

Вопросы развития электронного архива научных публикаций

Ковязина Елена Васильевна, научный сотрудник, Институт вычислительного моделирования Сибирского отделения Российской академии наук, elena@icm.krasn.ru

В докладе выделены основные этапы развития электронного архива научных публикаций, определяющие трансформацию базы данных трудов сотрудников организации в систему управления электронным архивом научных публикаций. Приведены особенности и характерные признаки каждого этапа. Перечислен ряд проблем, организационного и технологического характера, и предложены возможные пути решения некоторых из них.

Важной частью работы каждой библиотеки научной или образовательной организации является формирование базы данных трудов сотрудников этой организации. Конечной целью такого ресурса является сохранение научного наследия и отражение в публикациях результатов научных исследований. В свете последних государственных инициатив полнота и точность представленных в таком ресурсе данных приобретает жизненно важное для организации значение. Следуя за развитием технологий, база трудов сотрудников проходит от начала формирования некоторые обязательные этапы развития. Каждый этап имеет свои характерные признаки и особенности. Результатом этапа является законченный информационный ресурс, соответствующий поставленным на данном этапе целям. Полученные по завершении каждого этапа ресурсы могут дополняться новыми данными при переходе к следующему этапу развития, но могут продолжать расширяться на фиксированном уровне, если это соответствует концепции информационного ресурса. Каждый этап может дополнять новым содержанием уже имеющийся ресурс, но может и породить новый, с иным содержательным наполнением или на новой технологической платформе. Попробуем рассмотреть упомянутые этапы более подробно.

I. Библиографическая база данных трудов сотрудников организации

Содержательной основой библиографической базы данных трудов послужила традиционная бумажная картотека трудов сотрудников. Картотека пополнялась описаниями монографий, сборников и журнальных публикаций, большинство которых институт получал по подписке, и сверялась с годовым отчетом Института. Базовым требованием, предъявляемым к электронной базе данных, являлось обеспечение полноты и точности выборки публикаций автора.

Данный этап формирования базы данных является наиболее хорошо разработанным этапом, многократно реализованным в различных организациях. К тому же, библиографическая база данных лишена юридических ограничений, связанных с авторским правом. Отбор трудов осуществляется из имеющихся в

библиотеке печатных документов. Полученный перечень публикаций сверяется с годовым отчетом организации и дополняется недостающими трудами. Точность и полноту описания публикаций гарантируют службы корпоративной каталогизации и сопутствующий поиск в Интернет. Однако и на этом этапе формирования базы возникает ряд проблем, требующих иногда нетривиального подхода. Одна из таких проблем – наличие разночтений фамилии автора, возникающих при транслитерации фамилии в латиницу и наоборот. Понятно, что для обеспечения полноты выборки публикаций ученого необходимо учитывать все возможные разночтения. Небольшая на первый взгляд проблема не имеет простого решения. Во множестве руководств к базам данных научного цитирования автору предлагается указывать фамилию в том виде, в котором она записана у него в загранпаспорте. Возможно, такой совет хорош в случае, когда автор транслитерирует фамилию сам. Однако на практике издательства, редакции, временные коллективы и госдепартаменты транслитерируют фамилию автора самостоятельно в соответствии с одной из множества имеющихся в мире систем транслитерации. Проблема нашла отражение в ряде публикаций, например, [1]. В общепринятых форматах записей семейства MARC разночтение фамилии автора вводится в подполе поля автора. Следовательно, для учета всех повторений разночтения подполе должно быть повторяющимся, либо это повторение нужно организовывать специальными средствами в достаточном количестве.

Большинство библиотек на данном этапе работы с трудами сотрудников стремились наиболее полно представить в библиотеке все публикации в печатном виде. Однако, при нынешней издательской политике осуществить это достаточно трудно. Снижается количество авторских экземпляров публикации, и автору в конечном итоге предоставляется не более двух экземпляров его работы. Практически умерла практика авторских оттисков. Следствием этих трудностей явилось тенденция использования электронных копий публикаций в процессе формирования фонда трудов. На первом этапе накопления использовались:

- а) Интернет-ссылки на онлайн-документы, лежащие на сайтах редакций и личных сайтах авторов;
- б) отсканированные копии статей, предоставленные автором или полученные по МБА;
- в) электронные варианты публикаций в авторской редакции.

