

Электронный архив научных публикаций: этапы развития

Выделены основные этапы развития электронного архива научных публикаций, определяющие трансформацию БД трудов сотрудников организации в систему управления им. Изложены особенности и характерные признаки каждого этапа. Перечислен ряд проблем организационного и технологического характера и предложены возможные пути решения некоторых из них.

Ключевые слова: электронный архив научных публикаций, базы данных трудов сотрудников, этапы развития, Российский индекс научного цитирования, импакт-фактор журнала.

Актуальной и перспективной задачей, стоящей перед библиотекой каждого университета или академического института, является формирование библиографической или полнотекстовой базы трудов сотрудников. Главная цель такого ресурса – сохранение научного наследия и отражение в публикациях результатов научных исследований. В свете последних государственных инициатив полнота и точность представленных в таком ресурсе данных приобретают жизненно важное значение для организации.

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

Следуя за развитием технологий, база трудов сотрудников проходит от начала формирования некоторые обязательные этапы развития, имеющие свои характерные признаки и особенности. Результат каждого этапа – законченный информационный ресурс, соответствующий поставленным целям. Полученные по завершении каждого этапа ресурсы могут дополняться новыми данными при переходе к следующему этапу развития, но также 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.

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 «Учёный секретарь».

Библиографическая БД трудов сотрудников организации

Содержательной основой библиографической БД трудов послужила традиционная (бумажная) картотека трудов сотрудников. Картотека пополнялась описаниями монографий, сборников и статей из журналов, большинство которых институт получал по подписке, и сверялась с годовым отчётом Института. Базовым требованием, предъявляемым к электронной БД, было обеспечение полноты и точности выборки публикаций автора, причём желательно одним запросом.

Этот этап формирования БД хорошо разработан и многократно реализован в различных организациях. К тому же библиографическая БД лишена юридических ограничений, связанных с авторским правом. Тем не менее возник ряд трудностей – как технологических, так и организационных. Например, отсутствие в САБ ИРБИС повторяющихся подполей не позволяет учесть все возможные разночтения фамилий авторов (а их может быть достаточно много) при транслитерации в латиницу сложной фамилии. Такое 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 БД. Снижается количество авторских экземпляров публикации, и автору в итоге предоставляется не более двух. Кроме того, «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 отчёте Института. Выходом из создавшейся ситуации стало привлечение электронных копий публикаций в процесс формирования библиографических описаний.

Для извлечения из публикации необходимых данных использовались в порядке важности: а) интернет-ссылки на онлайн-документы 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.

Библиографическая БД трудов сотрудников, содержащая ссылки на полный электронный текст

С увеличением в БД трудов сотрудников доли описаний, содержащих ссылки на полные тексты, оформились перспективы приобретения ею новых качеств. Важной составляющей работы по формированию БД 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 БД трансформировалась в два достаточно независимых, но связанных с помощью встроенных ссылок информационных ресурса – библиографическую БД и полнотекстовый электронный архив. БД 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 определены цели и задачи, правила подачи документов в архив, приемлемые форматы, условия хранения документов. «Положение» обсуждено и принято на Учёном совете Института.

Для того чтобы стимулировать авторов своевременно предоставлять публикации в архив, были внесены изменения в «Положение о рейтинговых стимулирующих надбавках», закрепляющие факт – учёту в рейтинге подлежат исключительно переданные в архив документы.

2. *Доступ к документам архива.* Одной из наиболее трудноразрешимых проблем стало взаимодействие с авторами для обеспечения легитимности электронного архива. В случае, когда все соавторы являлись действующими сотрудниками Института, публикация считается служебным произведением, и согласно статье 1295 ГК РФ Институт является её правообладателем. Авторским вознаграждением в этом случае может считаться рейтинговая стимулирующая надбавка в размере, определённом «Положением о рейтинговых стимулирующих надбавках». Однако нельзя не принимать во внимание также авторские договоры с издательствами, ограничивающие авторов в правах в течение оговоренного издательством срока.

