH. Для общедоступных элементов репозитория EPrints представляет наружу их метаданные в формате DC. Если какие-либо службы OAI запрашивают другой формат метаданных, например MODS [23], система может корректно обработать и этот запрос.

Данные из EPrints могут экспортироваться в следующих форматах метаданных:

- BibTeX библиографический формат метаданных;
- OpenURL ContextObject – стандарт метаданных ANSI/NISO Z39.88-2004 [24] для контекстнозависимых сервисов, обычно полнотекстового поиска;
- OpenURL Dissertation – тот же стандарт, специализированный для ресурсов типа диссертации;
- OpenURL Journal – тот же стандарт, специализированный для ресурсов типа журнал;
- Dublin Core - Дублинское ядро, стандарт метаданных ANSI/NISO Z39.85-2001 (а также стандарт ISO 15836-2003) [21];
- DIDL - Digital Item Declaration Language, при помощи которого в MPEG-21 описываются сложные электронные объекты [23];
- EndNote – распространенный в научном сообществе библиографический формат ссылок цитирования, используется в одноименном коммерческом продукте [25];
- HTML Citation – HTML-формат цитирования для документов, используемый для просмотра или поиска документов в Eprints 3.0;
- METS - стандарт кодирования и передачи метаданных [26];
- MODS - схема метаданных описания объекта;
- Reference Manager – формат метаданных для

создания и управления архивами и библиографическими описаниями, экспорт в этот формат позволит использовать метаданные Eprints 3.0 в системе Reference Manager (системе того же класса, что и EndNote);

- Refer – формат, построенный в соответствии со специально отформатированным документом (troff), может использоваться практически любой программой и является довольно обобщенным форматом библиографий [27];
- Simple Metadata (SimpleMDE) – этот набор метаданных является подмножеством полного возможного набора метаданных и используется, когда выполняется быстрая аннотация [28];
- ASCII Citation – обычный текстовый формат;
- EP3 XML – экспорт в XML.

DSpace. Для реализации функций экспорта и импорта в системе предусмотрены плагины переходов (crosswalk plugins). Это программные модули, осуществляющие перевод между метаданными DSpace-объектов и определенным внешним представлением. Как правило это пары плагинов для импорта и экспорта. Например, из формата метаданных MODS во внутренний формат DSpace и наоборот. Используемые плагины перечислены в конфигурационном файле.

Инсталляция DSpace 1.4.1 включает следующие плагины переходов, используемые в рамках протокола OAI-PMH:

- METS - стандарт кодирования и передачи метаданных.
- MODS - схема метаданных описания объекта.
- QDC – квалифицированное Дублинское Ядро, которое, как было упомянуто выше, является основным набором метаданных системы DSpace [22].
- DIDL – Digital Item Declaration Language.

5 Использование

В этом разделе рассмотрим роли пользователей, процесс депонирования и внешнее использование, т.е. просмотр и поиск.

5.1 Пользователи, группы и права

EPrints. Изначально конфигурация EPrints представляет четыре группы пользователей определяемые правами доступа:

- минимальный пользователь может просматривать содержимое репозитория, подписываться на списки рассылки, создавать хранимые поиски;
- депонент обладает правами минимального пользователя, владеет собственной рабочей областью, куда он может загружать

- элементы, и может подавать элементы из рабочей области редактору на рассмотрение;
- редактор обладает правами депонента, и может принимать, отклонять или удалять элементы, поданные пользователями для размещения в репозитории;
- администратор обладает правами редактора, а также может управлять пользовательскими учетными записями и элементами репозитория.

Существует возможность конфигурировать права доступа для каждой группы пользователей, например ограничивать темы, доступные отдельному редактору для проверки, или пропускать шаг редакторской проверки в случае, если депонирование инициируется хорошо известной доверенной группой пользователей. Такая конфигурация может быть осуществлена посредством редактирования соответствующих файлов. В процессе депонирования для каждого элемента выбирается, будет ли он общедоступным, либо разрешение просматривать его будет дано только зарегистрированным пользователям, или только редакторам и администраторам (т.е. только штату репозитория).

DSpace. Многие функции DSpace, например просмотр и поиск документов в системе, могут выполняться анонимно, но чтобы выполнить внесение документов пользователю нужно зарегистрироваться.