Накопление электронных документов создало предпосылки для перехода к следующему этапу развития информационного ресурса.

II. Библиографическая база данных трудов сотрудников, содержащая ссылки на полный электронный текст

С увеличением доли описаний, содержащих ссылки на полные тексты в базе трудов сотрудников, оформились перспективы приобретения ею новых качеств. Важной составляющей работы по формированию базы данных становилось

требование привязки к записям полных текстов. В отдельных случаях это были тексты в авторской редакции, т.е. последние варианты публикации после согласования текста с рецензентами и редакцией. Собрание электронных текстов становилось «полным пакетом» трудов сотрудников и оформлялось в *электронный архив научных публикаций* организации. По сути, прежняя библиографическая база данных трансформировалась в два достаточно независимых, но связанных с помощью встроенных ссылок информационных ресурса – библиографическую базу данных и полнотекстовый электронный архив. База данных использовалась для управления и поиска в архиве, но обслуживание, организация и правовое оформление архива представлялось отдельной задачей. Решение этой задачи включает несколько обязательных составных частей:

1. Пополнение архива документами. С обыденной точки зрения именно эта задача представляется наиболее трудной – для сбора трудов авторам необходимы побудительные мотивы иногда довольно строгие. В нашей организации архив формировался на основе «Положения об электронном архиве научных публикаций Института», в котором определялись цели и задачи, правила подачи документов в архив, приемлемые форматы, условия хранения документов. «Положение» было обсуждено и принято на Ученом совете Института. Для того чтобы дополнительно стимулировать авторов своевременно предоставлять публикации в архив, были внесены изменения в «Положение о рейтинговых стимулирующих надбавках», закрепляющие факт, что учету в рейтинге подлежат исключительно переданные в архив документы.

2. Доступ к документам архива. Одной из наиболее трудоемких задач является взаимодействие с авторами для обеспечения легитимности электронного архива. В случае, когда все соавторы - сотрудники организации, и работа, отраженная в публикации, выполнялась в период работы сотрудника в Институте, публикация считается служебным произведением, и в силу статьи 1295 ГК РФ Институт является правообладателем. Авторским вознаграждением в этом случае может считаться рейтинговая стимулирующая надбавка в определенном «Положением о рейтинговых стимулирующих надбавках» размере. Однако нельзя не принимать во внимание также авторские договора с издательствами, ограничивающие авторов в правах в течение оговоренного издательством срока. Компромиссный вариант заключается в том, что для документа формируется библиографическое описание, его электронный вариант размещается в архиве, а доступ к нему закрыт на весь срок действия договора с издательством. Сложнее проблема решается, если публикация имеет соавторов, работающих в других организациях. Видимо, с соавторами необходимо заключать договора на передачу неисключительных прав безвозмездно (если Институт выступал ведущей организацией по гранту или договору, по которому соавторы получали вознаграждение) или с точно оговоренной суммой вознаграждения и условиями использования произведения. Еще сложнее проблема решается в случае зарубежной публикации, так как в этом случае авторский договор составлен в соответствии с законодательством той страны, где выходит публикация. В этом случае, по-видимому, необходимо руководствоваться системой классификации

издательств по вопросам самоархивирования [2-3]. В общем случае следует сделать вывод, что вопрос открытости доступа необходимо решать дифференцированно, индивидуально для каждого документа в архиве.

Технически в настоящее время разделение доступа к документам в архиве Института решается средствами, встроенными в САБ ИРБИС версии 2012.1 и предоставляемыми форматом pdf. На документ могут налагаться ограничения на печать и копирование, и, кроме того, каждый документ получает один из 4 статусов доступа:

- а) открытый,
- б) только в локальной сети Института,
- в) только в помещении библиотеки,
- г) нет доступа.

