

## **Информационные ресурсы открытого доступа и возможности их интеграции в единое информационное пространство**

*Г.О.Еременко*

*Научная электронная библиотека eLIBRARY.RU*

В процессе работы с электронными ресурсами каждая научная, образовательная, медицинская организация сталкивается с целым рядом вопросов и проблем. Эти вопросы можно разделить на две группы: связанные с потреблением электронной информации, с получением доступа к информационным ресурсам, с одной стороны, и с производством электронной информации и ее распространением, с другой.

О способах решения первой проблемы уже много говорилось в рамках наших конференций. Здесь основные задачи заключаются в поиске финансирования для приобретения ресурсов, часто довольно дорогих, и в организации консорциумов с целью получения скидок и оптимизации расходов в расчете на одну организацию. Я не буду рассматривать более подробно способы решения этих задач.

А вот вопрос создания и распространения собственных электронных ресурсов давайте рассмотрим подробнее. Каждая организация рано или поздно приходит к мысли, что неплохо было бы собирать информационные ресурсы, рождающиеся в организации (научные статьи, отчеты, диссертации, учебно-методические пособия, материалы конференций и т.д.), в электронном виде и как-то организовать к ним доступ, поиск, распространение и т.д. Некоторые организации уже неплохо продвинулись в области оцифровки своих внутренних информационных ресурсов, некоторые только начинают эту работу. Часто возникает вопрос: вот мы собрали у себя в организации коллекцию научных статей наших сотрудников или коллекцию диссертаций. Что с ними делать? Как лучше организовать их распространение?

Начнем с начала. Предположим, организация решила начать создавать такой архив внутренней информации. С чего начинать, как организовать доступ и распространение. Ну понятно, что вначале нужно ресурсы оцифровать – отсканировать, если они имеются только в печатном виде, или преобразовать к нужному формату – если есть электронные версии. Конечно, хорошо было бы, по-возможности, после сканирования провести распознавание документа – это поможет потом при создании его библиографического описания, а также для организации полнотекстового поиска. Также не нужно забывать о своевременном решении проблемы с авторскими правами на данный документ.

Что делать дальше с созданными ресурсами? Можно их разместить на сервере организации и организовать доступ через локальную сеть. Ресурсы будут доступны пользователям организации, но это не решает проблемы их распространения. Здесь возникает закономерный вопрос – а зачем их вообще распространять?

Во-первых, для широкого информирования научно-образовательной общественности о работах, проведенных в организации. Чем более доступна эта информация, тем более известна данная организация, тем больше цитирование научных работ, и, соответственно, тем больше инвестиций, грантов и т.д. Я думаю, эта цель стоит того, чтобы уделять вопросам распространения такой информации достойное внимание.

С другой стороны, можно взглянуть на этот вопрос иначе. Если организация производит действительно ценную и качественную информацию

(например, научные журналы, книги, научные пособия и т.д.), то почему бы не превратить ее в дополнительный источник финансирования, т.е. почему бы не продавать доступ к этой информации? Это тоже вполне реально.

Что обычно организация пытается сделать для распространения своих ресурсов. Конечно, выставить их в Интернет на своем сайте. Казалось, бы задача решена? Но, к сожалению, не все так просто. То есть проблема в этом случае решается только частично. Ведь чтобы добраться до этой информации, пользователь Интернет должен ее найти. Конечно, можно и нужно ее зарегистрировать в различных поисковых машинах Интернет. Но, к сожалению, сейчас это уже мало что дает. При существующем объеме проиндексированной этими машинами информации чаще всего надеяться на то, что Ваш сайт попадет в первую сотню результатов поиска, выводимых пользователю, не приходится. А дальше пользователю просматривать уже не хватит терпения.

Каким же образом можно решить эту проблему? Нужно использовать специальные протоколы и поисковые системы, специально предназначенные для работы с такой структурированной информацией. Нужно сделать библиографические описания ресурсов, чтобы можно было искать по отдельным полям: по авторам, по названию документов и т.д.

Дальше нужно определиться с архитектурой системы поиска, которой мы собираемся воспользоваться. Здесь существуют три подхода, различающиеся по степени централизации ресурсов, метаинформации и поисковых сервисов. Вначале я их перечислю, а потом мы их рассмотрим подробнее и сравним.

1) В первом случае все ресурсы, библиографические описания, и поисковый сервис находятся в организации. Будем называть ее ресурсным центром. По такому принципу работают системы, построенные на протоколе Z39.50, довольно широко распространенном в крупных библиотеках, в том числе российских. При этом поиск осуществляется путем непосредственного обращения ко всем или к выбранным пользователем электронным библиотекам с последующим сведением полученных из разных библиотек результатов в единый список. Так работают, например, системы в проектах АРБИКОН и СИГЛА.

2) Ресурсы остаются в организации, а метаинформация копируется на центральный поисковый сервер. Поиск осуществляется на центральном сервере, за полным текстом документа пользователь обращается в ресурсный центр. Такой подход реализован, например, в системах, поддерживающих протокол OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting). В России такие работы по харвестингу проводятся, насколько мне известно в рамках проекта СИГЛА.