Компромиссный вариант заключается в следующем: для документа формируется библиографическое описание, электронный вариант размещается в архиве, но доступ к нему закрыт на весь срок действия договора с издательством.

Проблема решается сложнее, если публикация имеет соавторов, работающих в других организациях. Видимо, с соавторами необходимо заключать договоры на безвозмездную передачу неисключительных прав (если Институт выступал ведущей организацией по гранту или договору, по которому соавторы получали вознаграждение) или с точно оговоренными суммой вознаграждения и условиями использования произведения. Во всех случаях следует сделать вывод: вопрос открытости доступа необходимо решать дифференцированно, индивидуально для каждого документа в архиве.

Технически разделение доступа к документам в архиве Института решается средствами, встроенными в САБ ИРБИС (версия 2012.1) и предоставляемым форматом pdf. На документ могут налагаться ограничения на печать и копирование, и, кроме того, каждый документ получает один из четырёх статусов доступа: открытый; только в локальной сети Красноярского научного центра СО РАН; только в помещении библиотеки; нет доступа.

В перспективе на этом этапе развития также возможно решать вопрос о смене (либо параллельном развитии) технологической платформы управления архивом. Например, при желании 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 [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 БД 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 БД.

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

Особое 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, № 15, 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 в методике определён ряд показателей, в том числе 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 учёта российских научных публикаций определен единственный ресурс – Российский индекс научного цитирования (РИНЦ), функционирующий на платформе *e-library*.

Необходимость оценивания организации, а также данные, извлечённые из predetermined 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 подсчёта 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-фактор за предыдущие годы, 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 даёт возможность ввода пристатейной библиографии, однако подавляющее большинство библиотек никогда не вводило её в библиографическое описание даже в случае полной аналитической росписи статьи. Теперь такой ввод становится крайне желательным. Учёт же источников, цитирующих публикацию, представляется и вовсе фантастическим в связи с большой трудоёмкостью отслеживания таких источников даже для единственного активно публикующегося автора. При подсчёте цитирований авторов целой организации трудозатраты возрастают многократно. Однако этот этап характерен тем, что все дополнительные данные вводятся в БД вручную.

Электронный архив научных публикаций как динамически изменяющаяся, многосвязная БД

Трудоёмкость многообразных задач, возникающих на предыдущих этапах развития архива

научных публикаций, побудила библиотечное сообщество искать решения в интернет-технологиях. В Интернете документы определяются с помощью различных идентификаторов, например DOI – идентификатора электронного документа. 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 подсчёта 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 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 БД – это актуальный и востребованный информационный ресурс. С одной стороны, он обеспечивает сохранность и полное отражение научных трудов Института в единой БД, с другой – предоставляет необходимые сведения для достоверной оценки публикационной активности, сверки данных в отчётах и РИНЦ, библиометрических исследований.

Отмеченные в докладе проблемы при создании такого ресурса ждут решения, а предлагаемые программно-технологические разработки – широкого внедрения и распространения.

Список источников

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 Всерос. объедин. конф., 28–30 окт. 2009 г., С.-Петербург. – С.-Петербург, 2009. – С. 56–60. – Режим доступа: <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 науч.-практ. семинар «Информационное обеспечение науки: новые технологии»: материалы семинара. – Екатеринбург, 2010.
3. **Свирюкова В. Г.** К вопросу о состоянии ресурсной и методической базы для подсчёта библиометрических показателей научной активности (индекса цитируемости учёного, организации, журнала) / В. Г. Свирюкова // Там же.
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 конф. – С.-Петербург, 2012. – Режим доступа: <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 междунар. науч.-практ. конф. – Москва: ВИНТИ РАН, 2012. – Режим доступа: <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 науч.-практ. 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 конф. – Москва, 2010. – Режим доступа: <http://elibrary.ru/projects/conference/morocco2010/presentations/kovyazina.ppt>