

ОТКРЫТЫЕ АРХИВЫ НАУЧНОЙ ИНФОРМАЦИИ

Филозова Ирина Анатольевна

*Старший преподаватель, научный сотрудник;
ГОУ ВПО Международный Университет природы, общества и человека «Дубна»,
Институт системного анализа и управления;
141980, Московская обл., г. Дубна, ул. Университетская, 19;
Объединенный институт ядерных исследований,
Лаборатория информационных технологий;
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
e-mail: fia@jinr.ru.*

Статья посвящена концепции Открытого Доступа и Открытых Архивов информации. Обсуждается развитие онлайн-инфраструктуры научно-образовательного сообщества в рамках идеологии и средств Открытых Архивов.

Ключевые слова: научная публикация, документ, метаданные, институциональные репозитории, открытый доступ, открытые архивы, Open Access Initiative (OAI), протокол OAI-PMH, депонирование, самоархивирование, электронные коллекции, электронные библиотеки.

OPEN ARCHIVES OF THE SCIENTIFIC INFORMATION

Filozova Irina

*Senior teacher, scientist;
Dubna International University of Nature, Society, and Man,
Institute of system analysis and management;
141980, Dubna, Moscow reg., Universitetskaya str., 19;
Join institute for nuclear research,
Laboratory of information technologies;
141980, Dubna, Moscow reg., Joliot-Curie str., 6;
e-mail: fia@jinr.ru.*

This article is on the concept of Open Access and Open Archives of the Information. The development of the online infrastructure for scientific and educational community within the frameworks of the ideology and tools of the open archives are discussed.

Keywords: scientific publishing, document, metadata, institutional repositories, open access, open archives, Open Access Initiative (OAI), the protocol OAI-PMH, deposit, self-archiving, digital collections, digital libraries.

В научном сообществе исторически существует два основных способа научной коммуникации: конференция и публикация. Результаты проведенных научных исследований, выраженные в форме публикации, становятся доступны для специалистов и общественности. «Публикация выступает как первичный источник сведений о научном знании, отношениях между учеными, ... той конечной реальностью, из которой исследователь черпает свои представления о науке ... изображения науки в различных исследовательских традициях ... становятся объектами изучения лишь постольку — поскольку сведения о них имеются в научной публикации.» [4].

В последние годы традиционные каналы распространения научных результатов посредством публикации в научных журналах претерпевают глубокие изменения, связанные с переходом от парадигмы традиционной публикации к созданию открытых архивов научной продукции. Это обусловлено несколькими причинами: экономическими, технологическими, организационными, социальными [3].

Изменения в обеспечении научных библиотек, происходящие в последнее время, явились следствием так называемого «журнального кризиса», суть которого в том, что стоимость подписки на научные журналы в мире росла в 2 – 3 раза быстрее, чем темпы роста бюджетов научных библиотек и темпы инфляции [7]. Таким образом, многие даже крупные библиотеки вынуждены отказываться от подписки на поддерживаемое ранее количество наименований научных журналов. А издательства повышают стоимость подписки при уменьшении числа подписчиков. Например, стоимость подписки на 2009 г. журнала «Annals of Physics» издательства Elsevier составила 11 000 \$ США.

В то же время бурное развитие электронных публикаций позволяет научным, исследовательским и образовательным организациям создавать и поддерживать архивы своей научной продукции. Возможность хранения гетерогенных материалов в электронном формате повышает интерес к их широкому применению. Развитие сетевых и программных технологий расширяет список возможностей: индексирование новых объектов – таблиц, изображений, диаграмм и пр., обновление отдельной части веб-страницы (AJAX), ведение личного сетевого дневника, обмен информацией по технологии peer-2-peer (P2P), пополнение новостей с помощью системы RSS. Электронные архивы более удобны в управлении, их легче поддерживать и сопровождать, чем традиционные. Стоимость содержания электронной версии журнала ниже, чем его печатного аналога.

Среди социальных причин специалисты называют понимание современным обществом факта «двойной платы за исследования» и удорожания журнальной продукции без изменения качества. Сначала общество платит исследователю за работу, затем оно платит научным журналам за доступ к результатам этой работы.

Открытый доступ

По этим причинам на западе в 90-е гг. возникло и бурно развивается мощное общественное движение за открытый доступ к результатам исследований (Open Access to Research, <http://www.eprints.org/openaccess/>). Открытый доступ (ОД) – это способ научного общения путем реализации права автора произведения на доведение до всеобщего сведения таким образом, что любое лицо может получить доступ к произведению из любого места и в любое время по собственному выбору [6]. Решение о представлении автором своей публикации в Интернет является добровольным. ОД не противоречит авторскому праву, способствует известности автора, закреплению научного приоритета, усилению научного влияния.

