

## О ЕДИНОМ ЦИФРОВОМ ПРОСТРАНСТВЕ НАУЧНЫХ ЗНАНИЙ

© 2019 г. А.Б. Антопольский<sup>1\*</sup>, Н.Е. Каленов<sup>2\*\*</sup>,  
В.А. Серебряков<sup>3\*\*\*</sup>, А.Н. Сотников<sup>4\*\*\*\*</sup>

<sup>1</sup>Институт научной информации по общественным наукам РАН, Москва, Россия

<sup>2</sup>Библиотека по естественным наукам РАН, Москва, Россия

<sup>3</sup>Вычислительный центр ФНЦ "Информатика и управление" РАН, Москва, Россия

<sup>4</sup>Межведомственный суперкомпьютерный центр РАН – филиал ФНЦ "Научно-исследовательский институт системных исследований", Москва, Россия

\*E-mail: ale5695@yandex.ru, \*\*E-mail: nek@benran.ru,

\*\*\*E-mail: serebr@ultimeta.ru, \*\*\*\*E-mail: asotnikov@jscc.ru

Поступила в редакцию 19.11.2018 г.

Поступила после доработки 21.11.2018 г.

Принята к публикации 26.11.2018 г.

В статье предпринимается попытка комплексного анализа понятия и феномена единого цифрового пространства научных знаний (ЕЦПНЗ) как более узкой формулировки понятия единого российского электронного пространства знаний, которое предложено в ряде правительственных документов. Предлагаются определения основных понятий, связанных с данным феноменом. Рассматриваются цели, задачи, состав, структура, границы и принципы создания ЕЦПНЗ. Приводится краткий обзор существующих научных информационных ресурсов и сервисов, которые имеют черты элементов пространства научных знаний. Делается вывод, что имеющиеся информационные ресурсы могут быть источниками ЕЦПНЗ, но не его составляющими. Предлагается структура ЕЦПНЗ, включающая статическую и динамическую составляющие (базис и надстройку), различные уровни представления знания, а также ядро и периферию. При этом ядро должно включать формализованные знания, над которыми задано множество операций, а периферия – источники этих знаний (в том числе документы), отобранные по критериям качества. Логическую структуру ЕЦПНЗ должна представлять его онтология, сформированная на основе информационных языков и метаданных существующих научных информационных ресурсов. Предлагается ряд шагов по организации работы по созданию ЕЦПНЗ, а также его наполнению и поддержке.

*Ключевые слова:* информационные ресурсы, цифровое пространство, научные электронные библиотеки, сервисно-поисковые системы, онтология научного знания.

DOI: <https://doi.org/10.31857/S0869-5873897728-735>

В ряде нормативных документов последних лет появилось понятие *единое российское электронное пространство знаний* (ЕРЭПЗ)<sup>1</sup>, которое рассматривается как некоторая цель или перспектива создания, развития и интеграции информацион-

ных ресурсов и сервисов в области науки, культуры и образования в современной цифровой среде. Однако до сих пор не опубликовано ни одного документа, в котором бы присутствовала содержательная трактовка этого понятия и были бы рассмотрены цели, задачи, состав, структура, границы и принципы создания этого пространства, хотя в некоторых публикациях уже обсуждались отдельные аспекты создания ЕРЭПЗ, в частности,

АНТОПОЛЬСКИЙ Александр Борисович – доктор технических наук, главный научный сотрудник ИНИОН РАН. КАЛЕНОВ Николай Евгеньевич – доктор технических наук, главный научный сотрудник БЕН РАН. СЕРЕБРЯКОВ Владимир Алексеевич – доктор физико-математических наук, главный научный сотрудник ВЦ ФИЦ "Информатика и управление" РАН. СОТНИКОВ Александр Николаевич – доктор физико-математических наук, заместитель директора МСЦ РАН – филиал ФНЦ НИИСИ РАН.

<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 № 78-ФЗ, ст. 18.1 (ред. от 03.07.2016).

правовые [1], а также вопросы лингвистического обеспечения [2].

В настоящей статье делается попытка подойти к этой проблеме комплексно, с позиций системного анализа.

