

# Единое российское электронное пространство знаний как ключевая задача информатики в обозримом будущем

А.Б. Антопольский

Институт научной информации по общественным наукам РАН

ale5695@yandex.ru

## Аннотация

Формулируются ключевые проблемы развития научных коммуникаций и информационного пространства науки в обозримом будущем. Обсуждается позиция по этим проблемам мирового научного сообщества, выраженная в ряде деклараций и коллективных докладов. Обсуждается постановка задачи создания «Единого российского электронного пространства знаний», имеющаяся в правительственных документах. Указывается на дифференциацию научных коммуникаций на публикацию и данные, для развития которых стоят различные проблемы. Для публикаций центральной стала проблема открытого доступа, для данных – это проблема их извлечения, представления и повторного использования. Обсуждаются пути модернизации существующих научно-информационных служб. Излагаются предварительные результаты исследований ИНИОН РАН по интеграции информационных ресурсов по языкознанию как части единого цифрового пространства РАН.

**Ключевые слова:** Инфосфера, пространство знаний, информационные ресурсы

**Библиографическая ссылка:** Антопольский А.Б. Единое российское электронное пространство знаний как ключевая задача информатики в обозримом будущем // Информационное общество: образование, наука, культура и технологии будущего. Выпуск 3 (Труды XXII Международной объединенной научной конференции «Интернет и современное общество», IMS-2019, Санкт-Петербург, 19 – 22 июня 2019 г. Сборник научных трудов). — СПб: Университет ИТМО, 2019. С. 9 – 23. DOI: 10.17586/2587-8557-2019-3-9-23

## 1. Введение

Исследования и разработки последних 20-25 лет в научной инфосфере, под которой мы понимаем совокупность ресурсов, сервисов и институций, обслуживающих научную коммуникацию, производят общее впечатление бессистемности и хаотичности, вытекающих из отсутствия у научно-информационного сообщества общего видения будущего научной коммуникации, вытекающего из тенденций ее развития в мировой практике.

Попробуем сформулировать некоторые вопросы, ответы на которые должны определить этот образ будущего.

Как должна выглядеть научная инфосфера через 10-15-20 лет?

Вся ли она будет цифровой?

Какую роль в научной коммуникации должны играть научные и вузовские библиотеки?

Какова перспектива ведущих ресурсов научной информации, таких как Национальная электронная библиотека, Научная электронная библиотека, Электронная библиотека диссертаций, Киберленинка и др.?

Как рационально распределять функции в единой цифровой среде между генераторами, агрегаторами, издателями и другими участниками научной коммуникации?

Вытеснят ли социальные сети другие формы коммуникации?

Нужны ли тематические, корпоративные, региональные и другие электронные библиотеки, и базы данных, или эти функции лучше реализуются в рамках универсальных систем?

Есть ли будущее у библиографических и реферативных служб?

Как будут развиваться информационные языки и онтологии?

Какая экономическая модель будет определять научную коммуникацию?

Удастся ли изменить систему копирайта?

Какие сервисы останутся за человеком, а какие перейдут к искусственному интеллекту?

Какие технологии предпочтительнее развивать и расширять сферу их применения?

Безусловно, без внятного ответа на эти и многие другие вопросы наши усилия по созданию информационных ресурсов, сервисов и институций, на которые сейчас затрачивается большой объем общественного труда, окажутся потраченными впустую.

В мировой практике регулярно делаются коллективные попытки ответить хотя бы на некоторые из этих вопросов. Последним примером может служить доклад группы европейских экспертов «Будущее научных изданий и научных коммуникаций» [1], опубликованный в январе 2019 г.

Подчеркнем, что авторы этого документа в центр проблемы оптимизации научной коммуникации ставят задачу адекватной оценки научных результатов. Методы оценки на основе импакт-факторов журналов и вообще библиометрические методы, очевидным образом приносят неудовлетворительные результаты. А изменение этих методов требует социальных инноваций.

## 2. Единое российское электронное пространство знаний

Возможным ориентиром для решения или хотя бы содержательного обсуждения перечисленных проблем могло бы стать понятие Единого российского электронного пространства знаний (ЕРЭПЗ)<sup>1</sup>, которое появилось в нормативных документах последних лет. В частности, в Федеральном законе «О библиотечном деле» сказано:

«Целью создания Национальной электронной библиотеки являются сохранение исторического, научного и культурного достояния народов Российской Федерации, обеспечение условий для повышения интеллектуального потенциала Российской Федерации и популяризации российской науки и культуры, формирование основы для создания единого российского электронного пространства знаний.

И далее:

Положение о Национальной электронной библиотеке утверждается Правительством Российской Федерации и определяет:

порядок формирования единого российского электронного пространства знаний на основе Национальной электронной библиотеки»

Ранее в некоторых публикациях уже обсуждались отдельные аспекты создания ЕРЭПЗ, в частности, правовые вопросы [2], а также вопросы формирования лингвистического обеспечения ЕРЭПЗ [3]. Попытка сформулировать коллективное мнение об этом понятии представлена в работе [4]. В этих работах понятие ЕРЭПЗ рассматривалось как направление движения различных проектов, охватывающих различные сферы научной коммуникации.

---

<sup>1</sup> Впервые это понятие введено в указе Президента РФ № 808 от 24.12.2014 г. «Об утверждении Основ государственной культурной политики» и затем использовано в Федеральном законе «О библиотечном деле» от 29.12.1994 N 78-ФЗ ст. 18.1 (ред. от 03.07.2016).

