

DSpace как платформа для институционального репозитория

Во второй половине двадцатого века информационное взаимодействие в академическом сообществе проходило по традиционной для данной сферы схеме. Ход, а затем и результаты проведенных исследований публиковались в монографиях или в научных рецензируемых журналах по теме исследования. Публикации являлись материальным свидетельством научной значимости и успеха исследования и исследователей. Редакторы и рецензенты журналов служили хранителями научных знаний и, следовательно, хранителями и даже гарантами статуса исследователей.

С появлением новых информационных технологий, сетевых технологий и, как следствие, доступности механизмов быстрого обмена информацией и публикаций в электронных изданиях, привычная схема информационного взаимодействия в академическом сообществе начала изменяться.

Во-первых, новые технологии добавили скорости традиционным методам научной коммуникации. Обмен данными, как вербальными, так и большими массивами экспериментальных данных, значительно ускорился.

Архивы препринтов в некоторых областях физики и математики стали обычным способом общения ученых - однако большинство авторов, практикующих препринт, всё же ожидают, что статья появится в печатной версии журнала.

Во-вторых, растет количество важных исследований, ход и результаты которых, не будут формально опубликованы. Например, официальные документы и технические отчеты финансируемых различными фондами и организациями исследований в области компьютерных наук, аналитики, квантовых или биотехнологий.

В-третьих, повышение доступности первичного исследовательского материала в сети уменьшает зависимость от наличия традиционно опубликованных материалов. Широкому кругу ученых предоставляется возможность проверить и воспроизвести результаты, проанализировать данные, лежащие в основе исследования, получить опыт моделирования и визуализации.

Печатные статьи постепенно становятся своего рода окончательным результатом исследования: выполняют функцию гаранта научного признания, но не научной коммуникации. Роль печатного издания меняется [1].

Сотрудники практически каждого высшего учебного заведения занимаются научной работой. ВУЗы производят сотни тысяч неопубликованных документов, множество технических отчетов, официальных и прочих материалов во всех без исключения областях науки, техники, бизнеса, экономики и пр.

Учебные заведения дополнительно производят большое количество первичных цифровых материалов исследования в виде наборов данных (статистических, картографических и т.д.), мультимедийных данных, источниками которых могут быть радары, телескопы, спутники или даже фото-видео хроника. Так же в учебных заведениях производится и распространяется программное обеспечение для моделирования, визуализации и другие механизмы выполнения или распространения научных исследований и их результатов.

Многие сотрудники публикуют подобные материалы на сайтах подразделений ВУЗа, и это, безусловно, преумножает интеллектуальный актив учреждения.

В знак признания новых тенденций в сфере научных коммуникаций и роли библиотеки в сборе и сохранении и распространении информации для нынешних и будущих исследователей был создан проект DSpace.

Система DSpace является результатом совместной разработки библиотеки Массачусетского технологического института и Hewlett Packard Laboratories. Первоначальной целью разработчиков было построение открытого институционального архива для поддержки исследований Массачусетского технологического института.

DSpace располагает инструментами для загрузки, управления, распространения и описания цифровых материалов.

Эти инструменты включают в себя интегрированные подсистемы для различных представлений цифрового материала и связанных с ними метаданных, гибкие схемы метаданных, систему индексирования и поиска, систему архивного управления пакетами, инструмент для создания политик доступа и управления, систему генерации постоянных идентификаторов, и пр.

В наши дни развитием и поддержкой системы DSpace занимается организация DuraSpace.

Некоммерческая организация DuraSpace была основана в 2009 году [2], в результате объединения усилий организаций Fedora Commons [3] и DSpace Foundation [4] - крупнейших поставщиков программного обеспечения с открытым исходным кодом для управления и обеспечения доступа к цифровому содержанию¹.

Организация получила багаж технологий с открытым кодом, разработанных при участии библиотекарей, архивариусов, исследователей и технологов, которые разделяют идеи создания и сохранения долгосрочного доступа к цифровому наследию мира. Для хранителей знаний, программное обеспечение с открытым исходным кодом [5] имеет ряд важных преимуществ по сравнению с проприетарным программным обеспечением. Продукты с открытым исходным кодом разрабатываются на основе принципов свободного обмена идеями и ресурсами между участниками процесса разработки. Сообщества DSpace и Fedora использовали этот процесс для создания программной платформы, которая используется в более чем тысяче учреждений в ста странах мира.[6]

В число учреждений, использующих систему DSpace входят как зарубежные (Мичиганский университет, Колумбийский университет, Массачусетский технологический институт, Северо-Западный университет Чикаго и пр.), так и отечественные высшие учебные заведения (Уральский федеральный университет, Уральский государственный лесотехнический университет, Белгородский государственный университет, Сибирский федеральный университет, Тверской государственный университет, Удмуртский государственный университет, Южно-Уральский государственный университет, Ярославский государственный университет и пр.).

С технической стороны система DSpace сложна, но в то же время прозрачна и понятна. Для хранения данных используется свободная СУБД PostgreSQL, ядро системы написано на языке Java и для его сборки и запуска используются свободные инструменты Java Development Kit, Apache ANT, Apache Maven и Apache Tomcat. DSpace работает со всеми стандартными для библиотечной сферы протоколами, такими, как стек протоколов Z39.50. Система может работать практически под любой операционной системой, в том числе и под открытыми системами UNIX и GNU/Linux. Не смотря на

¹ <http://www.fedora-commons.org/>, <http://www.dspace.org/>

кажущуюся сложность системы по причине использования серьёзных программных инструментов, её установка и настройка занимает совсем немного времени и не требует особой квалификации. Студенты пятого курса Института информатики и информационных технологий Уральского государственного педагогического университета справляются с установкой в среднем за сорок минут при наличии документа, описывающего процесс [7].