Основные принципы открытого доступа сформулированы в следующих документах:

- **Будапештская инициатива «Открытый доступ»** (Budapest Open Access Initiative) была провозглашена на собрании, проведенном Институтом «Открытое Общество» 14 февраля 2002 г. в Будапеште [2]. Определение открытого доступа, данное в ходе Будапештской инициативы (BOAI): «Это право пользователя читать, выгружать, копировать, распространять, печатать, осуществлять поиск или проставлять гиперсвязи к полному тексту статей».
- **Берлинская декларация об открытом доступе к научным и гуманитарным знаниям** (Open Access to Knowledge in the Sciences and Humanities) была принята в октябре 2003 г. [1]. Модель Открытого доступа, разработанная в рамках Берлинской декларации, развивает идеи самоархивирования и формирования институциональных репозиториев (ИР).

Согласно Берлинской декларации авторы и правообладатели даруют всем пользователям свободное, неограниченное право доступа и разрешение на копирование, использование, распространение, передачу и публичную демонстрацию, а также на распространение работы в электронной форме для любых законных целей с соответствующим указанием авторства, а также право на изготовление небольшого количества печатанных копий для своего личного использования. Полная версия работ и всех сопутствующих материалов депонируется (публикуется), по крайней мере, в одном сетевом репозитории, использующем принятые технические стандарты, который поддерживается научным сообществом, государственным агентством или другим авторитетным институтом, уполномоченным обеспечивать открытый доступ, беспрепятственное распространение, сетевое взаимодействие и долгосрочное хранение [1].

Модели открытого доступа

Осуществление открытого доступа реализуется по двум взаимодополняющим направлениям: *журналы открытого доступа* и *архивы (репозитории) открытого доступа*. Журналы открытого доступа публикуют статьи, получившие экспертную оценку. Репозитории осуществляют сбор документов – публикаций в состоянии неотрецензированного и отрецензированного препринта, опубликованного постпринта и его доработанных после опубликования версий.

В отношении к реферируемым научным документам существует два основных направления открытого доступа, получившие название «*золотой путь*» (*gold road*) и «*зеленый путь*» (*green road*). «*Золотой путь*» означает, что журнал не взимает плату за доступ конечного пользователя (читателя) к опубликованным материалам. «*Зеленый путь*» разрешает размещать прорецензированные материалы в общедоступные институциональные и тематические репозитории.

Существуют и другие финансовые модели – финансирование публикаций благотворительными фондами, оплата автором или организацией, которую он представляет, корпоративная модель (например, Sponsoring Consortium for Open Access Publishing in Particle Physics – SCOAP).

В ходе последующих Берлинских конференций были разработаны рекомендации для научных и образовательных учреждений по реализации принципов Берлинской декларации [7]:

1. Рекомендовать ученым, научным сотрудникам и исследователям депонировать свои опубликованные работы в репозитории открытого доступа.
2. Поддерживать ученых, научных сотрудников и исследователей в их стремлении опубликовать результаты их научных работ в существующих по данной тематике журналах открытого доступа.
3. Обеспечивать всемерную поддержку создания и развития таких журналов.

Архивы открытого доступа

Архивы открытого доступа становятся востребованными для научных институтов и вузов. Формирование таких электронных архивов рассматривается организациями как платформа для демонстрации своей научной продукции. Таким образом, научным институтам и образовательным учреждениям целесообразно создавать и развивать архивы своей научной продукции помимо традиционных Web-сайтов.

Открытые архивы строятся по принципу распределенной системы с общим сетевым протоколом и программным обеспечением, что позволяет говорить о единой информационной системе.

В ходе 69-й Генеральной конференции ИФЛА на семинаре «Информационные технологии и работа группы метаданных Dublin Core» были сформулированы принципы, на которых базируется идеология «Открытого архива» [7]: консолидация в мировом масштабе архивов научных материалов; свободный доступ к архивам (к метаданным); согласованные интерфейсы архивов и поставщиков информации; простота использования; применение существующих стандартов – HTTP, XML, Dublin Core, MARC, MARCXML.