Недавно появилось предусмотренное законом Положение о Национальной электронной библиотеке [5], в котором предлагается следующее определение ЕРЭПЗ:

«Пространство знаний - единое российское электронное пространство знаний, представляющее собой совокупность взаимно интегрированных на основе Национальной электронной библиотеки информационных систем и иных информационных ресурсов, сформированных на базе научного, исторического и культурного достояния народов Российской Федерации, образцов зарубежных научных, культурных и исторических ценностей и функционирующих на основе единых информационных технологий и принципов, которые обеспечивают семантическую взаимосвязь их содержимого, а также инструменты поиска и извлечения информации по запросу пользователей».

Это определение носит достаточно общий характер и в принципе тоже могло бы охватывать все области научной коммуникации.

Однако далее в этом документе имеется специальный раздел, посвященный ЕРЭПЗ, который мы приводим полностью:

«III. Порядок формирования пространства знаний на основе Национальной электронной библиотеки

15. Пространство знаний формируется Министерством культуры Российской Федерации на технологической основе Национальной электронной библиотеки.

16. Координация создания и развития пространства знаний возлагается на Межведомственный совет по развитию Национальной электронной библиотеки.

17. Целями формирования пространства знаний являются создание единой, целостной и авторитетной совокупности накопленных человечеством знаний, повышение интеллектуального потенциала Российской Федерации, популяризация российской культуры и науки, в том числе за рубежом.

Приоритет при формировании пространства знаний должен отдаваться документам на русском языке и языках народов Российской Федерации.

18. Основными принципами формирования пространства знаний являются:

а) отсутствие ограничения доступа пользователей к информации, содержащейся в пространстве знаний;

б) достоверность и авторитетность информации;

в) семантическая взаимосвязь информации.

19. Пространство знаний формируется на основе:

а) Национальной электронной библиотеки;

б) информационных ресурсов, содержащих в том числе электронные копии музейных предметов и музейных коллекций музеев Российской Федерации, электронные копии не имеющих ограничений по доступу архивных документов, электронные копии аудиовизуальных документов, находящихся в ведении организаций, уполномоченных на постоянное хранение аудиовизуальных документов в соответствии с законодательством Российской Федерации.

20. Пространство знаний формируется в соответствии с планом, утверждаемым Правительством Российской Федерации, и состоит из следующих этапов:

а) создание технологической основы;

б) создание системы классификации, поиска и извлечения информации;

в) обеспечение на основе Национальной электронной библиотеки электронного взаимодействия информационных систем и информационных ресурсов, входящих в пространство знаний, функционирующих по единым с Национальной электронной библиотекой принципам и правилам;

г) обеспечение семантической взаимосвязи содержащихся в пространстве знаний электронных копий документов посредством использования системы классификации, поиска и извлечения информации, а также информационно-телекоммуникационных сетей, в том числе сети "Интернет".

21. Оператор обеспечивает:

- а) семантическую взаимосвязь электронных копий документов, содержащихся в пространстве знаний;
- б) определение требований к форматам данных и метаописаний, содержащихся в пространстве знаний;
- в) электронное взаимодействие информационных ресурсов, входящих в пространство знаний;
- г) проведение мероприятий по популяризации пространства знаний.

22. Оператор с привлечением экспертных советов обеспечивает энциклопедическую систематизацию отбираемых документов».

По поводу приведенных формулировок возникает множество вопросов, например, в связи с тезисом об «энциклопедической систематизации». Однако обсуждение их в настоящем докладе представляется неуместным. Важно подчеркнуть следующее: авторы документа однозначно отождествляют пространство знаний с собранием библиотечных, архивных, музейных и аудиовизуальных документов. С таким подходом согласиться, конечно, нельзя, хотя понять авторов можно. Они строят концепцию ЕРЭПЗ на основе тех возможностей, которые в настоящее время обеспечивают информационные системы, находящиеся в ведении Министерства культуры, а это почти исключительно документные системы. Авторы также не рассматривают связь пространства знаний с проблемой нового знания и оценки научных результатов, что является главной задачей современной научной коммуникации.

### 3. Научные публикации и научные данные

Современная инфосфера предлагает значительно более широкий спектр информационных продуктов, чем традиционные документные системы. Все более важное место занимают продукты, которые обобщенно называют «данные». Вопросам управления данными, включения их в научные коммуникации, в том числе использования данных для оценки эффективности научной деятельности посвящены многие концептуальные документы, определяющие вектор развития научной инфосферы.

Так, в коллективном документе по управлению открытыми научными данными [6], который разработан консорциумом университетов Великобритании, следующим образом определяется понятие научных данных применительно к социальным и общественным наукам:

«Научные данные это: статистические данные, подборки цифровых изображений, аудиозаписи, стенограммы интервью, данные обследования и наблюдения на местах с соответствующими аннотациями, интерпретации, произведения искусства, архивы, найденные предметы, опубликованные тексты или рукописи и др.».

Гораздо большее разнообразие структур и типов научных данных представлено в информационных ресурсах, отражающих результаты в естественных и инженерных науках.

Понятие научных данных дополняет понятие научных публикаций, которое ранее рассматривалось как основная форма научной коммуникации. Понятно, что пространство знаний, в том числе с тем определением, которое приведено выше, нельзя трактовать только как собрание документов.

Это также ярко проявляется в движении за открытую науку, когда устанавливается две формы раскрытия научных результатов:

- открытый доступ к научным журналам, репозиториям и электронным библиотекам (т.е. к научным документам);
- открытые данные, в том числе первичные результаты научных исследований, чаще всего born digital.

Отметим, что работа с данными никогда не входила в сферу библиотечно-информационной деятельности, ориентированной на работу с публикациями.