3) Все ресурсы и метаинформация копируются на центральный поисковый сервер. Поиск и схема авторизации пользователей к ресурсам реализуется на центральном сервере, но управляется и контролируется организацией. По такому принципу работает Научная электронная библиотека eLibrary.Ru.

Давайте теперь сравним эти подходы:

	Z39.50	OAI-PMH	eLibrary.Ru
Полные тексты	ресурсный центр	ресурсный центр	сервисный центр
Метаинформация	ресурсный центр	сервисный центр	сервисный центр
Поиск	распределенный	в центральной БД	в центральной БД
Обновление	немедленно	с задержкой	с задержкой

метаданных			
Единая рубрикация	отсутствует	возможна	возможна
Доступность ресурсных центров в момент запроса	потеря данных	зависит для полных текстов	не зависит
Нормализация метаданных	невозможна	возможна	возможна
Скорость поиска	низкая	высокая	высокая
Единая система авторизации	невозможна	невозможна	возможна
Единая система статистики	невозможна	возможна по метаданным	возможна по метаданным и полным текстам
Сложность установки и поддержания	высокая	средняя	низкая
Единая биллинговая система	невозможна	невозможна	возможна
Надежность хранения ресурсов	низкая	дублирование по метаинформации	дублирование по метаинформации и полным текстам
Использование пристатейной библиографии	невозможно	невозможно	возможно
Расходы	ПО + поддержка	ПО + поддержка	подписка

В целом, сравнивая эти подходы можно сделать вывод что общее качество сервиса растет по мере увеличения степени централизации системы, тогда как скорость обновления информации уменьшается. Впрочем, поскольку процесс сбора информации из ресурсных центров в сервисный центр автоматизирован, эти задержки могут быть сведены к минимуму. При этом ресурсы, подверженные частым обновлениям – опрашиваются чаще, более стабильные ресурсы – реже.

Какой же способ распространения выбрать?

Здесь все зависит от поставленных задач. Все три варианта возможны и будут продолжать использоваться.

Серьезные поставщики информации (например, крупные библиотеки) на данный момент времени должны уметь работать и по протоколу Z39.50, и по протоколу OAI-PMH, чтобы обеспечить максимально возможное распространение своей информации.

Если ресурсный центр только создается, то, наверное, лучше использовать OAI. Он проще, современнее, обеспечивает более высокое качество сервиса, имеет возможность для развития, позволяет интегрировать ресурс со многими другими открытыми ресурсами. Для обеспечения поддержки протокола OAI можно либо приобрести коммерческое программное обеспечение, поддерживающее этот протокол, либо разработать свое, либо использовать бесплатно распространяемое ПО для создания открытых архивов (например, DSpace, FEDORA и некоторые другие).

Если Вас интересует более централизованный вариант, который, как мы видели, имеет целый ряд дополнительных преимуществ, или, например, Вы собираетесь продавать свои информационные ресурсы, можно рассмотреть вариант сотрудничества с Научной электронной библиотекой. Сбор информации из ресурсных центров мы тоже осуществляем через протокол OAI-PMH, но

расширенный с целью решения дополнительных задач, а именно: возможности передачи индексов полных текстов, самих полных текстов, пристатейной библиографии, информации о правах и условиях распространения. При этом полные тексты могут либо передаваться в НЭБ, либо оставаться в ресурсном центре, либо дублироваться (что также очень полезно с точки зрения наличия резервных копий информации). Такая схема позволяет нам построить на базе передаваемой информации дополнительные сервисы, такие как централизованный поиск по полным текстам, навигацию по пристатейным ссылкам, настройку прав распространения ресурсов, организацию платного доступа к ним на базе подписки или доставки документов и другие.

Дело в том, что протокол OAI построен на базе XML и для обеспечения совместимости требует обязательной поддержки схемы описания метаданных Dublin Core. В то же время не запрещается и даже рекомендуется дополнительная поддержка других, более сложных форматов описания метаданных. Если какая-то группа организаций договорилась об использовании дополнительного формата, они могут легко расширить возможности своего взаимодействия для решения каких-то специфических задач, но оставаясь при этом в поле ресурсов, доступных через протокол OAI. Например, если Вы собираетесь создавать общую базу диссертаций, Вам имеет смысл добавить в схему XML поля, специфические для описания и поиска диссертаций. Известны случаи, например, когда с помощью OAI был организован сбор статистики использования ресурсов из разных организаций.

Наконец, хотелось бы сказать несколько слов о том, как можно стимулировать создание и наполнение информацией электронных репозитариев организаций. Для активного привлечения организаций к этому процессу красивых слов о необходимости распространения информации недостаточно. В некоторых странах, например, в Великобритании и некоторых других, для этого принимаются решения на национальном уровне, устанавливающие, что результаты всех работ, проведенных на государственные средства, в том числе на средства научных грантов, должны быть обязательно опубликованы в одном из открытых архивов. Это рассматривается и как показатель уровня научной деятельности организации, и влияет на ее дальнейшее государственное финансирование. Такие решения могут приниматься и на уровне отдельных регионов, научных центров или отдельных организаций.