О каждом пользователе DSpace хранит следующую информацию:

- адрес электронной почты;
- фамилию и имя;
- пароль;
- список коллекций, относительно которых пользователь уведомляется о новых поступлениях;
- признак саморегистрации, т.е. создана ли была учетная запись пользователя автоматически системой либо, например, в результате обращения к администратору.

DSpace обладает более продвинутой системой прав пользователя по сравнению с EPrints, которая тесно связана с применяемой моделью данных. Выделены следующие группы пользователей: депоненты, администраторы, участники процесса депонирования, подписчики и пользователи с разрешением просматривать не общедоступные элементы. Группы пользователей можно также рассматривать в качестве ролей.

Гибкая система прав доступа в DSpace позволяет ограничивать доступ к различным частям архива. Каждому разделу архива можно назначить группу пользователей, которым разрешается доступ к данному разделу. Каждой коллекции назначается множество из отдельных пользователей и групп, которые будут депонентами для этой коллекции, будут иметь доступ к содержимому, играть роль

редакторов, или, наконец, администрировать коллекцию.

Пользователь может быть ассоциирован с несколькими группами одновременно. Каждый пользователь группы наделяется соответствующими правами. Управление группами и отдельными пользователями осуществляется посредством веб-интерфейса и не требует, во-первых, специальных программистских навыков, и, во-вторых, доступа к операционной системе, где установлено программное обеспечение DSpace. Аналогично, редактирование прав доступа к разделу или коллекции осуществляется посредством веб-интерфейса. Конечно, чтобы иметь доступ к функциям управления, необходимо войти в систему с правами администратора.

Гибкая система авторизации DSpace основана на привязке действий к объектам (*политики ресурсов*) и спискам пользователей (групп), которые могут их выполнять. Существует две специальные группы – «администраторы», где пользователи могут совершать любые действия, и «анонимный», которая содержит всех пользователей системы. Назначение политики действия для элемента группе пользователей «анонимный» означает, что каждый пользователь системы может совершать это действие. Например, большинство элементов DSpace имеют политику для анонимных пользователей **ЧТЕНИЕ**.

Чтобы пользователь мог выполнить действие с объектом системы, он должен иметь разрешение, которое должно быть задано явно. В табл. 3 и 4 соответственно представлены перечень возможных действий, которые понимает система авторизации и политики ресурсов.

Таблица 3. Возможные действия в DSpace

ЧТЕНИЕ (READ)	Действие связано с тем, что пользователь узнает о существовании объекта в системе, и просматривает какие-либо связанные с ним метаданные
ЗАПИСЬ (WRITE)	Изменение метаданных, связанных с объектом. Здесь не включена возможность удаления
ДОБАВИТЬ (ADD)	Действие добавления объекта (например, элемента) в контейнер (например, в коллекцию). Чтобы внести элемент в коллекцию, пользователь должен иметь разрешение ДОБАВИТЬ в этой коллекции
ИЗЪЯТЬ (REMOVE)	Действие, связанное с изъятием объекта из контейнера

Обращаем внимание, что здесь нет действия **УДАЛИТЬ**. Для того, чтобы «удалить» объект из архива (например, элемент) нужно иметь разрешение **ИЗЪЯТЬ** на все объекты (в данном случае, коллекцию), которая его содержит.

В идеале, ничто поступающее в систему DSpace не должно когда-либо удаляться; однако, практические и юридические факторы иногда требуют удаления. Элементы могут быть удалены из

DSpace одним из двух способов. Они могут быть «изъяты», что означает, что они остаются в архиве, но полностью скрыты для обозрения. В этом случае, если пользователь пытается получить доступ к изъятому элементу, ему посылается сообщение, указывающее, что элемент был удален. В другом случае, элемент может также быть «удален», когда все его следы удалены из архива.