Приведем некоторые принципы управления научными данными, сформулированные в цит. выше Конкордате по открытым научным данным:

- хорошее управление данными имеет фундаментальное значение для всех этапов исследовательского процесса и должно быть создано с самого начала;
- управление данными имеет жизненно важное значение для того, чтобы данные были полезными для других и для сохранения данных в долгосрочной перспективе;
- данные, на которые ссылается публикация, должны быть доступны к дате публикации и должны быть в цитируемой форме;
- обучение работе с данными является сферой ответственности всех заинтересованных сторон; администрирование, архивирование, манипулирование и анализ данных требует набора навыков, отличных от тех, которые используются для сбора, генерации или измерения данных.

В дополнение к сказанному можно указать, что в результате мониторинга информационных ресурсов по социальным и гуманитарным наукам, имеющихся в учреждениях РАН и отраженных в Навигаторе информационных ресурсов по общественным наукам (НИРОН) [7], представлено множество типов электронных информационных ресурсов, лишь часть из которых представляет собрания электронных копий традиционных документов. Основные типы этих ресурсов таковы:

- Электронные библиотеки, архивы периодики, труды конференций;
- Библиографии, каталоги, описи;
- Репозитории, другие коллекции неопубликованных документов;
- Оцифрованные архивные и музейные коллекции, фонды, выставки и проч.
- Фактографические и другие объектно-признаковые БД, реестры, регистры, перечни;
- Web-сайты и страницы;
- Корпуса размеченных текстов;
- Лексикографические ресурсы;
- Просопографические ресурсы (сведения о персонах);
- Электронные представления рукописей (кодинологические системы)
- Энциклопедии, справочники, указатели;
- Онлайн-курсы;
- Первичные научные данные (журналы наблюдений и пр.);
- Географические информационные системы
- Экспертные системы, базы знаний
- Банки изображений, аудио- и видеодокументов
- Интерактивные продукты, 3D-модели, продукты виртуальной и дополненной реальности;
- Аккаунты в социальных сетях.

Очевидно, что лишь небольшая часть этих категорий информационных ресурсов охватывается в том варианте пространства знаний, которое определено в цит. выше «Положении о НЭБ».

#### **4. О цифровом пространстве научных знаний**

Таким образом, возникает необходимость в разработке концепции пространства знаний по возможности охватывающая всю совокупность научных информационных продуктов и форм научной коммуникации. Для отличия от понятия пространства знаний, рассмотренного в Положении о НЭБ, предлагается термин Единое цифровое пространство научных знаний (ЕЦПНЗ), как это сделано в работе [3]. Такая концепция, кроме идей организации или интеграции существующих ресурсов, должна предусматривать некую модель оценки поступающих научных документов и выявления нового знания, т.е. схему

пополнения и развития знаний, выходящую за пределы библиотечного и архивного комплектования.

В настоящем докладе делается попытка предложить некоторый план действий по созданию ЕЦПНЗ. Частично моделирование этих действий производится в рамках Комплексного гранта РФФИ, ссылка на который приведена выше.

## 5. Существующее положение

Последние годы автор посвятил исследованию инфосферы академического сектора общественных наук России, включая проведение мониторинга и инвентаризации этих ресурсов. Результаты этих исследований изложены в ряде статей [8-10], а также в монографии, подготовленной в соавторстве с Д.В. Ефременко [11].

Объектом этих исследований является изучение основных тенденций развития инфосферы. Кратко их можно сформулировать следующим образом:

- научные и образовательные коммуникации неуклонно переходят в цифровую форму, хотя этот процесс оказался медленнее ожидаемого и, очевидно, в обозримом будущем сохранятся обе формы коммуникации;
- наука и образование переходят к открытым формам коммуникации (открытому доступу), что существенно меняет экономические модели функционирования научных журналов и других научно-информационных продуктов;
- ресурсы НТИ, отражающие результаты научной деятельности демонстрируют увеличение разнообразия форм или «библиоразнообразия» [12]; роль традиционных публикаций сокращается, а возрастает роль первичных и промежуточных результатов, условно называемых «данными»;
- библиотеки демонстрируют тенденцию к разделению на 2 типа (культурно-досуговый и научно-информационный), причем второй тип функционально сближается с органами НТИ (ГПНТБ России, БЕН РАН, ВПТБ, ЦНМБ, ЦНСХБ и др.);
- генераторами общедоступных ресурсов НТИ стало множество научных и образовательных структур разного уровня, вплоть до отдельных ученых и преподавателей, в результате число таких ресурсов разного типа возросло многократно (десятки тысяч сайтов, ЭБ, БД и других ресурсов);
- ресурсы НТИ играют важную роль в оценке эффективности научных и образовательных организаций (инфометрика, вебометрика, альтметрика), однако принципы такой оценки подвергаются критике;
- появились инструменты (семантические технологии, data mining, антиплагиат и др.), позволяющие существенную долю усилий по обработке и обобщению научной информации.

Заметим, что появление новых форм информационных коммуникаций и ресурсов, практически не привело к исчезновению прежних. Хотя роль некоторых информационных продуктов, например, печатных библиографических и реферативных изданий, значительно сократилась за последние 20 лет. Так что в настоящее время мы наблюдаем сосуществование традиционных и новых форм информационных ресурсов, которое, очевидно, сохранится в обозримом будущем.

Представляется очевидным, что формирование ЕЦПНЗ должно основываться на существующих информационных ресурсах, как традиционных, так и новых, поскольку именно в этих ресурсах зафиксирован опыт многих поколений информационной деятельности по сбору, обработке научной информации и представлению знаний, заключенных в этой информации.

Соответственно способ и форма использования существующих ресурсов в ЕЦПНЗ должны различаться в зависимости от типа ресурсов.