В общем случае под термином *репозиторий* понимают хранилище данных, организованных в файлы, доступные для дальнейшего распространения по сети. В них хранят дистрибутивы программного обеспечения с учетом версионности, библиотеки программ, документацию с историей изменения и другой служебной информацией. Условно различают тематические (предметные) и институциональные (учрежденческие) репозитории. Они могут существовать и вне системы Открытого доступа. Такие репозитории реализуются как набор сервисов для членов сообщества, по управлению и распространению электронных материалов, созданных в данной организации. По определению сетевой энциклопедии Википедии (*Wikipedia*) «Институциональный репозиторий – это место в сети для сбора, обеспечения сохранности и распространения в электронном формате интеллектуальной продукции учреждения, в частности, научного учреждения». Институциональный репозиторий предполагает сбор «своих» публикаций, т.е. публикаций своих сотрудников, чем и отличается от тематической электронной библиотеки (ЭБ). Электронные библиотеки – это распределенные информационные системы, позволяющие обрабатывать, хранить, распространять, анализировать и организовывать поиск в разнообразных коллекциях электронных документов через глобальные сети передачи дан-

ных. Основным объектом хранения данных в ЭБ является документ, снабженный метаданными. Под электронным документом в таких системах понимается элемент данных или ресурс (информационный объект), наиболее полно отражающий потребности пользователя данной ЭБ. Но эти отличия весьма размыты. Институциональные репозитории могут создаваться и как часть ЭБ, поддерживаемой организацией, или быть реализованы в форме ЭБ. Пример такой электронной библиотеки в области физики высоких энергий – архив публикаций на сервере CERN Document Server (CDS) Европейского центра ядерных исследований (ЦЕРН).

Таким образом, в материалах по открытым архивам употребляется следующая терминология:

Открытые Архивы (ОА) – это распределенная совокупность информационных объектов, доступных в среде Интернет, для которой поддерживается репозиторий описывающих их стандартизованных метаданных. ОА представляют собой информационные системы, построенные на технологии OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).

Репозиторий — хранилище метаданных в открытых архивах.

Институциональные репозитории открытого доступа – публично доступные архивы научных, исследовательских и образовательных организаций, в которых члены сообщества размещают свои опубликованные и подготовленные к печати статьи и другие материалы научно-исследовательской и научно-организационной деятельности.

Репозитории метаданных доступны для авторизованных приложений-сборщиков метаданных, функционирующих в соответствии с протоколом OAI-PMH. В репозиториях по запросам сборщиков метаданных выполняется набор операций, определенных протоколом OAI-PMH. В качестве примера глобальных сборщиков метаданных можно назвать Scirus (<http://www.scirus.com>), myOAI (<http://www.myoai.com>), OAIster (<http://oaister.umdl.umich.edu>). Регистрация в OAI-реестрах, которые собирают метаданные со всех зарегистрированных архивов и осуществляют поиск по ним, не обязательна, но желательна.

Некоторые крупные реестры OAI-совместимых репозиториев:

1. Реестр открытых архивов: <http://www.openarchives.org/Register/BrowseSites>.
2. Реестр репозиториев открытого доступа: <http://roar.eprints.org/>.
3. Каталог открытых репозиториев *OpenDOAR / Directory of Open Access Repositories* – <http://www.opendoar.org/>.
4. Европейский реестр OAI-PMH совместимых репозиториев: <http://www.openarchives.eu/home/home.aspx>.

Протокол сбора метаданных / Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH)

Технологически модель Открытых архивов опирается на протокол сбора метаданных, базирующийся на протоколе HTTP. Протокол использует шесть типов запросов и HTTP-методы POST и GET (Таблица 1).

Различают два типа запросов по протоколу OAI-PMH:

1. Выдать все записи.
2. Выдать записи, которые *<условие>*,

где *<условие>* уточняет принадлежность к набору, временные параметры и др. характеристики.

Таблица 1. Запросы Protocol for Metadata Harvesting

Команда	Описание
GetRecord	Поиск записи в объекте репозитория. Аргументы уточняют идентификатор требуемой записи и формат метаданных.
Identify	Поиск информации о репозитории: человекочитаемое имя репозитория; базовый URL репозитория; поддерживаемая версия протокола OAI; адрес электронной почты администратора.
ListIdentifier	Поиск идентификаторов записей в репозитории. Дополнительные аргументы задают принадлежность к определенному набору, временные параметры – модификация, создание или удаление в определенный период.
ListMetadataFormats	Получение списка форматов метаданных, доступных в репозитории.
ListRecords	Сбор записей в репозитории. Дополнительные аргументы для селективного сбора – принадлежность записи к конкретному набору, временные параметры.
ListSets	Получение структуры наборов коллекций в репозитории.

Ответы кодируются в синтаксисе языка XML. Протокол предусматривает три типа классификации: по предметным рубрикам Немецкой библиографии, по формальному типу публикации (статья, книга, диссертация и т.п.), по формальному типу документа (изображение, текст, звук и т.п.).

Информационная модель OAI-PMH

Информационная модель OAI-PMH описывает ресурс в терминах библиографических данных [5]. Информационный ресурс представлен в репозитории ЭЛЕМЕНТОМ, имеющим уникальный ИДЕНТИФИКАТОР_ЗАПИСИ и содержащий метаданные в различных форматах, на основе которых формируются ЗАПИСИ: РЕСУРС ↔ ЭЛЕМЕНТ {ИДЕНТИФИКАТОР_ЗАПИСИ; ЗАПИСИ}



Ресурс – традиционный библиотечный объект (книга, статья) или другие сущности (изображения, звук, музейный экспонат и пр.).