Таблица 4. Политики ресурсов в DSpace

Раздел (Community) ADD/REMOVE	Добавить или изъять коллекции или подразделы
Коллекция (Collection) ADD/REMOVE	Добавить или изъять элементы (ADD=разрешение на занесение элемента)
DEFAULT_ITEM_READ	Наследуется как READ всеми вносимыми элементами
DEFAULT_BITSTREAM_READ	Наследуется как READ файлами всех вносимых элементов
COLLECTION_ADMIN	Администраторы коллекции могут редактировать элементы в коллекции, изымать, отображать другие элементы в этой коллекции.
Элемент (Item) ADD/REMOVE	Добавить или изъять наборы
READ	Можно просматривать элемент (метаданные элемента всегда видны)
WRITE	Можно модифицировать элемент
Набор (Bundle) ADD/REMOVE	Добавить или изъять файлы в наборе
Файл (Bitstream) READ	Просмотреть файл
WRITE	Модифицировать файл

Делая общий вывод, можно сказать, что с точки зрения ролей и прав доступа EPrints лучше подходит для однородных репозиториях, где не имеют значения права пользователя, выходящие за рамки обычного, т.е. не требуется назначение прав разнообразным группам пользователей для доступа к различным частям репозитория. Такая система проста и не требует настройки прав пользователей. Что же касается DSpace, то эта система обладает более гибкой системой прав доступа, которая позволяет ограничивать доступ к различным частям архива. Возможность администрирования и управления процессом депонирования посредством веб-интерфейса является удобной и простой в использовании в DSpace.

5.2 Депонирование

EPrints. Когда новый пользователь регистрируется в системе, ему выделяется отдельная рабочая область, куда он может загружать свои документы. В процессе депонирования (занесения) нового

элемента пользователь должен выполнить следующие шаги (здесь рассматривается последовательность действий при изначальной конфигурации):

1. Выбрать тип документа.
2. Загрузить файлы, при необходимости создать наборы файлов.
3. Ввести описание элемента.
4. Продолжить описание – выбрать тему, соответствующую элементу.
5. Подать элемент на редакторскую проверку, принимая лицензионное соглашение.

На любом шаге процесс депонирования элемента можно прервать без потери данных, введенных на предыдущих шагах. Некоторые поля являются обязательными, что означает невозможность подачи элемента на редакторскую проверку в случае их незаполнения.

DSpace. По аналогии с EPrints, у пользователя есть рабочая область, где хранятся загруженные документы. Последовательность депонирования в DSpace описана ниже (конфигурация по умолчанию):

1. Выбрать коллекцию, которой элемент будет принадлежать.
2. Выбрать опции, влияющие на множество полей, доступных для ввода на последующих шагах.
3. Ввести значения основных полей метаданных.
4. Ввести ключевые слова, классифицирующие тему элемента, а также ввести дополнительные поля метаданных.
5. Загрузить файлы.
6. Проверить, все ли в порядке с загруженными файлами и их форматом; отредактировать в случае наличия ошибок.
7. Проверить, все ли в порядке во всех данных, введенных ранее; отредактировать в случае наличия ошибок.
8. Принять лицензионное соглашение.

Подобно EPrints, прерывание процесса депонирования безопасно с точки зрения сохранности введенных данных.

В основном, последовательность депонирования совпадает для обеих систем. EPrints обладает немного более дружелюбным интерфейсом, так как, во-первых, набор полей ввода не определяется на дополнительном шаге выбором соответствующих опций, а строится по выбранному типу документа; и, во-вторых, количество шагов меньше из-за лучшей группировки полей ввода. Однако, система DSpace более производительна, поэтому процесс депонирования здесь в целом проходит быстрее.

5.3 Процесс поглощения и занесения

На рис. 2 проиллюстрирован процесс поглощения в системе DSpace. Приложение «Пакетный импорт элементов» (Batch Item Importer) преобразовывает внешний SIP (Submission Information Package – пакет

подачи информации) [14], XML-документ метаданных и несколько файлов контента, в объект «процесса приема» (In Progress Submission). Аналогично для сборки объекта «процесса приема» используется передача элементов конечным пользователем через веб-интерфейс.

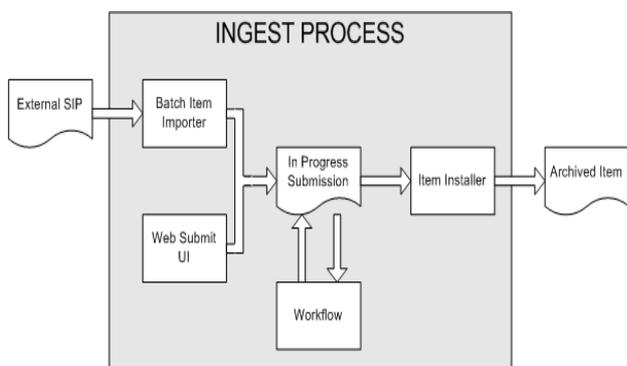


Рис. 2. Процесс поглощения в DSpace

В зависимости от политики коллекции, в которую нацелено внесение, может стартовать рабочий процесс (workflow). Что позволяет редакторам или контролерам осуществлять проверку вносимых материалов и гарантировать его пригодность для включения в коллекцию.

