

Электронный архив научных публикаций: этапы развития
Digital archive of scientific publications: The stages of development
Електронний архів наукових публікацій: етапи розвитку

Е. В. Ковязина

*Институт вычислительного моделирования СО РАН,
Красноярск, Россия*

Elena Kovyazina

*Institute of Computational Modelling,
Siberian Division of the Russian Academy of Sciences,
Krasnoyarsk, Russia*

О. В. Ковязіна

*Институт обчислювального моделювання СВ РАН,
Красноярськ, Росія*

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.

The main stages of development of the digital archive of scientific publications are highlighted. The characteristic features of each stage are described. A number of problems of organizational and technological nature is listed, and possible solutions for some of them are suggested.

В доповіді виділені основні етапи розвитку електронного архіву наукових публікацій, що визначають трансформацію бази даних праць співробітників організації у систему управління електронним архівом наукових публікацій. Наведено особливості і характерні ознаки кожного етапу. Перераховано ряд проблем організаційного і технологічного характеру, запропоновано можливі шляхи вирішення деяких з них.

Актуальной и перспективной задачей, стоящей перед библиотекой каждого университета или академического института, является формирование библиографической или полнотекстовой базы трудов сотрудников соответствующей организации. Главной 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.

Практически каждая библиотека научной или образовательной организации уже сформировала базу трудов её сотрудников, но результаты этой работы столь сильно различаются по форме и содержанию, что создать, например, интегрированный распределенный ресурс на основе этих отдельных частных ресурсов практически не представляется возможным. В чем причина этих различий? Можно ли прийти к каким-либо соглашениям, позволяющим объединить ресурсы на основе общих критериев? Для ответа на поставленные вопросы полезно исследовать историю развития такого типа ресурса и выделить характерные черты, различающие варианты его реализации.

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 дополняться новым содержанием уже имеющийся ресурс, но может и порождать новый, с иным содержательным наполнением или на новой технологической платформе.

Рассмотрим выделенные этапы более подробно. Отметим, что исследование проводилось на примере базы трудов сотрудников Института вычислительного моделирования СО РАН, реализованной в системе автоматизации библиотек (САБ) ИРБИС и АРМ Ученый секретарь.

1. 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, причем желательно одним запросом.

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. Тем не менее, возник ряд трудностей как технологического, так и организационного порядка. Так, например, отсутствие в САБ ИРБИС повторяющихся подполей не позволяет учесть все возможные разночтения фамилий авторов, которых при транслитерации в латиницу сложной фамилии может быть достаточно много. Данное технологическое ограничение приводит к выпадению некоторых публикаций в процессе поиска. Для обеспечения полноты потребовалось редактирование вложенных рабочих листов авторов с целью включения в них дополнительных повторений разночтений фамилии.

В организационном плане следует отметить, что при нынешней издательской политике трудно обеспечить полноту отражения трудов в базе данных. Снижается количество авторских экземпляров публикации, и автору в конечном итоге предоставляется не более двух экземпляров его работы. Практически умерла практика авторских оттисков. Как следствие, в библиотеку перестали поступать печатные труды сотрудников, что привело к значительным трудностям при сверке данных в библиографических описаниях. Не отличаются точностью и данные, представленные в годовом отчете Института. Определенным выходом из создавшейся ситуации явилось привлечение в процесс формирования библиографических описаний электронных копий публикаций. Для извлечения из публикации необходимых данных использовались в порядке важности:

- 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. Организационно 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» обсуждено и принято на Ученом совете Института. Для того чтобы 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 трудноразрешимых проблем явилось взаимодействие с авторами для обеспечения легитимности электронного архива. В случае, когда все соавторы являлись сотрудниками Института, 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 договоры 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. 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 КНЦ СО РАН,
- 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, EPints и т.п., а также протоколов ОАИ-РМН), реализующего эту инициативу [1]. 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 предопределенных источников, породили целую волну библиометрических исследований, например, [2-4]. 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. Учет же источников, цитирующих публикацию, представляется и вовсе фантастическим, в связи с большой трудоемкостью отслеживания таких источников даже для единственного активно публикующегося автора. При подсчете цитирований авторов целой организации трудозатраты возрастают многократно. Однако 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. В сети Интернет документы определяются с помощью различных идентификаторов, например, идентификатора электронного документа DOI. В базах данных научного цитирования документ также определяется идентификатором. 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. 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 [5]. К сожалению, аналогичная работа с Российским индексом научного цитирования (РИНЦ) невозможна по причине известных недостатков этой системы, на которые неоднократно указывалось в ряде публикаций и докладов конференций [6-7].

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 Института 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. 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>
2. 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.
3. 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.
4. 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>
5. 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>
6. 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.
7. 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>