В перспективе на этом этапе развития также возможно решать вопрос о смене (либо параллельном развитии) технологической платформы управления архивом. Например, при наличии доброй воли организации присоединиться к Инициативе архивов открытого доступа (ОАИ) логично оформлять архив на совместимой с международными требованиями программно-технологической платформе, т.е. на основе свободно распространяемого программного обеспечения (DSpace, EPrints и т.п., а также протоколов ОАИ-РМН), реализующего эту инициативу [2-4]. Для перехода на новую программную платформу необходимо обеспечить импорт документов в новую систему в формате DC или MARC21. Для организации полнофункциональной автоматизированной системы дифференциации доступа предпочтительно использовать специализированные Интернет-протоколы (например, LDAP для персональных данных) или их сочетания, либо собственное оригинальное программное обеспечение, а также дополнительные программные средства для решения частных задач.

Отметим, что характерной особенностью этого этапа развития является выделение электронного архива как отдельного информационного ресурса, своеобразного хранилища данных, требующего специфических организационно-технологических методов работы с ним. Дальнейшее развитие библиографическая база данных получает как система управления электронным архивом научных публикаций, что отражено в её названии. Электронный архив информационно обогащает БД, а ссылки на полные тексты делают их неразрывно связанными. Однако данный факт не исключает одновременное использование других систем управления архивом, таких как уже упомянутый ОАИ или полнотекстовая база данных.

III. Электронный архив научных публикаций, дополненный сведениями о библиографии и текущем импакт-факторе журнала-источника публикации

Особенное значение и новый творческий потенциал работа с электронным архивом приобрела с момента выхода в свет постановления Правительства Российской Федерации №312 «Об оценке результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-

конструкторские и технологические работы гражданского назначения» (Собрание законодательства Российской Федерации, 2009, N15, ст. 1841) и приказа Министерства образования и науки Российской Федерации № 406 «Об утверждении типового положения о комиссии по оценке Результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения, и типовой методики оценки результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские технологические работы гражданского назначения». Составной частью методики оценки научной организации является пункт 1.2 «Публикационная активность» раздела 1 «Научный потенциал и эффективность научных исследований». Для оценки публикационной активности методикой определяется ряд показателей: количество публикаций, цитируемость по каждой публикации и импакт-фактор журнала, где опубликована работа. Учитываются только те журнальные статьи, которые зафиксированы ограниченным кругом баз данных научного цитирования. Их перечень достаточно широк для зарубежных и небольшого количества российских публикаций – это Web of Science (WoS), Scopus и целый ряд ресурсов биологического, химического и медицинского направлений. Для учета российских научных публикаций определен единственный ресурс - Российский индекс научного цитирования (РИНЦ), функционирующий на платформе e-library. Необходимость оценивания организации, а также данные, извлеченные из predetermined источников, породили целую волну библиометрических исследований, например, [5-8]. С одной стороны, эти исследования позволяют извлечь массу полезных сведений, направленных на оптимизацию работы библиотеки. С другой стороны, требования методики оценки породили потребность учитывать дополнительные данные о публикациях сотрудников, такие как:

1. Рейтинг журнала в списке источников публикаций, определяемый его текущим импакт-фактором.
2. Пристатейные списки литературы, как источник информации об используемых сотрудниками информационных ресурсах.
3. Источники, цитирующие публикации Института, как фактор подсчета индекса цитирования.

Импакт-фактор журнала – величина динамическая и изменяется ежегодно в соответствии с индексом цитирования публикаций конкретного журнала. То есть, импакт-фактор «привязан» к текущему году, а значит должен регистрироваться вместе с публикацией. Если из WoS можно извлечь импакт-факторы предыдущих лет, то восстановить его для российских публикаций очень затруднительно, если вообще возможно. Как следствие, необходимо иметь специальное поле в библиографическом описании документа, где фиксируется импакт-фактор источника статьи на момент публикации. В имеющихся стандартах такое поле отсутствует.

САБ дает возможность ввода пристатейной библиографии, однако подавляющее большинство библиотек никогда не вводило такие списки в библиографическое описание даже в случае полной аналитической росписи статьи.