Когда завершен процесс внесения пакетный или через веб-интерфейс, то включается следующий шаг поглощения – рабочий процесс (workflow) либо «инсталлятор элемента» (Item Installer). В DC добавляется сообщение о «происхождении» (provenance message), которое включает имена файлов и контрольные суммы вносимого контента. Подобным же образом, всякий раз, когда рабочий процесс изменяет состояние (например, редактор разрешил внесение), добавляется аналогичное предложение «происхождения». Это позволяет проследить, как элемент изменился с тех пор, как его внес пользователь. Также в DSpace применяется система историй (History system), однако «происхождение» облегчает получение доступа в конкретный момент.

Как только успешно завершен рабочий процесс, объект «процесса приема» поглощается «инсталлятором элемента», который конвертирует его в полностью соответствующий репозиторию DSpace элемент. «Инсталлятор элемента»:

- назначает дату внесения;
- добавляет значение метаданных “date.available” в запись метаданных DC;
- добавляет дату публикации, если ничего другого нет;
- добавляет, сообщение «происхождения» (включая контрольные суммы файла);
- назначает постоянный идентификатор дескриптора (Handle);
- добавляет элемент в целевую коллекцию, и добавляет соответствующую политику авторизации;

- добавляет новый элемент к индексам поиска и просмотра.

5.4 Внешнее использование

DSpace по умолчанию не обладает таким красивым интерфейсом как EPrints. Однако, на основе DSpace существуют репозитории с прекрасным дизайном. DSpace позволяет конечным пользователям обнаруживать контент множеством способов:

- по внешней ссылке, например, дескриптору (Handle);
- поиск по одному или нескольким ключевым словам в метаданных или извлеченном полном тексте;
- просмотр по индексам названия, автору, дате и предметной теме.

Поиск – существенный компонент обнаружения объектов в любой библиотечной системе. Учитывая высокие пользовательские требования к поисковому машинному, предоставление как можно большего числа поисковых возможностей – важная особенность DSpace. Модуль индексации и поиска DSpace имеет очень простой API, который предоставляет возможность индексации нового контента, регенерации индекса и выполнения поиска по всему архиву, разделу, подразделу или коллекции. За API стоит свободно распространяемая поисковая Java-машинка Lucene. Lucene предоставляет поиск по полям метаданных, удаление стоп-слов, выполнение поиска слова во всех его морфологических формах, возможность инкрементно добавлять новый индексируемый контент без регенерации всего индекса. Индексы поиска конфигурируемы, что позволяет организациям настраивать индексируемые поля метаданных. Система предоставляет простой и расширенный поиск (рис. 3).



Рис. 3. Главная страница раздела

Другой важный механизм обнаружения материалов в DSpace является просмотр – процесс, где пользователь просматривает конкретный индекс, например индекс названия, и в процессе поиска интересующих элементов проводит по нему навигацию. Подсистема просмотра обеспечивает простой API, где предоставляется возможность определить индекс и его подраздел. Затем

подсистема просмотра раскрывает часть интересующего индекса. Индексы, которые могут быть просмотрены: название элемента (рис. 4), автор элемента, дата выпуска элемента и предметные термины. Дополнительно, просмотр может ограничиваться до элементов внутри определенной коллекции или раздела.



Рис. 4. Просмотр по названиям

EPrints. Последняя версия EPrints имеет привлекательный и удобный веб-интерфейс. При модификациях в настройках системы обычно остаются неизменными следующие возможности:

- доступно несколько типов навигации по репозиторию. Изначальная конфигурация включает просмотр по году издания и по теме;
- простой поиск;
- расширенный поиск, где для многокритериального поиска представлено множество полей метаданных. Здесь же доступен полнотекстовый поиск;
- каждое поле ввода сопровождается кнопкой с изображением вопросительного знака, щелчок по ней открывает краткое пояснение о том, что должно быть введено в это поле;
- доступна лента RSS по содержимому архива;
- можно просмотреть последние поступления.