С другой стороны, быстрый рост количества научно-информационных продуктов и ресурсов, потребовало разработки систем их обобщения (интеграции и агрегации). В

последние годы ведется целый ряд проектов, которые призваны в той или иной степени агрегировать (интегрировать) научную и образовательную информации в электронной форме. Перечислим некоторые из них:

- национальная электронная библиотека <https://нэб.пф;>
- национальный энциклопедический портал на основе Большой Российской энциклопедии: <http://www.encyclopedia.ru/cat/books/book/73437/3;>
- национальный портал открытого образования <http://npod.ru/about;>
- электронная библиотека диссертаций <http://diss.rsl.ru;>
- интегратор научной периодики: Научная электронная библиотека и Российский индекс научного цитирования <https://elibrary.ru;>
- интеграторы библиографической информации: ЛИБНЕТ <http://www.nilc.ru>, ЭКБСОН <http://www.vlibrary.ru>, АРБИКОН <https://arbicon.ru>;
- интеграторы материалов научных конференций <http://konferencii.ru;>
- интегратор научных ресурсов открытого доступа Киберленинка <https://cyberleninka.ru;>
- репозитории научно-образовательной информации: НОРА <https://openrepository.ru>, СОЦИОНЕТ <https://socionet.ru>, Федеральная резервная система банков знаний <https://noosphere.ru;>
- государственный каталог Музейного фонда РФ <https://goskatalog.ru/portal;>
- центральный фондовый каталог Росархива <http://cfc.rusarchives.ru/CFC-search;>
- электронная библиотека «Научное наследие России» <http://e-heritage.ru/index.html>.

Одновременно создается множество других ресурсов НТИ. в том числе полнотекстовых ЭБ, энциклопедических, словарных, аналитических, фактографических БД, тематических, проблемных, видовых АИС.

Так, каталог «Наука в Рунете» (<https://elementy.ru/catalog>) содержит около 10 тыс. ссылок, каталог Рамблера топ-100 раздел «Наука» (<https://top100.rambler.ru/navi/?theme=1113>) - 2,3 тыс. ссылок, Навигатор информационных ресурсов по общественным наукам (<http://niron.inion.ru>) - св. 3 тыс. ссылок.

Число российских периодических изданий, индексируемых в РИНЦ - 5.3 тыс., а общее количество российских журналов св. 14. тыс. (<https://elibrary.ru>).

Общая оценка числа электронных информационных ресурсов по науке и образованию, функционирующих в России, - 50-100 тыс., причем их число неуклонно возрастает.

Среди них много информационных систем федерального уровня и значения, которые поддерживаются такими авторитетными организациями как МГУ, ВИНТИ, ИНИОН, РГБ, ГПНТБ России, БЕН РАН, ФИПС, ЦНМБ, ЦНСХБ, Президентская библиотека им. Б.Н. Ельцина, или таких популярных как Википедия, Интегрум, Киберленинка. Однако много и небольших ресурсов, содержащих тем не менее, уникальную научную информацию.

С другой стороны, все эти ресурсы создаются бессистемно и хаотично, и их интеграция в единое информационное пространство представляет собой очень сложную проблему.

Если оценивать нормативную базу национальной системы НТИ, то следует констатировать что Постановление Правительства РФ от 24 июля 1997 г. № 950 «Об утверждении Положения о государственной системе научно-технической информации» в настоящее время, безусловно, устарело, хотя формально действует. Подробнее позиция автора по этому вопросу изложена в монографии [11].

Нужно отметить, что последнее время в рамках вновь принятых нормативных и концептуальных правительственных документов появилось несколько новых проектов, которые можно рассматривать как направленных на организацию национальной научной инфосферы. Так, в Национальном проекте «Наука», в Федеральном проекте «Развитие передовой инфраструктуры для проведения исследований и разработок в Российской Федерации» п. 1.4. сформулирован как «Цифровая система управления сервисами научной инфраструктуры коллективного пользования», включая цифровые коллекции и банки

данных» [13]. Важно отметить также поручение Президента, предусматривающее разработку инфраструктуру научно-технической информации [14].

## 6. Исходные принципы построения ЕЦПНЗ

Далее формулируются некоторые принципы создания ЕЦПНЗ, которые автор считает принципиальными. Конечно, они отражают личное мнение автора и не претендуют на завершенность и истинность.

- ЕЦПНЗ должно формироваться из существующих в России ресурсов научной информации на основе конвергенции, не нарушая их функционирования;
- ЕЦПНЗ должно включать централизованное ядро (онтологию научного знания и Базу знаний) и множество распределенных ресурсов публикаций и данных; ресурсы, используемые для формирования ядра ЕЦПНЗ должны быть отобраны по критериям качества;
- ЕЦПНЗ должно строиться на принципах открытой науки;
- онтология научного знания должна наследовать существующие информационные языки, номенклатуры и системы метаданных;
- информационные объекты, включаемые в Базу знаний ЕЦПНЗ, должны отвечать критериям достоверности, проверяемости, авторитетности и научности;
- технологии формирования ядра ЕЦПНЗ должны обеспечивать возможность коллаборации участников.
- ЕЦПНЗ должно выполнять функции ГСНТИ в части координации деятельности по сбору, обработке, предоставлению в общественный доступ и обеспечения сохранности НТИ.
- ЕЦПНЗ должно использоваться для оценки РНТД, но с учетом модернизации систем оценки.

Целью работ по формированию ЕЦПНЗ является создание современной системы научно-технической информации, ориентированной на решение приоритетных задач социально-экономического и научно-технологического развития России, обеспечение свободного доступа российских пользователей к актуальной, полной, достоверной научной и образовательной информации, создаваемой в России, преимущественно в электронной (цифровой) форме.