Теперь такой ввод становится крайне желательным, но порождает целый ряд вопросов. Например, такого плана: вводить ли библиографию из монографий сотрудников? Такой вопрос способен поставить в тупик любого библиографа в силу фантастической трудоемкости работы и слабо уловимого её смысла. Хотя возможны компромиссные варианты, учитывающие только интересы сбора трудов сотрудников организации, то есть вводить из монографии только ссылки на труды «своих» сотрудников. Учет же источников, цитирующих публикацию, представляется и вовсе нереальным, из-за трудностей отслеживания таких источников даже для единственного активно публикующегося автора. При подсчете цитирований авторов целой организации трудозатраты возрастают многократно. Отметим, что данный этап характерен тем, что все дополнительные данные набираются в базу данных вручную.

IV. Электронный архив научных публикаций как динамически изменяющаяся, многосвязная база данных

Трудоемкость многообразных задач, возникающих на предыдущих этапах развития архива научных публикаций, побудила библиотечное сообщество искать решения в Интернет-технологиях. Справедливости ради следует отметить, что большинство поставленных на этом этапе задач, базы научного цитирования предлагают в качестве дополнительного сервиса, такого как например, ResearcherID в WoS или ScienceIndex в РИНЦ, что нашло отражение в ряде руководств, таких как [9]. Однако большинство таких сервисов рассчитаны на индивидуальное использование авторами, вводимые данные отчуждаются, а сервис для организаций обычно недешево стоит.

В силу вышеуказанных причин многие организации пытаются самостоятельно осваивать и использовать предоставляемые Интернет средства. Эти инициативы порождают много вопросов и затруднений. Одно из них связано с множественностью используемых в сети Интернет идентификаторов документов: часто обсуждаемого идентификатора электронного документа DOI [10], идентификатора документа РИИ, принятого Elsevier, идентификатор AN агрегатора OVID, идентификаторы SCOPUS и РИНЦ, а также множество иных менее известных идентификаторов. Безусловно, в первую очередь нас интересуют идентификаторы баз данных научного цитирования. Но и таких достаточно много. Для чего мы можем их использовать? Обращение к документу при помощи идентификатора позволяет извлекать данные с помощью постоянно действующих запросов и динамически изменять содержимое полей цитирования и импакт-фактора в электронных архивах. Безусловно, сначала для этого необходимо ввести значения идентификаторов в записи документов. На основе идентификаторов можно извлекать данные по цитированию каждой статьи и организовывать связи между документами и списками библиографии. Такие формы работы с базами цитирования используются за рубежом для подсчета различных рейтингов. В России также имеется опыт извлечения информации по цитированию с помощью идентификаторов публикации из WoS и Scopus [8]. К со-

жалению, аналогичная работа с Российским индексом научного цитирования (РИНЦ) затруднительна по причине известных недостатков этой системы, на которые неоднократно указывалось в ряде публикаций и докладов конференций [6-7], хотя в настоящее время заметен определенный прогресс в преодолении этих недостатков, что делает еще более актуальным извлечение различных данных о документе с помощью его идентификатора.

Выводы. Электронный архив научных публикаций, как динамическая, многосвязная база данных является актуальным и востребованным информационным ресурсом. С одной стороны, он обеспечивает сохранность и полное отражение научных трудов организации в единой базе данных, с другой стороны, предоставляет необходимые сведения для достоверной оценки публикационной активности, сверки данных в отчетах и РИНЦ, библиометрических исследований. Отмеченные проблемы при создании такого ресурса ждут своего решения, а предлагаемые программно-технологические разработки – широкого внедрения и распространения.