6 Технические различия

6.1 Установка и необходимое программное обеспечение

Установка и исходная конфигурация для обеих систем может быть выполнена в течение одного рабочего дня при наличии:

1. Опыта установки программного обеспечения на базовой операционной системе.
2. Собственно установочных файлов системы электронных библиотек.
3. Другого необходимого программного обеспечения (см. табл. 5).

EPrints. После установки всего программного обеспечения, необходимого для функционирования системы, исходная конфигурация и установка EPrints производится с помощью одного интерактивного скрипта. В результате создается первый репозиторий с конфигурацией по умолчанию. Кроме того, несколько периодических заданий нужно добавить в таблицу cron: для полнотекстовой индексации, отправки уведомлений по электронной почте и генерирования обзоров по репозиторию.

DSpace. После установки программного обеспечения, необходимого для DSpace, систему нужно распаковать и сконфигурировать, отредактировав конфигурационный файл. После чего добавить задания в cron таблицу: периодическая очистка базы данных, индексация, рассылка почты и сбор статистики.

Таблица 5. Необходимое программное обеспечение

	EPrints	DSpace
<i>Операционная система</i>	Unix- подобная	Unix- подобная
<i>Веб-сервер</i>	Apache (включая mod_perl)	Apache Tomcat или эквивалент
<i>Сервер баз данных</i>	MySQL	PostgreSQL или Oracle
<i>Библиотеки языков программирования</i>	Perl с модулями perl-Unicode-String, perl-XML-LibXML	Java, Apache Ant

6.2 Поддержка многоязычности

EPrints. Почти все языко-зависимые части интерфейса собраны в отдельных файлах простой структуры. Эти файлы делятся на три группы: (1) системные, не зависящие от конкретного репозитория фразы; (2) части интерфейса, которые могут отличаться в разных репозиториях, они представляют собой исходную конфигурацию для создания репозитория; (3) конфигурация каждого репозитория, вначале просто копия конфигурации по умолчанию. Файлы каждой группы хранятся в разных папках, которые содержат подпапки для каждого языка (рис. 5).

К сожалению, cgi-скрипт *set_lang* не включен в последнюю версию EPrints, и поэтому функция переключения языков доступна только при помощи скрипта из предыдущей версии. Для того чтобы пользователи могли легко переключать язык, необходимо в шаблоны веб-интерфейса вставить соответствующие ссылки на скрипт.

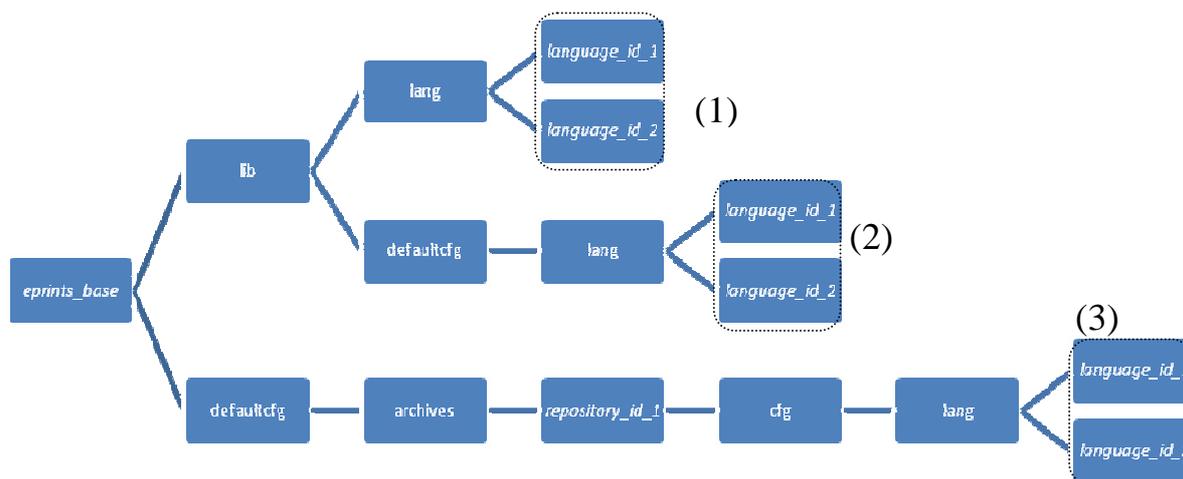


Рис. 5. Языко-зависимые файлы в EPrints

В целом, локализация EPrints проводится легко. И она достаточно полна в том смысле, что обладая полным переводом языко-зависимых файлов, только несколько фраз в интерфейсном тексте остаются непереуведенными (скорее всего пропущенные разработчиками по недосмотру).