Основные задачи ЕЦПНЗ:

- научно-информационное обеспечение достижения целей национального развития России;
- информационное обеспечение стратегического планирования и управления научно-технологическим и инновационным развитием России, разработки и реализации научно-технологической и инновационной политики, приоритетных программ и проектов социально-экономического и научно-технологического развития России;
- координация и оптимизация научно-информационной деятельности в России, в том числе деятельности по созданию научных и образовательных электронных ресурсов;
- международное сотрудничество в области информационного обеспечения науки и образования;
- популяризация и пропаганда науки, научных знаний и научно-технических достижений;
- наукометрические исследования научно-образовательных институций России, включая выявление новых и перспективных научных направлений.

Перечисленные выше задачи ЕЦПНЗ можно назвать внешними. Кроме того, при создании ЕЦПНЗ необходимо обеспечить оптимизацию самой научной инфосферы, при помощи решения задач, которые можно назвать внутренними. К последним можно отнести следующие:

- мониторинг научных и образовательных электронных ресурсов (НОЭР), создаваемых в России;
- экспертизу НОЭР и отбор наиболее качественных НОЭР как составной части ЕЦПНЗ. К этому направлению можно отнести и деятельность по оптимизации системы научных журналов;
- согласованное развитие качественных и востребованных НОЭР;
- организацию доступа к НОЭР, навигацию и поиск в них из единого окна;
- создание новых информационных продуктов и сервисов, обеспечивающих предоставление актуальной, достоверной информации в современной и удобной форме;
- разработку семантических технологий, обеспечивающих обработку и преобразование НОЭР, извлечение знаний, проверку научной информации на новизну и других задач, диктуемых развитием науки;
- согласованную программу оцифровки традиционных НОЭР;
- архивирование и сохранность НОЭР;
- обмен с международными системами представления знаний;
- стандартизацию информационных процессов и продуктов;
- разработку и реализацию экономической модели функционирования научно-информационных сервисов;
- совершенствование правового обеспечения ЕЦПНЗ.

## **7. Принципы и последовательность формирования ЕЦПНЗ**

Изложенное выше современное состояние научной инфосферы диктует стратегию ее развития и формирования ЕЦПНЗ на основе конвергенции существующего хаотического множества информационных систем и ресурсов как упорядоченную совокупность знаний и данных. Реализацию этой стратегии можно представить себе в несколько этапов.

### **7.1. Первый этап – мониторинг инфосферы**

Первый этап формирования ЕЦПНЗ включает проведение инвентаризации научно-образовательной инфосферы, организацию постоянного мониторинга и учета научно-образовательных ресурсов и сервисов, а также научных институций–генераторов и операторов этих ресурсов и сервисов. Результатом первого этапа должен стать общегосударственный каталог научно-образовательных электронных ресурсов. Этот каталог может быть единым, но вероятнее, будут реализованы несколько каталогов по основным научным направлениям:

- естественные и точные науки;
- социальные и гуманитарные науки,
- технические науки;
- медицинские науки;
- сельскохозяйственные науки.

Прототипом такого каталога может послужить созданный в ИНИОН РАН Навигатор информационных ресурсов по общественным наукам (НИРОН) [7].

Конечно, мониторинг научных ресурсов и формирование каталогов может осуществляться и на более детальном уровне – по отдельным научным дисциплинам или по отдельным категориям НОЭР. Например, в настоящее время разрабатывается проект создания специализированного Центра лингвистических ресурсов на базе Института русского языка им. В.В. Виноградова РАН. Очевидно, что первичной задачей этого центра должен быть мониторинг лингвистических ресурсов России, их учет и каталогизация.

Также возможно организовать учет и мониторинг по типам НОЭР, или по региональным отделениям научных учреждений. Например, учет и мониторинг

университетских репозиториях осуществляет НЭИКОН в рамках проекта НОРА. Важно, чтобы мониторинг научно-информационных ресурсов различных типов и категорий осуществлялся согласованно.

Мониторинг инфосферы и формирование каталога должно производиться регулярно, например, ежегодно.

## **7.2. Второй этап - оценка и отбор НОЭР**

В настоящее время количество российских НОЭР, претендующих на научную, образовательную и просветительскую функцию, весьма велико, по экспертной оценке, их число составляет 50-100 тыс., при этом их качество далеко не всегда соответствует современным стандартам и пользовательским требованиям. Поэтому необходим второй этап – оценка и отбор наиболее качественных НОЭР.

Критериями отбора НОЭР должны стать различные показатели – в том числе:

- пользовательские (открытость, видимость, цитируемость, скачиваемость);
- содержательные (уникальность данных, полнота, актуальность, достоверность);
- правовые (легитимность, правовой статус, режим доступа);
- информационно-технологические (качество поиска, стандартные форматы, надежность ПО).

Важным критерием для отбора НОЭР должны стать целесообразность включения НОЭР в процессы интеграции и агрегации, а также их использования для формирования ядра ЕЦПНЗ. Назовем отобранные НОЭР основными ресурсами ЕЦПНЗ.

Отбор НОЭР может включать также их агрегацию. Например, многочисленные электронные библиографические указатели, каталоги и базы данных, создаваемые в научных и образовательных учреждениях, могут входить в состав основных ресурсов НТИ через библиографические агрегаторы, такие как ЛИБНЕТ, ЭКБСОН, АРБИКОН.

Вероятно, в число основных без отбора должны быть включены НОЭР, имеющие официальный статус государственных информационных систем, такие как Национальная электронная библиотека, АИС ФИПС или Государственный каталог Музейного фонда.

