



СКОВОРОДИНА Ирина Сергеевна  
Консорциум «Контекстум»,  
директор департамента агрегации  
научно-образовательного контента  
✉ skovorodina@akc.ru



ЕФРЕМОВ Павел Владимирович  
ООО «Национальный цифровой  
ресурс «Руконт»,  
генеральный директор

## ТЕХНОЛОГИИ СОЗДАНИЯ, АГРЕГАЦИИ И ИСПОЛЬЗОВАНИЯ НАУЧНОГО И ОБРАЗОВАТЕЛЬНОГО КОНТЕНТА

*В статье представлен обзор докладов конференции, которая проводилась 30 октября 2013 г. в Вычислительном центре им. А.А. Дородницына РАН. Главная идея конференции – представить концепцию системы агрегации и использования результатов научной деятельности РАН. Авторы статьи сфокусировали внимание на основных темах: создание информационно-технологической платформы; развитие и совершенствование систем агрегации; технологические решения в области поиска информации и анализа научного текста; использование DOI и ORCID.*

**Н**аправление агрегации и использования результатов научной деятельности активно поддерживается Министерством образования и науки РФ в рамках исполнения долгосрочных государственных программ и фундаментальных научных исследований, создания единой федеральной базы данных, включающей результаты НИОКР и технологических работ, выполняемых за счет федерального бюджета, Единого информационного пространства Минобрнауки, а также обсуждается в ведомствах и организациях, которые генерируют и используют научно-образовательный контент.

30 октября 2013 г. в Вычислительном центре им. А.А. Дородницына РАН (ВЦ РАН) состоялась научно-практическая конференция «Технология создания, агрегации и использования научного и образо-

вательного контента», организованная Консорциумом «Контекстум» совместно с ВЦ РАН, Национальной ассоциацией исследовательских и научно-образовательных электронных инфраструктур «e-APEN-A». Конференция проводилась при поддержке Министерства образования и науки Российской Федерации и Российского фонда фундаментальных исследований в рамках НИР «Разработка научно-методических и нормативно-организационных основ создания системы агрегации и использования результатов научной деятельности государственных академий наук». В ходе выполнения НИР организаторами конференции была разработана концепция системы агрегации и использования результатов научной деятельности, функционал которой можно рассматривать как совокупность взаимодействующих компонентов, развивающихся в самостоятельные технические задания и проекты

Основные вопросы, обсуждаемые участниками конференции, были посвящены ключевым проблемам агрегации и обработки больших массивов научной информации, накопленной в научных и образовательных учреждениях, созданию инфокоммуникационных систем, содержащих информацию о научных исследованиях и их результатах, сервисов подготовки и оценки научного контента в рамках современной организации научно-исследовательских работ. В работе приняли участие представители ведущих научных учреждений Российской академии наук, научных библиотек и издательских подразделений федеральных и исследовательских университетов, отраслевых вузов, а также специалисты коммерческих IT-компаний.

Основные выступления были подчинены главной идее конференции – представить концепцию системы агрегации и использования результатов научной деятельности РАН. Были обсуждены пути ее создания и функциональности. Доклады, подготовленные сотрудниками Вычислительного центра РАН, Института системного анализа РАН, Центрального экономико-математического института РАН, Межведомственного компьютерного центра РАН, Института информационных технологий Высшей школы экономики, Центра информационных технологий и систем, Национального фонда подготовки кадров совместно с представителями коммерческих структур – Агентством «Книга-

сервис», Центральным коллектором библиотек «БИБКОМ», входящих в Консорциум «Контекстум», можно рассматривать как предложения по внедрению и дополнению функционала или сервисов для больших объемов данных либо интеграции этой системы с уже существующими другими информационными ресурсами.