Однако, в EPrints нет возможности отображать метаданные отдельного элемента репозитория на разных языках. Это было бы полезно не только для просмотров и поисков, но даже в большей степени для сбора метаданных. В системе EPrints только поле Тема может быть представлено на нескольких языках, либо с помощью редактора тем, либо посредством соответствующего XML файла. Также нет возможности менять структуру библиографических ссылок в системе в зависимости от выбранного языка, хотя соответствующие стандарты отличаются в разных странах.

Внутренне EPrints использует кодировку UTF-8 для всех полей метаданных и их значения могут быть введены на любом языке, но пока еще не все функции системы обрабатывают эту кодировку правильно. Например, ошибки такого типа были обнаружены в формате почтовых сообщений, рассылаемых системой, а также в функционировании расширенного поиска при использовании нелатинских символов в поле Автор.

DSpace. Прежде всего опишем основную версию системы, а затем дополнения и исправления от сторонних разработчиков. В DSpace есть только один ресурсный файл, содержащий большинство фраз интерфейса. Чтобы локализовать интерфейс, необходимо иметь перевод этого файла с видоизмененным именем. Когда пользователь входит в систему через веб-интерфейс, язык определяется автоматически по предпочтениям, указанным в веб-браузере. Переключение языков недоступно. Шаблоны писем электронной почты и файлы помощи не включены в локализацию. Нет также возможности ввести новости и описания коллекций на разных языках.

Однако существует несколько сторонних дополнений к системе, добавляющих переключение языков, возможность локализовать сообщения электронной почты и тексты справки. Имена коллекций и разделов, их описания, новости сайта остаются непереуводимыми.

В отличие от EPrints, значения метаданных могут быть добавлены на нескольких языках, хотя и не в обычном процессе депонирования, а на дополнительном шаге редактирования метаданных (это означает, что нужно провести дополнительные необычные манипуляции). Тем не менее, только первое из введенных значений используется для отображения на экране, вне зависимости от языка. Единственным исключением является возможность вводить альтернативные заголовки документа.

Общий вывод таков, что локализованный интерфейс и многоязычный сайт можно создать для обеих систем. В EPrints это сделать проще благодаря удачному архитектурному решению. В DSpace такой уровень локализации достигается с помощью сторонних дополнений. Обе системы не поддерживают многоязычные метаданные в интерфейсе.

7 Заключение

EPrints и DSpace являются системами одного и того же класса, предоставляющими полный спектр функциональности для создания электронных репозиторияев. Обе системы поддерживают OAI-PMH, но отличаются по структуре модели данных. Идея разделов в DSpace хороша, но и поддержка в EPrints разнообразных классификаций имеет свои преимущества. EPrints поддерживает больше форматов метаданных, но тут недостает поддержки расширенного Дублинского Ядра. EPrints более удобен для локализации, однако обе системы не поддерживают многоязычного представления метаданных.

8 О проекте

Анализ ряда программных продуктов, реализующих системы электронных библиотек был проведен в рамках проекта по информатизации НАН Украины «Разработка проектных решений автоматизированного библиотечного сервиса в интересах организаций НАН Украины». Было выбрано несколько систем, среди которых рассматривались DSpace и EPrints. Обе системы были успешно установлены и запущены как действующие службы DSpace в локальной сети, а EPrints доступен по адресу <http://eprints.isoftware.kiev.ua>.