Литература

1. Главчева, Ю.Н. Создание предпосылок для включения периодических изданий в наукометрические базы данных [Электронный ресурс] / Ю.Н.Главчева // Slideshare. – Харьков, 2011. - URL: http://library.kpi.kharkov.ua/Prezent/5_prerequisites_for_inclusion.pdf
2. Московкин В.М. Институциональные электронные архивы открытого доступа к научному и гуманитарному знанию: структурно-статистический анализ [Текст] [Электронный ресурс] /В.М.Московкин, Д.И.Тюпа // Новый Коллегиум. – 2006. - №5. - С. 3-11. URL: http://admpwd.bsu.edu.ru/_files/Nauka/CoopOnLine/Arch.doc
3. Московкин В.М. Открытый доступ: механизмы, инструменты и первые результаты принятия Белгородской декларации об открытом доступе к научному знанию и культурному наследию [Электронный ресурс] /В.М.Московкин // 9-ая Международная междисциплинарная научно-практическая школа-конференции «Современные проблемы гуманизации и гармонизации управления»: материалы конференции. - Харьков, 2009. – URL: <http://www.myshared.ru/slide/158035/>
4. Мбого, И.А. Инициатива открытых архивов: информационные, методические и технологические аспекты [Текст] [Электронный ресурс] / И.А. Мбого, И.Н. Карнуп, А.В. Чугунов // Интернет и современное общество. - XII Всероссийская объединенная конференция, 28-30 октября 2009 г., Санкт-Петербург. - СПб., 2009. - С. 56-60. - URL: <http://ims2009.nw.ru/thesis/IMS%202tom.pdf>
5. Трескова, П.П. Показатели научной продуктивности и исследовательская активность уральской науки в реферативной базе Web of Science / П.П.Трескова // XIV научно-практический семинар «Информационное обеспечение науки: новые технологии»: материалы семинара. – Екатеринбург, 2010.
6. Свирюкова В.Г. К вопросу о состоянии ресурсной и методической базы для подсчета библиометрических показателей научной активности (индекса цитируемости ученого, организации, журнала) / В.Г.Свирюкова // XIV научно-практический семинар «Информационное обеспечение науки: новые технологии»: материалы семинара. – Екатеринбург, 2010.
7. Мазов, Н.А. Web of Science и Scopus: новые технологические решения для наукометрических исследований персоны (организации) [Электронный ресурс] / Н.А.

- Мазов // Система обеспечения российских организаций научно-технической информацией в электронном виде. Итоги и перспективы проекта МОН: материалы конференции. – СПб., 2012. - URL:
<http://conf.neicon.ru/index.php/science/mon2012/paper/view/29/26>
8. Мазов, Н.А. Идентификация библиографических метаданных научных публикаций в различных базах данных: проблемы и решения [Электронный ресурс] / Н.А. Мазов // Научное издание международного уровня: проблемы и решения при подготовке и включении в индексы цитирования и реферативные базы данных: материалы международной научно-практической конференции. – М.: ВИНТИ РАН, 2012. - URL: <http://www.viniti.ru/download/russian/conf/MAT/16-10.ppt>
 9. Чаплыгин, А.Г. Работа исследователя (автора публикации) с национальными и международными библиометрическими системами: методическое пособие [Текст] [Электронный ресурс] / А.Г. Чаплыгин. М.: МГУ, 2012. – URL:
<http://www.fbm.msu.ru/sci/manuals/bibliometr.doc>
 10. Соловяненко, Д.В. Перспективы внедрения системы DOI в библиотечной сфере [Электронный ресурс] / Д.В. Соловяненко // Научная библиотека академического учреждения: опыт, проблемы и перспективы информационного обеспечения исследовательской деятельности: материалы научно-практической конференции. – Симферополь, 2009. – URL: <http://www.center.crimea.ua/library/colovynenko.pdf>
 11. Каленов, Н.Е. О Российском индексе цитирования [Электронный ресурс] / Н.Е. Каленов // XIV научно-практический семинар «Информационное обеспечение науки: новые технологии»: материалы семинара. – Екатеринбург, 2010.
 12. Ковязина, Е.В. Об оценке деятельности научной организации с помощью Российского индекса научного цитирования (РИНЦ) // SCIENCE ONLINE: электронные информационные ресурсы для науки и образования: материалы конференции. – М., 2010. – URL:
<http://elibrary.ru/projects/conference/morocco2010/presentations/kovyazina.ppt>