- Здесь рассматривается не пространство знаний в целом, а его научная часть (подпространство научных знаний), тем самым выделяются те информационные объекты, которые верифицированы мировым научным сообществом, они отделяются от информационных объектов, которые носят идеологический, религиозный и другой спорный с научной точки зрения характер.

- Из наименования предмета обсуждения исключается признак *российский*, поскольку научное знание трансгранично по своему существу и выделение российского фрагмента информационного научного пространства в контексте последующих рассуждений не имеет смысла.

- Представляется более адекватной и соответствующей современным представлениям формулировка *цифровое*, а не *электронное* пространство.

Таким образом, предметом рассмотрения в данном случае будет *единое цифровое пространство научных знаний* (ЕЦПНЗ).

**Об основных понятиях.** Говоря о создании цифрового пространства научных знаний, необходимо прежде всего определить каждое понятие, входящее в это словосочетание. Пространство в его математическом (формализованном) понимании есть совокупность элементов с определёнными над ними операциями (векторное пространство, Гильбертово пространство, Банахово пространство и пр.). Элементами цифрового научного пространства должны выступать некоторые формализованные характеристики научного знания. Конкретные научные знания специфичны для отдельных областей науки. Но почти в каждой области можно выделить два класса знаний — теоретические и экспериментальные.

*Теоретические знания* — совокупность постулатов, методов исследования и базовых результатов, полученных на основе применения этих методов. Необходимо отметить, что если постулаты и методы исследований являются формальными, декларированными и признанными обществом устоявшимися элементами, то базовые результаты теоретических исследований — понятие относительное, которое определяется экспертными оценками научного сообщества.

*Экспериментальные знания* — совокупность практических результатов применения теоретических знаний, подтверждённых практикой.

Очевидно, что такое деление относительно, поскольку множество теоретических знаний возникло в результате экспериментов (законы физики, механики и других естественных наук).

Под ЕЦПНЗ мы будем понимать компьютерную среду, обращаясь к которой пользователь (начиная от учёного и кончая учеником средней школы) должен получить ответы на вопросы, касающиеся тех или иных областей науки. Эта среда должна содержать достоверную фундаментальную и научно-популярную информацию; состоять из ряда подпространств, вообще говоря, связанных между собой, относящихся к отдельным научным направлениям; каждое подпространство должно иметь фундаментальную (статическую) основу, включающую базовые постулаты и результаты данного научного направления, и динамическую часть, включающую информацию о новейших достижениях в данной научной области.

Прежде чем перейти к анализу принципов построения и функциональности ЕЦПНЗ, рассмотрим существующие элементы мирового научного информационного пространства, которые в той или иной степени обладают вышеперечисленными свойствами ЕЦПНЗ.

**Существующее положение.** В последние десятилетия происходит быстрое развитие научных информационных ресурсов и сервисов, которые образуют научное информационное пространство. Возникают их новые типы, меняют свою форму и содержание традиционные ресурсы и способы коммуникации. Подробный анализ современных тенденций развития научного информационного пространства применительно к сфере общественных наук содержится в монографии [3]. Существенное влияние на научную коммуникацию оказывают такие процессы, как цифровизация (дигитализация) научной коммуникации, включая научные издания. Практически все научные журналы представлены в электронной форме.

Важным фактором стало возникновение специальных научных и научно-практических дисциплин, основное содержание которых — формирование и использование цифровых научных ресурсов. Примерами могут служить биоинформатика, геоинформатика, историческая информатика, компьютерная лингвистика и др. Одна из основных тенденций современных научных информационных процессов — движение в направлении открытой науки, изменяющей экономические и организационные формы научной коммуникации. Активно формируются репозитории открытого доступа, которые начали теснить традиционные журналы. Бурно развиваются сетевые информационные сервисы. Прежде всего это научные социальные сети, такие как Academia ([www.academia.edu](http://www.academia.edu)), ResearchGate ([www.researchgate.net](http://www.researchgate.net)), Social Science Research Network ([www.ssrn.com](http://www.ssrn.com)), Mendeley ([www.mendeley.com](http://www.mendeley.com)), Citeulike ([www.citeulike.org](http://www.citeulike.org)), Myexperimen