Тонкая настройка специальных функций требует большего времени и специальной подготовки, но и проводится она специалистами – библиотекарями-автоматизаторами.

Опыт внедрения системы DSpace в Уральском федеральном университете положительный. Итогом работы электронного архива Уральского государственного университета и его преемника, Уральского федерального университета в течение десяти лет, стало почетное 358 место в рейтинге Web of World Repositories (TOP Institutional Repositories) по итогам 2011 года. В связи с изменением доменного имени и связанными с этим техническими особенностями индексирования, на сентябрь 2012 года архив занимает лишь 1112-ое место в мире и 6-ое в России [8]. Самостоятельные подсчеты рейтинга по системе webometrics позволяют говорить об уверенном третьем месте в России и вхождении в мировой TOP400. Цитируемость архива, частота и объёмы его актуализации, процент документов в открытом доступе – критерии, влияющие на рейтинг. Если эти критерии количественно высоки, значит материалы архива актуальны и востребованы. Значит, коммуникации научного сообщества посредством архива ведутся и материалы архива цитируются. Следовательно, инструмент выполняет свои задачи.

Система DSpace обладает рядом уникальных, отсутствующих в альтернативных системах функций.

Первая особенность – полнотекстовый поиск. Система поддерживает извлечение данных из большого количества форматов файлов. Как правило, система работает с файлами PDF и после проведения процедуры индексации, позволяет производить поиск строки не только по метаданным, но и по данным – по содержимому файлов, прикрепленных к записям в электронном архиве.

Вторая особенность – открытость системы, и как следствие – расширяемость. Допустим, данные, загружаемые в систему имеют текстовое наполнение, но не имеют распознанного слоя (фотографии архива газет, карты, PDF файлы с изображениями). Не составляет большого труда модифицировать систему индексации текстового содержания, описанную выше, так, чтобы она

использовала систему оптического распознавания символов (OCR), функционирующую на кластере [9]. Таким образом, можно автоматически распознать надписи на картах, в нотных тетрадах, газетный текст и пр. информацию, по которой впоследствии можно будет производить поиск в электронном архиве.

Третья особенность – возможность обмена коллекциями между разными электронными архивами. Подобный функционал делает коммуникации в академической среде еще более быстрыми и удобными, стирая границы, а использование стандартных протоколов, позволяет строить большие гетерогенные архивы из систем DSpace и альтернативных решений.

Четвертая особенность – возможность регистрации пользователей с разными правами для доступа к разным типам данных. Подобный функционал очень полезен для обеспечения доступа определенного круга лиц к закрытой в силу специфики исследования информации, информации, защищенной авторским правом и пр.

DSpace – актуальная, развивающаяся и удобная система, имеющая мощное сообщество разработчиков и пользователей. По мнению многих экспертов, DSpace обеспечивает лучшую платформу, для долгосрочного хранения цифровых материалов, используемых в академических исследованиях [10].

Литература

1. Smith M. DSpace: An Institutional Repository from the MIT Libraries and Hewlett Packard Laboratories : [электронный ресурс] / М. Smith // Lecture Notes in Computer Science. - 2002. - N 2458. - P. 543-549. - Режим доступа: http://dspace.mit.edu/bitstream/handle/1721.1/26706/Smith_2002_DSpace.pdf?sequence=1
2. Fedora Commons and DSpace Foundation Join Together to Create DuraSpace™ Organization : [электронный ресурс] - Режим доступа: <http://duraspace.org/node/710>
3. Fedora Commons Awarded \$4.9M Grant to Develop Open-Source Software for Building Collaborative Information Communities : [электронный ресурс] // Science. - 2007. - Aug. 7. - Режим доступа: <http://www.moore.org/pa-newsitem.aspx?id=2172>
4. HP and MIT Create Non-profit Organization to Support Growing Community of DSpace Users : [электронный ресурс]. - Режим доступа: <http://www.hp.com/hpinfo/newsroom/press/2007/070717a.html>
5. Why Open Source? : [электронный ресурс]. - Режим доступа: http://duraspace.org/why_open_source
6. DuraSpace : [электронный ресурс]. - Режим доступа: <http://en.wikipedia.org/wiki/DuraSpace>
7. Ефимов А. А. Установка DSpace 1.8 в операционной системе Windows : [электронный ресурс]. - Режим доступа: <http://ideafix.name/?p=359>
8. Ranking web of repositories Central & Eastern Europe : [электронный ресурс]. - Режим доступа: http://repositories.webometrics.info/en/Ranking_Europe/Central_Europe
9. Ефимов А. А. Доступный суперкомпьютер в академической среде, или облачные вычисления в учебных аудиториях / А. А. Ефимов // Новые образовательные технологии в вузе : материалы VII междунар. науч.-метод. конф. - Екатеринбург, 2010. - Ч. 1. - С. 14-17.
10. Nixon W.J. DAEDALUS, Initial experiences with EPrints and DSpace at the University of Glasgow [электронный ресурс]. - Режим доступа: https://dspace.gla.ac.uk/bitstream/1905/197/1/initial_experiences_eprints_dspace.pdf