Целесообразным является создание специализированных научно-информационных центров по отдельным научным дисциплинам или проблемам, в тех случаях, когда для этого есть условия

Дискуссионным является включение в основные НОЭР российских ресурсов на иностранных языках (например, переводные журналы), НОЭР, создаваемые совместно с иностранными участниками, русскоязычные ресурсы, владельцами которых являются нерезиденты и другие спорные категории ресурсов.

В любом случае методика отбора НОЭР и организация экспертизы НОЭР должны стать результатом широкого обсуждения научно-информационным сообществом и быть зафиксированы в нормативно-техническом документе.

Основные НОЭР, прошедшие экспертизу и отбор, должны получить официальный статус и соответствующую поддержку, включая государственное архивирование. Как уже было отмечено, основные НОЭР должны стать источником ядра ЕЦПНЗ. Вероятно, в рамках экономической модели ЕЦПНЗ основные НОЭР должны получить определенные предпочтения.

## **7.3. Третий этап – формирование ядра ЕЦПНЗ**

Формирование ядра ЕЦПНЗ должно стать третьим и наиболее сложным этапом создания национальной системы научной информации. Понятно, что наличие централизованного ядра придает цельность и системность всему множеству информационных ресурсов и сервисов научной инфосферы. Ядро ЕЦПНЗ включает две части - Онтологию научных знаний и Базу знаний.

Онтология научного знания должна унаследовать терминологию, номенклатуру, парадигматику и концептуальные схемы информационных языков и систем метаданных всех основных НОЭР ЕЦПНЗ, с тем, чтобы обеспечить поиск и навигацию в этих ресурсах из единого окна. Основное условие – формирование онтологии не должно приводить к переработке основных НОЭР. Поиск в основных НОЭР следует обеспечивать с помощью поисковых средств, которые были заложены в этих НОЭР.

Естественно, кроме специфических видов поиска, должен быть предложен и обычный сквозной лексический поиск для тех НОЭР, которые формируют поисковый лексический индекс.

Онтология должна обеспечивать поиск и навигацию в базе знаний ЕЦПНЗ.

База знаний ЕЦПНЗ представляет собой совокупность энциклопедических и справочных сведений, представленных как в виде формализма, пригодного для компьютерной обработки, так и в человекочитаемом виде. В качестве прототипа базы знаний ЕЦПНЗ можно указать на DBpedia.

Процедуру формирования исходной версии Базы знаний можно представить себе следующим образом:

- формирование перечня наиболее авторитетных и популярных энциклопедий и справочников по различным наукам;
- загрузку этих энциклопедий в единую среду, их интеграция;
- выявление противоречий, ошибок;
- содержательное редактирование выявленных ошибок;
- индексирование информационных объектов Базы знаний при помощи онтологии;
- разработка технологии пополнения Базы знаний.

База знаний может быть, как централизованной, так и разделенной на области наук. В последнем случае должна быть сформирована междисциплинарная (общенаучная) база знаний.

Формирование Базы знаний должно осуществляться на принципах открытости, актуальности, достоверности, авторизованности, тематического универсума НТИ.

При формировании Базы знаний могут различаться просветительский, базовый (образовательный) и профессиональный уровни. Навигация в Базе знаний на основе онтологии также должна различаться по этим уровням.

В Базе знаний должна быть возможность отражения различных (в т. ч. противоречивых) научных теорий и концепций. Однако должен быть предложен механизм предотвращения включения в Базу знаний лженаучной и паранаучной информации. Очевидно, решающая роль в этом должна принадлежать РАН, как основной экспертной организации страны.

Информация, содержащая государственную или коммерческую тайну, в Базу знаний не включается.

Принципиально важно, чтобы ядро ЕЦПНЗ разрабатывалось на основе коллаборативных технологий, чтобы в их формировании приняли участие любые ученые России, независимо от их места работы или статуса.

На основе базы знаний и онтологии могут разрабатываться и реализовываться дополнительные (монетизированные) сервисы, сокращающие затраты на функционирование ЕЦПНЗ.

#### **7.4. Четвертый этап – выявление нового знаний**

Четвертым этапом формирования ЕЦПНЗ является разработка системы выявления нового знания на основе потока научной информации, генерируемого российскими научными организациями.

Эта система должна быть основана на существующих информационно-аналитических и реферативных службах. Специалисты-аналитики должны иметь в своем распоряжении сервис, позволяющий определить наличие новизны в документе или ином научно-

информационном продукте по сравнению с Базой знаний. Если такая новизна обнаружена, новые информационные объекты (энциклопедические статьи и справочные данные) вводятся в Базу знаний. При появлении новых понятий, терминов, имен или связей между ними они также вводятся в онтологию.

Этот сервис должен опираться на использование наиболее развитых семантических технологий.

Процедуры выявления нового знания и включение новых информационных объектов в ядро ЕЦПНЗ должны непосредственно отражаться в критериях оценки эффективности научной деятельности. Система оценки научных результатов на основе выявления нового знания должна постепенно заменять существующие способы оценки на основе публикационной активности и цитируемости, которые демонстрируют устойчивое снижение адекватности и полезности.

Очевидна сложность и противоречивость предлагаемого подхода. Однако необходимость формирования новой системы научно-технической информации, отражающей реальность современного этапа научно-технического развития, и цифровой среды научной коммуникации, не вызывает сомнений.

Четвертый этап реально может быть реализован только после появления приемлемого прототипа Базы знаний. До тех пор существующая система аналитических и реферативных служб должна сосуществовать с процедурами формированием Базы знаний.