([www.myexperiment.org](http://www.myexperiment.org)). Примерами российских научных социальных сетей являются Соционет (<https://socionet.ru>), Учёные России ([www.russian-scientists.ru](http://www.russian-scientists.ru)), Научная сеть ([www.scipeople.ru](http://www.scipeople.ru)), Социальная научная сеть ([www.science-community.org](http://www.science-community.org)). Научные социальные сети предоставляют многочисленные сервисы по обмену разнообразной информацией, размещению результатов исследований, их рецензированию и комментированию, поиску вакансий и т. п.

Активно формируются и развиваются научные электронные библиотеки, которые также предоставляют широкому исследовательскому сообществу разнообразные сервисы, в частности:

- быстрый доступ к необходимым информационным ресурсам, в том числе поиск по различным дисциплинам и источникам как в специальных базах данных, так и в полнотекстовых электронных изданиях многих крупных научных издательств мира;
- использование новых технологий научных исследований, эффективного инструментария для их проведения; предоставление различных встроенных или надстроенных сервисов — приложений, превращающих электронные библиотеки в исследовательский полигон (виртуальная обсерватория, виртуальная химическая или биологическая лаборатория и т. п.);
- сохранность научной информации и предотвращение утраты ценных научных коллекций для будущих поколений учёных;
- обеспечение возможностей научного сотрудничества не только в региональном, ведомственном, национальном, но и в международном масштабе.

Примером современной научной электронной библиотеки может служить Национальная цифровая библиотека по науке США (The National Science Digital Library — NSDL, <https://nsdl.orgcommons.org>), которая основана Национальным научным фондом (NSF) в 2000 г. с целью обеспечить доступ к сервисам высокого качества по многим областям науки, технологий и техники. NSDL агрегирует ресурсы из цифровых библиотек и проектов, поддерживаемых NSF, а также формирует собственные обзоры вебсайтов по своим направлениям.

Мощные электронные библиотеки поддерживают крупнейшие зарубежные научные издательства Elsevier, Springer, Wiley и др., предоставляя коммерческий доступ к издаваемым ими журналам и книгам.

Наиболее популярной в научной среде является открытая электронная библиотека SciHub, однако она оперирует в основном нелегитимным контентом, который формируется путём скачивания и получения от пользователей электронных копий статей, защищённых авторскими правами.

В качестве примера отечественной электронной библиотеки можно привести научную электронную библиотеку (<https://elibrary.ru/>), содержащую метаданные о нескольких миллионах отечественных публикаций и предоставляющую коммерческий и свободный доступ к значительному количеству их полных текстов. В последние годы развивается электронная библиотека "Киберленинка" (<https://cyberleninka.ru/>), построенная на основе принципов открытой науки, содержащая более 1,3 млн полных текстов.

Основным недостатком как зарубежных, так и отечественных электронных библиотек является примитивный поисковый интерфейс, построенный по аналогии с поисковыми машинами типа Google и не обеспечивающий приемлемую точность поиска. Зачастую поисковые запросы чувствительны к количеству присутствующих в них пробелов, что свидетельствует о низком качестве их программного обеспечения. Так, если попытаться найти в "Киберленинке" книгу Д. К. Фаддеева "Вычислительные методы линейной алгебры", вводя в строку запроса "Фаддеев Д. К. Вычислительные методы линейной алгебры", электронная библиотека выдаст 30 документов; если же между инициалами вставить пробел, будут выданы 80 записей, но ни среди первой партии, ни среди второй искомым книги не будет.

Своего рода исключением из общего числа электронных библиотек является "Научное наследие России" [4–6]. Эта библиотека создавалась изначально как объектно-ориентированная информационно-поисковая система, интегрирующая научные ресурсы институтов памяти (библиотек, архивов, музеев) и предоставляющая возможность многоаспектного поиска таких объектов, как "персона", "публикация", "музейный предмет", "архивный документ", и навигации по найденным ресурсам.