Концепция системы агрегации и использования результатов научной деятельности РАН была представлена в совместном докладе **Марата Биктимирова** (зам. директора Вычислительного центра им. А.А. Дородницына РАН) и **Виталия Иванова** (председатель совета директоров Консорциума «Контекстум»). Ее цель – создание информационно-технологической платформы и условий для ее внедрения на основе учета мотиваций участников процессов производства и использования научного контента РАН для качественного и своевременного решения актуальных задач развития науки. Каждый из компонентов концепции может быть вполне самостоятельным проектом: система подготовки и рецензирования научных публикаций, технологии «открытого связывания данных», интеграция с государственными едиными системами управления НИОКР и единой информационной системой Минобрнауки и науки РФ, проект ИРИС по научному наследию России, внедрение идентификатора DOI (Digital Object Identifier), создание сводного цифрового архива российских научных периодических изданий, технологические решения организации лицензионно-договорной работы и ведения реестров авторских прав и др. Но вместе все они составляют единую систему.

На конференции были представлены доклады, связанные с развитием и совершенствованием систем агрегации. Так, например, представители **Центра информационных технологий и систем** предложили интегрировать единую государственную информационную систему учета научно-исследовательских, опытно-конструкторских и технологических работ гражданского назначения (ЕГСУ НИОКР), информационную систему «Карта науки России» и единую информационную систему Минобрнауки с системами РАН для повышения уровня информационного обеспечения сведениями о результатах научной деятельности, осуществляемой учеными, научными коллективами и орга-

низациями государственных академий наук, и мониторинга публикационной активности и использования научного контента, создаваемого учеными, научными коллективами и организациями государственных академий наук.

Интеграцию информационных ресурсов различных информационных фондов (библиотек, архивов, музеев) продемонстрировал представитель **Международного культурного центра РАН** на примере электронной библиотеки «Научное наследие России» – межведомственной, некоммерческой, постоянно действующей и развивающейся информационной системы научно-образовательного назначения на технологической платформе дата-центра.

Новые технологичные решения в области поиска информации и анализа научного текста на примере системы «Eхastus Expert» были предложены **Институтом системного анализа РАН**. Технология позволяет исследовать структуры предметных областей на основе анализа научных публикаций, показать динамику научных исследований на линейке времени, выявить возможное дублирование, плагиат, частичное заимствование, а также отследить преемственность (или выявить факт ее отсутствия), провести автоматическую оценку качества научных текстов (т.е. выявить научные, квазинаучные, ненаучные тексты) и многое другое.

Новое направление в подготовке и использовании научного контента в рамках современной организации научно-исследовательской деятельности представили **Центральный экономико-математический институт РАН** и **Консорциум «Контекстум»**. На примере функционирующей с 2000 года системы «Соционет» показаны возможности анализа результатов научной деятельности отдельного ученого (лаборатории, института), в частности, обработки сведений по использованию результатов научных исследований. Таким образом, возникают новые мощные механизмы самообновления и саморазвития на основе функционирования мощного виртуального агрегатора научных метаданных из большого количества разнородных источников.

Вопросы использования технологий «открытых связанных данных» (Linked open data) и открытого доступа (Open Access) в процессах создания и использова-

ния информационных научных ресурсов поднимались представителями **Института информационных технологий ВШЭ** и **Национального цифрового ресурса «Руконт»**. Открытый доступ дает возможность обработать содержание большого массива контента, определить, насколько один документ похож на другой, независимо от того, в каких предметных областях они созданы. Такой анализ дает возможность создать кластеры, сгруппировать похожие документы.

Использование международного идентификатора цифрового объекта (DOI) и международного открытого идентификатора автора Open Researcher and Contributor Identifier (ORCID) в процессе производства и распространения научного цифрового контента прослежено **Павлом Арефьевым** (Национальный фонд подготовки кадров). Внедрение идентификаторов DOI и ORCID в издательскую и учетную практики в российском научном сообществе повысит экспортный потенциал российской научной периодики, обеспечит выход на международные рынки, будет способствовать исполнению указа Президента № 599 от 7 мая 2012 г. «О мерах по реализации государственной политики в области образования и науки».