## **8. Принципы и последовательность формирования ЕЦПНЗ**

Моделирование некоторых аспектов создания ЕЦПНЗ предусмотрено в рамках Комплексного гранта РФФИ «Исследование и разработка принципов, методов и средств конвергенции естественнонаучных и социогуманитарных ресурсов как составляющих единого электронного пространства знаний». В этом гранте участвуют 4 организации РАН (ФИЦ ИУ, НИИСИ, БЕН и ИНИОН), каждая из которых реализует свой частный проект. В данном докладе описываются исследования, которые выполняются ИНИОН РАН совместно с Институтом русского языка РАН в рамках гранта № 18-00-00218 «Интеграция научно-информационных ресурсов учреждений РАН (на примере языкознания) как части единого цифрового пространства РАН». В рамках этого гранта предусмотрено выполнение следующих работ:

- Разработка стратегии для действий создателей информационных ресурсов РАН в области языкознания по интеграции информационных ресурсов в единое цифровое информационное пространство РАН.
- Разработка методики оценки научных ресурсов РАН на примере ресурсов по языкознанию.
- Проведение инвентаризации, классификации и каталогизации информационных ресурсов РАН по языкознанию.
- Участие в комплексном проекте по разработке онтологии научного знания на основе имеющихся в ИНИОН РАН и ИРЯ РАН информационных ресурсов по языкознанию, тезаурусов, рубрикаторов и систем метаданных.

Стратегия интеграции научно-информационных ресурсов должна строиться на основе анализа современного состояния инфосферы по гуманитарным и социальным наукам и прогноза ее развития на обозримый период. Стратегия должна включать комплекс, правовых, экономических, технологических и организационных мер, обеспечивающих в перспективе интеграцию научно-информационных ресурсов учреждений РАН в цифровое информационное пространство знаний. В стратегии должен быть учтен опыт предыдущих разработок - ЕНИП РАН и ИСИР РАН. Однако в условиях современного состояния инфосферы РАН, дорожная карта разрабатывается впервые. В качестве модели для реализации подобной стратегии разрабатывается концепция создания Центра лингвистических ресурсов РАН.

Методика оценки качества научно-информационных ресурсов должна строиться на основе сочетания экспертных и инфометрических (альтметрических, вебометрических и др.) показателей, обеспечивающих комплексную оценку научных информационных ресурсов. Методика оценки качества научно-информационных ресурсов должна учитывать наиболее перспективные типы научно-информационных ресурсов (электронные библиотеки, базы данных, интерактивные и коллаборативные сервисы и др.). Методика должна быть ориентирована на использования в различных процессах интеграции, обеспечения сохранности информационных ресурсов и управления инфосферой. Методика оценки качества научно-информационных ресурсов разрабатывается впервые.

Мониторинг научно-информационных ресурсов учреждений РАН должен предусматривать различные методы сбора данных, обеспечивающих полноту и достоверность результатов инвентаризации, и завершаться формированием электронного каталога информационных ресурсов с фиксацией их организационной принадлежности, тематики и структурного типа. Мониторинг научно-информационных ресурсов учреждений РАН проводится впервые

Участие в комплексном проекте по интеграции ресурсов должно предусматривать формирование фрагментов различных категорий информационных ресурсов по лингвистической тематике, имеющих в ИНИОН и ИРЯ РАН, в соответствии с требованиями единой онтологии научного знания. Форматы и состав метаданных фрагментов различных категорий информационных ресурсов разрабатываются совместно с другими участниками комплексного проекта. Также совместно осуществляется формирование онтологии на основе представленных лингвистических средств и систем метаданных и загрузка представленных образцов в единую программно-технологическую среду. Комплексный проект интеграции учитывает опыт предыдущих разработок - ЕНИП РАН и ИСИР РАН, но обладает существенной новизной с учетом состава информационных ресурсов и новых технологий формирования онтологии и единого информационного пространства РАН.

## Литература

- [1] Future of Scholarly Publishing and Scholarly Communication: Report of the Expert Group to the European Commission. URL: <https://publications.europa.eu/en/publication-detail/-/publication/464477b3-2559-11e9-8d04-01aa75ed71a1>.
- [2] Вершин А.П. Единое российское электронное пространство знаний: вопросы права // Информационно-аналитический журнал «Университетская книга». 2016. URL: <http://www.unkniga.ru/biblioteki/bibdelo/6630-edinoe-rossiyskoe-elektronnoe-prostranstvo-znaniy-voprosy-prava.html>.
- [3] Антопольский А.Б., Белоозеров В.Н., Маркарова Т.С. О разработке онтологии на основе классификаторов научной информации и терминологических словарей // Информационные ресурсы России. 2017. № 5. С. 2-7.
- [4] Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О Едином цифровом пространстве научных знаний // Вестник РАН. 2019. Т.89, № 7. С. 728-735.
- [5] Положение о федеральной государственной информационной системе «Национальная электронная библиотека». Утверждено Постановлением Правительства РФ от 20 февраля 2019 г. № 169. URL: <http://government.ru/docs/35825/>.
- [6] The Concordat on Open Research Data, 1916. URL: <https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf>.
- [7] Навигатор информационных ресурсов по общественным наукам. URL: Режим доступа: <http://niron.inion.ru>.
- [8] Антонопольский А.Б. Инфосфера общественных наук: структура, границы, функции // Научно-техническая информация. Сер.1. 2017. № 4.