Наряду с электронными библиотеками в Интернете представлены многочисленные коммерческие и некоммерческие научные сервисно-поисковые системы, среди наиболее известных из которых можно выделить следующие.

*Google Scholar* — Академия Google (<http://scholar.google.ru/>) поисковая система научной литературы. Используя единую форму запроса, сбор информации о свободно скачиваемых научных публикациях, Академия Google позволяет найти работу, наиболее точно соответствующую запросу.

*Scholar.ru* — поисковая система научных публикаций (<http://www.scholar.ru>). Проект был создан для упрощения поиска документов научной тематики на русском языке, в первую очередь подготовленных в России. Основная цель этой поисковой системы — сбор информации о свободно скачиваемых научных публикациях. Проект

не рассчитан на хранение полных текстов статей в том или ином виде, вместо этого используется база ссылок на тексты документов с информацией о публикациях (аннотация, авторы и т.д.).

*Science Research Portal* (<http://www.science-research.com/scienceresearch>) – научная поисковая система, осуществляющая полнотекстовый поиск в журналах многих крупных научных издательств, таких как Elsevier, Highwire, IEEE, Nature и др.

*CiteSeer Publications Research Index* (<http://citeseer.ist.psu.edu/>) – научная поисковая система, индексирующая статьи в PostScript и PDF формате с научных веб-сайтов.

*Ingenta* (<http://www.ingentaconnect.com/>). Библиографическая база данных универсального типа, включающая описания статей из англоязычных журналов и сборников. Объём базы составляет более 13 млн записей.

Особую нишу в информационном пространстве занимают мультидисциплинарные коммерческие базы данных цитирования *Web of Sciences* (<https://webofknowledge.com>) и *Scopus* (<http://www.scopus.com>), включающие миллионы записей из научных журналов и сборников и предоставляющие пользователям достаточно мощный интерфейс, поддерживающий поиск и навигацию как по метаданным собственно статей, так и по элементам пристатейной библиографии. Это позволяет использовать указанные базы данных не только в качестве источников собственно научной информации, но и в качестве инструментов для проведения наукометрических исследований.

В мировом информационном пространстве поддерживается значительное число документальных и фактографических информационно-поисковых систем, относящихся к конкретным научным направлениям.

*Agris International Information System for the Agricultural Sciences and Technology* (<http://agris.fao.org>) – Международная информационная система по сельскохозяйственным наукам и технологиям. *GeoRef* (<http://www.agiweb.org/>) – библиографическая база данных, охватывающая мировую литературу по наукам о Земле. *Chemical Abstracts Service* (<http://www.cas.org/>) – химическая реферативная служба, издающая реферативный журнал "Chemical Abstracts" и поддерживающая крупнейшую химическую информационно-поисковую систему SciFinder. Этот перечень можно продолжать, но приведённые примеры уже дают представление о составе и функциях международных информационных научных ресурсов.

За рубежом достаточно широкое распространение получили интеграторы научных информационных ресурсов – организации, приобретающие у владельцев права на их распространение. Такие организации разрабатывают единый по-

исковый интерфейс для доступа к различным ресурсам и предоставляют коммерческий доступ к ним через одну точку входа. Среди основных игроков на этом поле можно назвать компании STN International [7] (она обеспечивает доступ к почти 200 базам данных по научно-технической информации и к более 100 коллекциям научных журналов; переведённый на русский язык список доступных через неё ресурсов представлен в [8]) и EBSCO [9], агрегирующую коллекции полнотекстовых электронных журналов и книг.