**Консорциум «Контекстум»** представил технологические подходы, направленные на решение организационно-правовых вопросов агрегации научного контента. Заключение письменных лицензионных договоров с большим количеством территориально удаленных правообладателей является технологически и организационно сложной задачей. Для решения данной проблемы Консорциумом при поддержке Минобрнауки была разработана информационная технология «Контекстум», обеспечивающая автоматизацию основных процессов ведения лицензионных договоров с правообладателями и соответствующих реестров авторских прав. Данная технология предназначена для решения технологических вопросов автоматизации, получения прав на РИД и обеспечения распространения научных трудов.

На базе ИС «Контекстум» создается целый комплекс интегрированных информационных ресурсов, к их числу относятся отраслевые научно-образовательные электронные библиотеки, региональные цифровые ресурсы, библиотека малоти-

ражных произведений «Золотая полка», представляющих высокую культурную ценность. Кроме применения в качестве универсального инструмента агрегации прав и контента в рамках создания специализированных электронных библиотек и цифровых ресурсов, ИС «Контекстум» может быть использована в качестве системной основы в составе прикладных информационных систем, обеспечивающих агрегацию российских результатов научной деятельности, например, в составе системы агрегации и использования РНД РАН.

Еще один новый проект **Консорциума «Контекстум»** по созданию национального сводного цифрового архива статей российских научных журналов был представлен участникам конференции. Реализация проекта обеспечит доступ ученых к уникальной научной информации, опубликованной в российской периодике. Создание цифрового архива статей реферируемых русскоязычных научных журналов проводится при соблюдении авторских прав в рамках совместного проекта. Партнерами выступают Консорциум «АРБИКОН», научные издательства, российские библиотеки, являющиеся держателями печатных версий научных журналов.

По своим количественным характеристикам проект создания сводного цифрового архива российских журналов превосходит все имеющиеся в мире аналоги. Например, архив американского проекта JSTOR составляет 1000 журналов с глубиной номеров с момента основания, россий-

ский проект охватит около 2000 наименований периодических изданий с аналогичной глубиной архива, но более высокой оперативностью опубликования – запаздывание не более года после выпуска печатной версии (в американском варианте – 3–4 года). Отличительной чертой проектируемого ресурса является предоставляемая пользователю возможность получить электронную копию статьи.

В реализации проекта заинтересованы и библиотеки, которые будут иметь возможность доступа к цифровым архивам периодики, и правообладатели, получающие цифровой архив всех номеров с момента создания журнала. По мере востребованности информации к оцифровке печатной версии архивных номеров издания планируется привлечь библиотеки, располагающие этими фондами.

Следует отметить высокую востребованность подобного ресурса на сегодняшний день, связанную с масштабным переходом сферы науки и образования на электронные носители и преимуществами с точки зрения доступа к информации и удобства ее анализа. Подобный проект, безусловно, поспособствует возвращению в научный оборот значительного массива информации, труднодоступной на сегодняшний момент.

Таков был актуальный круг вопросов, которые рассматривались и обсуждались на конференции. Основные векторы и направления будущей работы по заявленным вопросам определены в финальном документе – Резолюции конференции.



#### **Irina S. SKOVORODINA**

Consortium «Kontekstum», Director of the Department of aggregation of scientific and educational content

#### **Pavel V. EFREMOV**

The National Digital Resource “Rukont” LLC, CEO

### **Technologies of Creation, Aggregation and The Use of Scientific and Educational Content**

The article presents an overview of the conference, which was held October 30, 2013 at the Computer Center named after A.A. Dorodnitsyn RAS. The main idea of the conference - to introduce the concept of aggregation and use of the results of scientific activity of the RAS. The authors focused attention on the major themes: the creation of information-technology platform, the development and improvement of aggregation; technological solutions in the field of information retrieval and analysis of scientific text, the use of DOI and ORCID.