- [9] Антопольский А.Б. Информационные ресурсы общественных наук. Опыт организации мониторинга // Библиосфера. 2017. № 3. С. 78-84.
- [10] Антопольский А.Б. Определение границ при проведении мониторинга информационных ресурсов социально-гуманитарных наук // Информационные ресурсы России. 2017. №3. С. 6-10.
- [11] Антопольский А.Б. Ефременко Д.В. Инфосфера общественных наук России: монографии / под ред. Цветковой В.А. М.: Директ-Медиа, 2017. 676 с. ISBN 978-5-4475-9218-9. DOI 10.23681/468227.
- [12] Призыв Жюсье к открытой науке и библиоразнообразию. URL: <https://jussieucall.org/jussieu-call/>.
- [13] Перечень поручений по итогам заседания Совета при Президенте по науке и образованию, состоявшегося 27.11.2018 г. Утвержден 29 декабря 2018 года. Содержит 12 поручений. Пр-2558, п.3 а) 2 «О создании современной цифровой инфраструктуры для хранения и анализа научно-технической информации, а также для обмена такой информацией». URL: <http://kremlin.ru/acts/assignments/orders/59632>.
- [14] Паспорт национального проекта «Наука». URL: <http://government.ru/projects/selection/740/35565/>.

## **Common Russian Electronic Knowledge Space as a Key Task of Informatics in the Foreseeable Future**

A.B. Antopolskii

Institute of Scientific Information on Social Sciences  
of the Russian Academy of Sciences

Key problems of development of scientific communications and information space of science in the foreseeable future are formulated. The position of the world scientific community on these problems, expressed in a number of declarations and collective reports, is discussed. The problem of creating a "Common Russian electronic knowledge space", which is available in government documents, is discussed. The differentiation of scientific communications into publication and data, for the development of which there are various problems, is pointed out. For publications, the problem of open access has become Central; for data, it is the problem of extracting, presenting and reusing them. The ways of modernization of existing scientific and information services are discussed. The preliminary results of the INION RAS research on the integration of information resources in linguistics as part of the unified digital space of the RAS are presented.

**Keywords:** InfoSphere, knowledge space, information resources

**Reference for citation:** Antopolskii A.B. Common Russian Electronic Knowledge Space as a Key Task of Informatics in the Foreseeable Future // Information Society: Education, Science, Culture and Technologies of the Future. Vol. 3 (Proceedings of the XXII International Joint Scientific Conference «Internet and Modern Society», IMS-2019, St. Petersburg, June 19-22, 2019). - St. Petersburg: ITMO University, 2019. P. 9 – 23. DOI: 10.17586/2587-8557-2019-3-9-23

## **Reference**

- [1] Future of Scholarly Publishing and Scholarly Communication: Report of the Expert Group to the European Commission // URL: <https://publications.europa.eu/en/publication-detail/-/publication/464477b3-2559-11e9-8d04-01aa75ed71a1>.
- [2] Vershin A.P. Edinoe rossijskoe elektronnoe prostranstvo znaniy: voprosy prava // Informacionno-analiticheskij zhurnal «Universitetskaya kniga». 2016. URL:

- <http://www.unkniga.ru/biblioteki/bibdelo/6630-edinoe-rossiyskoe-elektronnoe-prostranstvo-znaniy-voprosy-prava.html> (In Russian).
- [3] Antopol'skij A.B., Beloozerov V.N., Markarova T.S. O razrabotke ontologii na osnove klassifikatorov nauchnoj informacii i terminologicheskikh slovarej// Informacionnye resursy Rossii. 2017, № 5, S. 2-7 (In Russian).
- [4] Antopol'skij A.B., Kalenov N.E., Serebryakov V.A., Sotnikov A.N. O Edinom cifrovom prostranstva nauchnyh znaniy // Vestnik RAN. 2019. Vol. 89, № 7. P. 728-735. (In Russian).
- [5] Polozhenie o federal'noj gosudarstvennoj informacionnoj sisteme "Nacional'naya elektron'naya biblioteka". Utverzhdeno Postanovleniem Pravitel'stva RF ot 20 fevralya 2019 g. № 169. URL: <http://government.ru/docs/35825/> (In Russian).
- [6] The Concordat on Open Research Data, 1916 // URL: <https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf>.
- [7] Navigator informacionnyh resursov po obshchestvennym naukam// URL: Rezhim dostupa: <http://niron.inion.ru> (In Russian).
- [8] Antonopol'skij A.B. Infosfera obshchestvennyh nauk: struktura, granicy, funkcii// Nauchno-tehnicheskaya informaciya, ser.1, 2017, № 4 (In Russian).
- [9] Antopol'skij A.B. Informacionnye resursy obshchestvennyh nauk. Opyt organizacii monitoringa // Bibliosfera, 2017, № 3, S. 78-84 (In Russian).
- [10] Antopol'skij A.B. Opredelenie granic pri provedenii monitoringa informacionnyh resursov social'no-gumanitarnyh nauk // Informacionnye resursy Rossii, 2017, №3 S. 6-10 (In Russian).
- [11] Antopol'skij A.B. Efremenko D.V. pod red. Cvetkovej V.A. Infosfera obshchestvennyh nauk Rossii: monografii. M.: Direkt-Media, 2017. S. 676 ISBN 978-5-4475-9218-9. DOI 10.23681/468227 (In Russian).
- [12] Prizyv ZHyus'e k otkrytoj nauke i biblioraznoobraziyu // URL: <https://jussieucall.org/jussieu-call/> (In Russian).
- [13] Perechen' poruchenij po itogam zasedaniya Soveta pri Prezidente po nauke i obrazovaniju, sostoyavshegosya 27.11.2018 g. Utverzhden 29 dekabrya 2018 goda. Soderzhit 12 poruchenij. Pr-2558, p.3 a) 2 «O sozdanii sovremennoj cifrovoj infrastruktury dlya hraneniya i analiza nauchno-tehnicheskoy informacii, a takzhe dlya obmena takoj informaciej» // URL: <http://kremlin.ru/acts/assignments/orders/59632> (In Russian).
- [14] Paspport nacional'nogo proekta«Nauka» // URL: <http://government.ru/projects/selection/740/35565/> (In Russian).