Отечественное информационное пространство не изобилует мощными документальными и фактографическими информационными системами и в основном представлено узкотематическими ресурсами, которые существуют благодаря энтузиастам. Оценить общее количество научно-информационных ресурсов, поддерживаемых отечественными организациями, достаточно сложно. Если академические информационные системы в сфере общественных наук достаточно полно представлены в Навигаторе информационных ресурсов по общественным наукам (<http://niron.inion.ru>), то по другим научным направлениям такой учёт отсутствует. В международном регистре баз данных по неорганической химии [10] перечислены 22 базы данных, поддерживаемые в России. Но реально доступны через Интернет только 10 из них: пять принадлежат Институту металлургии им. А.А. Байкова РАН, две – МГУ им. М.В. Ломоносова, две – Объединённому институту высоких температур РАН, одна – Институту экспериментальной минералогии СО РАН. Из общедоступных документальных справочных систем по естественным наукам можно отметить раздел сайта библиотеки по естественным наукам РАН (БЕН РАН) "Естественные науки в Интернете" [11], который содержит ссылки на сайты, осуществляющие интеграцию ресурсов по основным областям естественных наук, а также раздел сайта Центральной библиотеки (отдела БЕН РАН) в Пушкинском научном центре [12], отражающий различные ресурсы по физико-химической биологии, информационную систему по геологии и горному делу [13].

В качестве универсального справочного материала, представленного в Интернете, наиболее популярны и за рубежом, и в России такие ресурсы, как Wikipedia и Wikidata. Они, без сомнения, играют значительную роль в жизни современного информационного общества, но в качестве источника серьёзных научных знаний выступать не могут, поскольку содержат значительное количество неподтверждённых и недостоверных сведений.

Приведённые примеры показывают, что хотя каждый из существующих в мировом информационном пространстве ресурсов играет определённую роль, элементами единого цифрового

пространства научных знаний они выступать не могут, поскольку: во-первых, в подавляющем большинстве случаев ресурсы (за исключением узкотематических фактографических систем) ориентированы не на выдачу конкретных научных знаний, а на предоставление информации о публикациях, где эти знания отражены (в лучшем случае можно получить полный текст публикаций); во-вторых, наполнение политематических ресурсов (Wikipedia, Wikidata, Europeana) является случайным, их контент зачастую содержит неточную или неверную информацию; в-третьих, в большинстве существующих систем отсутствует развитый поисковый аппарат, он не снабжён серьёзными лингвистическими средствами (проблемно-ориентированные фактографические системы имеют собственные специфические лингвистические средства, однако ни о каком универсальном подходе к их формированию речи идти не может); в-четвёртых, ни в одной системе не присутствует многоуровневость представления контента, предусматривающая разделение контента на базовые теоретические знания, экспериментальные результаты, научно-популярные и учебные материалы.

Построение ЕЦПНЗ требует разработки специальных подходов и алгоритмов, базирующихся на новых принципах. Существующие информационные ресурсы должны стать источниками наполнения ЕЦПНЗ, но не могут являться его составляющими.

**Структура ЕЦПНЗ.** Как уже было сказано, пространство научных знаний должно включать две составляющие — постоянную (статическую) и динамическую. Постоянная составляющая — это базовые, проверенные временем и практикой теоретические и экспериментальные знания; динамическая составляющая охватывает появляющиеся новые знания. Эти составляющие можно рассматривать как две подсистемы, одна из которых (базис) отражает зафиксированные научные знания, другая (надстройка) — новую научную информацию, которая по истечении некоторого времени, после прохождения экспертного фильтра вливается в первую.

Базис и надстройка ЕЦПНЗ включают совокупность подпространств (фасетов) по различным научным направлениям, связи между которыми могут быть реализованы на уровне лингвистических средств (междисциплинарной общенаучной онтологии). Базис включает три взаимосвязанных уровня — начальный (например, рассчитанный на школьников), основной (для широкого круга пользователей, не специализирующихся в данной области), профессиональный (предназначенный для специалистов в данной области науки). Надстройка ориентирована на специалистов в данной

научной области и является источником текущей информации, необходимой для проведения научных исследований.

Начальный уровень — статический, он формируется путём извлечения информации из учебников, научно-популярной литературы; в нём могут быть широко представлены оцифрованные музейные экспонаты, мультимедийные популярные ресурсы.

Основной уровень базиса — квазистатический, он изменяется крайне редко при появлении существенных, принципиальных для данной области знания результатов (открытий) и формируется с использованием существующих энциклопедий, справочников, авторитетных монографий, проверенных временем учебников для высшей школы. Этот уровень включает, наряду с извлечёнными метаданными, ссылки на источники и оформленные в соответствии с законодательством об охране авторских прав их цифровые копии.