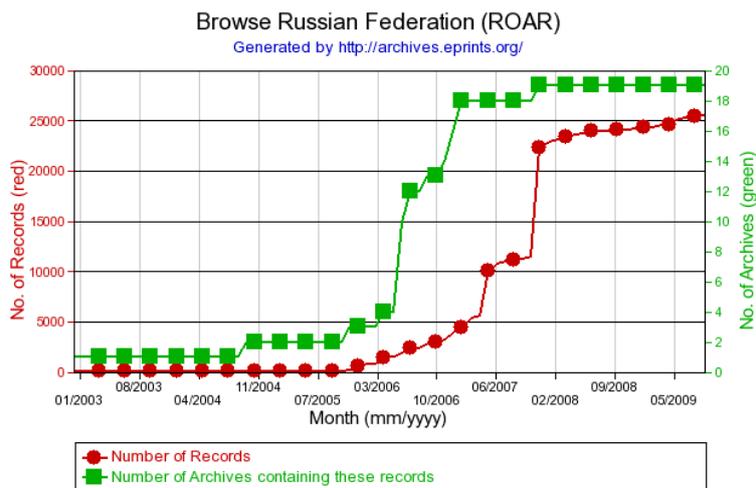
Элемент – контейнер, хранящий или генерирующий метаданные о ресурсе в различных форматах. На основе каждого формата ресурс может быть представлен в виде записи средствами OAI-PMH.

Идентификатор записи – уникальный идентификатор, однозначно определяющий элемент в репозитории, используется в запросах OAI-PMH.

Протокол OAI-PMH требует обязательного включения в описание ресурса набора метаданных Дублинское ядро / Dublin Core. DublinCore (DC) – формат метаданных, рекомендуемый 15 элементов для простого описания и поиска ресурсов. Произвольный формат метаданных описывается XML-Schema. Протокол рекомендует использование элемента записи метаданных для связывания записи с

идентификатором (URL, URN, DOI и пр.) документа. Это необходимо для доступа к полному тексту, т.к. идентификатор записи не является идентификатором документа.

По данным реестра репозитория открытого доступа ROAR (Registry of Open Access Repositories, <http://roar.eprints.org/>) на декабрь 2009 г. лидерами по количеству архивов открытого доступа являются США (301), Великобритания (158), Германия (112). Далее следуют Япония (71), Бразилия (70), Канада (52), Франция и Италия (50), Австралия (48). Активность России в этой сфере пока выглядит скромно – 28 репозитория. Отставание в данной области может повлиять на качество научной коммуникации. Поэтому представляется, что исследование данной проблематики, изучение зарубежного и отечественного опыта построения открытых архивов позволит преодолеть негативные тенденции. На графике, приведенном ниже, видно, что показатель роста числа зарегистрированных в реестре архивов практически не изменился с февраля 2008 г. по май 2009 г. Рост количества записей в них также несущественен.



Вероятно, это связано с тем, что в организациях, создавших свои архивы на некотором базовом массиве данных, не была разработана четкая политика их наполнения. Это один из самых острых и важных вопросов, не имеющих универсального решения. Ведь своевременное и качественное наполнение информационной системы – основа ее жизнеспособности.

Литература

1. Берлинская Декларация об открытом доступе к научному и гуманитарному знанию / Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. [Электронный ресурс]. URL: http://www.mpg.de/pdf/openaccess/BerlinDeclaration_en.pdf (дата обращения: 16.12.2009).
2. Будапештская инициатива «Открытый доступ» / Budapest Open Access Initiative. [Электронный ресурс]. URL: <http://www.soros.org/openaccess/ru/index.shtml> (дата обращения: 22.01.2010).
3. Земсков А. И, Шрайберг Я. Л. Системы открытого доступа к информации: причины и история возникновения // Научные и технические библиотеки. – 2008. – №8.
4. Мирский Э. М. Массив публикаций и система научной дисциплины // Системные исследования. – М.: Наука, 1977.
5. Новицкий А. В., Резниченко В. А., Проскудина Г. Ю. Создание научных архивов с помощью системы EPrints // Российский научный электронный журнал «Электронные библиотеки». [Электронный ресурс]. URL: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal> (дата обращения: 22.01.2010).
6. Федеральный закон от 20 июля 2004 г. N 72-ФЗ «О внесении изменений в Закон Российской Федерации «Об авторском праве и смежных правах».
7. Шрайберг Я. Л, Земсков А. И.. Модели открытого доступа: история, виды, особенности, терминология // Научные и технические библиотеки. – 2008. – №5.