Литература

- [1] Резниченко В.А., Проскудина Г.Ю., Овидий О.М. Создание цифровой библиотеки коллекций периодических изданий на основе Greenstone. Электронные библиотеки. 2005. — 8. Вып. 6. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2005/part6>.
- [2] Новицкий А.В., Резниченко В.А., Проскурина Г.Ю. Создание научных архивов с помощью системы EPrints. Электронные библиотеки. 2006. — Том 9. Вып. 4. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2006/part4/Novitski>
- [3] DSpace Federation Web site <http://dspace.org/>
- [4] EPrints Software <http://software.eprints.org/>
- [5] Registry of Open Access Repositories <http://roar.eprints.org/>
- [6] Open Archives Initiative. <http://www.openarchives.org/>
- [7] The Open Archives Initiative Protocol for Metadata Harvesting Version 2.0 of 2002-06-14. <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
- [8] Open Source BSD License. <http://www.opensource.org/licenses/bsd-license.php>
- [9] Kahn R., Wilensky R. A Framework for Distributed Digital Object Services, May 1995. <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>
- [10] Arms W.Y. Key Concepts in the Architecture of the Digital Library, D-Lib Magazine, July 1995. <http://www.dlib.org/dlib/July95/07arms.html>
- [11] Arms W.Y., Blanchi C., Overly E.A. An Architecture for Information in Digital Libraries, D-Lib Magazine, February 1997. <http://www.dlib.org/dlib/february97/cnri/02arms1.html>
- [12] Payette S., Lagoze C. Flexible and Extensible Digital Object and Repository Architecture, in Christos N. and Stephanidis C. eds., Research and Advanced Technologies for Digital Libraries: Proc. of the Second European Conference, ECDL '98, Heraklion, Crete, Greece, 1998, <http://www.cs.cornell.edu/payette/papers/ECDL98/FEDORA.html>
- [13] Thornton S., Ross W.: Virginia Dons FEDORA: A Prototype for a Digital Object Repository. DLib Magazine 6, 7/8 (July/August 2000). <http://www.dlib.org/dlib/july00/staples/07staples.html>
- [14] Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1, Blue Book, Issue 1, January 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- [15] EPrints GNU License <http://software.eprints.org/gnu.php>
- [16] Nixon W.J. DAEDALUS: Initial experiences with EPrints and DSpace at the University of Glasgow. Ariadne, Vol. 37 October 2003. <http://www.ariadne.ac.uk/issue37/nixon/intro.html>
- [17] The Greenstone Digital Library Software <http://www.greenstone.org/>
- [18] CERN Document Server Software (CDSware). <http://cdsware.cern.ch/>
- [19] Tansley R., Bass M., Stuve D., Branchofsky M., Chudnov D. The DSpace Institutional Digital Repository System: Current Functionality. In Proc. of JCDL 2003.
- [20] 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
- [21] ANSI/NISO Z39.85-2001. The Dublin Core Metadata Element Set. — National Information Standards Organization. (2001). <http://www.techstreet.com/cgi-bin/pdf/free/335284/z39.85-001.pdf>
- [22] Dublin Core Library Application Profile. <http://dublincore.org/documents/2002/09/24/libraryapplication-profile/>
- [23] Understanding Metadata. National Information Standards Organization. (2004). <http://www.niso.org>
- [24] ANSI/NISO Z39.88-2004. The OpenURL Framework for Context-Sensitive Services. National Information Standards Organization. (2005). http://www.niso.org/standards/resources/Z39_88_2004.pdf
- [25] EndNote. Bibliographies Made Easy. Getting Started Guide. Thomson. (2006), 86 p. <http://scientific.thomson.com/media/pdfs/ENXGettingStartedGuide.pdf>
- [26] Metadata Encoding and Transmission Standard (METS). <http://www.loc.gov/standards/mets/>
- [27] Ossanna J. F., Kernighan B. W., Heirloom Documentation Tools. Nroff/Troff User's Manual. 2007. <http://heirloom.sourceforge.net/doctools/troff.pdf>
- [28] Simple Metadata Annotation Specification. Version 6.2. Linguistic Data Consortium. (2004). http://projects.ldc.upenn.edu/MDE/Guidelines/SimpleMDE_V6.2.pdf

Comparison of repository systems EPrints 3.0 and DSpace 1.4.1

Kudim K.A., Proskudina G.Yu, Reznichenko V.A.

The basic facilities and features of most popular DSpace and EPrints open source systems for construction of scientific digital libraries are considered in the work. And also experience in creation of multilingual digital libraries on these basis is described. Comparative analysis of DSpace 1.4.1 and EPrints 3.1 is presented. Special attention is given to problems of localization, external formats compatibility and usability.