Профессиональный уровень базиса включает более глубокую по сравнению с основным уровнем статическую информацию по данному научному направлению, дополняемую подтверждённой экспертными оценками новой информацией.

Надстройка отражает новые результаты в той или иной научной области. Её контент формируется за счёт извлечения знаний из текущей опубликованной информации, прошедшей экспертизу (это могут быть статьи из престижных журналов, сборники научных статей с подтверждённым научным уровнем, патенты, материалы ведущих мировых научных конференций и пр.). Наполнение надстройки динамически меняется как благодаря новой информации, так и вследствие исключения из неё сведений, ценность которых не подтвердилась временем.

**Контент ЕЦПНЗ.** Источниками всех научных знаний являются информационные ресурсы — описания постулатов, теорий, экспериментов, полученных на их основе результатов в той или иной области науки, представленные на материальном носителе. В древности это были глиняные таблички, папирусы, потом появились печатные материалы, фото-, кино-, видеоматериалы, музейные модели, а в последнее время — цифровые носители, обладающие качественно иными возможностями с точки зрения компактного хранения, извлечения, поиска и предоставления знаний пользователям.

Базис ЕЦПНЗ должен формироваться в основном из имеющихся в России ресурсов — библиотечных, архивных, музейных, энциклопедических, фактографических, словарных и других, создаваемых научно-образовательным сообществом на основе конвергенции этих ресурсов, то есть базис выращивается из существующих систем

без нарушения их функционирования. В большинстве случаев информация, содержащаяся в этих ресурсах, является достоверной и проверенной временем, однако для решения вопроса о конкретном выборе того, что будет загружено в ЕЦПНЗ, необходима специальная экспертиза, которая может быть осуществлена только квалифицированными представителями научного сообщества в области, соответствующей данному подпространству.

Что касается надстройки, то эта динамическая часть ЕЦПНЗ формируется информационно-библиотечными специалистами (с привлечением учёных) на основе существующих мировых баз данных научной информации и коллекций полнотекстовых цифровых материалов. Некоторым прообразом технологии формирования этой части ЕЦПНЗ может стать подход, разработанный в БЕН РАН по созданию проблемно-ориентированных информационных систем [14].

Как базис, так и надстройка каждого подпространства ЕЦПНЗ состоит из ядра и периферии. *Элементами ядра* являются информационные объекты, содержащие метаданные, извлечённые из перечисленных выше ресурсов, которые определяют основные постулаты, методы и результаты исследований в конкретной области науки для базиса и надстройки. Каждый информационный объект должен, во-первых, содержать сведения типа "Что (кто), где, когда" и указывать на связи с другими объектами как данного, так и, возможно, других подпространств; во-вторых, иметь ссылку на источник информации. *Периферию ЕЦПНЗ* образуют источники информации, отобранные по критериям качества, — оцифрованные публикации, архивные материалы, образы музейных экспонатов, мультимедийные материалы, тематические базы данных, поддерживаемые в научных организациях.

**Онтология ЕЦПНЗ.** Единство пространства научных знаний обеспечивается поддержкой унифицированного представления связей между объектами, единых принципов описания информационных объектов, единых терминологических словарей и тезаурусов, единых принципов оцифровки источников информации. Тем самым в ЕЦПНЗ формируется онтология, включающая единые словари (тезаурусы) общенаучных терминов, а также словари номенклатурных наименований, географических названий, имён персоналий, событий, обозначения временных интервалов, тематики и другие категории научной лексики. Внутри каждого подпространства создаются онтологии данного научного направления, включающие ключевые слова и корреляции между ними, а также средства, благодаря которым обеспечивается взаимодействие различных систем классифи-

кации научных знаний в той или иной области. Понятия онтологий связаны соответствующими элементами обобщённых словарей ЕЦПНЗ.

**Операции над объектами ЕЦПНЗ.** В ЕЦПНЗ определены следующие операции над информационными объектами:

- поиск объектов по любым элементам метаданных и их комбинациям в отдельных подпространствах как в цифровом пространстве в целом, так и внутри его базиса или надстройки, с выдачей списков найденных объектов, отсортированных по заданному элементу; для числовых значений метаданных определены операции "равно", "не равно", "меньше", "больше"; для текстовых объектов — "равно" (точное совпадение элемента метаданных с заданным фрагментом текста с возможностью выбора значения из перечислимого множества), "начинается с...", "содержит", "не содержит", правое и левое усечение от начала слова; для данных типа "дата" — "равно", "до заданной даты", "после заданной даты". Поисковый запрос может содержать множество элементов, связанных операторами "и", "или", "и не";
- навигация (переход) от найденных объектов к связанным с ними как внутри подпространства, так и в пространстве в целом (с использованием возможностей общей онтологии ЕЦПНЗ);
- формирование различных коллекций объектов, отвечающих заданному свойству.

Следует отметить, что развитие ЕЦПНЗ как с точки зрения его наполнения, расширения номенклатуры входящих в него объектов, так и появления новых запросов потребует введения новых операций на множестве объектов ЕЦПНЗ.

**Наполнение и поддержка ЕЦПНЗ.** Единое пространство строится на принципах, сочетающих распределённую (децентрализованную) подготовку метаданных по единым правилам с централизованным их хранением при единой централизованной системе контроля качества. Материалы, относящиеся к периферии частей ЕЦПНЗ (оцифрованные публикации, архивные документы, музейные предметы и т. п.), могут храниться как у их владельцев, так и в центральном блоке пространства.

Технологически и технически ЕЦПНЗ включает три уровня. На верхнем уровне располагается диспетчерский блок, связанный со всеми центральными фасетными блоками, представляющими собой второй уровень пространства. На третьем уровне находятся блоки участников поддержки периферийной информации подпространств ЕРПНЗ, каждое из которых связано с соответствующим центральным подпространством второго уровня. Центральные блоки обеспечивают сбор и поддержку ядра каждого подпространства.

**Организационная структура разработки и поддержки ЕЦПНЗ.** Формирование ЕЦПНЗ как общенационального проекта должно координироваться специальным межведомственным советом, включающим ведущих учёных и определяющим основные направления разработки проекта, принципы участия в нём и необходимое финансирование; совет может быть создан под эгидой президиума РАН. Для проектирования ЕЦПНЗ создаётся проектный офис на базе одного из ведущих центров в области информатики (например, ФИЦ "Информатика и управление" РАН или ФНЦ НИИСИ РАН), обеспечивающий определение всех основных компонентов ЕЦПНЗ с привлечением специалистов из разных организаций. Первым этапом деятельности проектного офиса должна стать подготовка эскизного проекта ЕЦПНЗ, в рамках которого необходимо уточнить принципы его разработки и поддержки, правила взаимоотношений с существующими и создаваемыми информационными ресурсами (в том числе с Национальной электронной библиотекой), участниками создания ЕЦПНЗ, а также оценить необходимые финансовые затраты, выбрать стандарты представления информации.

Эскизный проект должен быть рассмотрен и утверждён Советом ЕЦПНЗ, после чего силами проектного офиса нужно разработать техническое задание на создание пилотной версии ЕЦПНЗ, включающей несколько подпространств по ряду естественных и гуманитарных научных направлений, а также на прототип единой онтологии ЕЦПНЗ. Это техническое задание также рассматривается и утверждается Советом ЕЦПНЗ, после чего осуществляются подготовка технорабочего проекта пилотной версии ЕЦПНЗ и её практическая реализация.

#### ЛИТЕРАТУРА

1. *Вершинин А. П.* Единое российское электронное пространство знаний: вопросы права // Информационно-аналитический журнал "Университетская книга". 2016. Декабрь. <http://www.unkniga.ru/biblioteki/bibdelo/6630-edinoe-rossiyskoe-elektronnoe-prostranstvo-znaniy-voprosy-prava.html>
2. *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.
3. *Антопольский А. Б., Ефременко Д. В.* Инфосфера общественных наук России / Под ред. В. А. Цветковой. М.; Берлин: Директ-Медиа, 2017.
4. *Калёнов Н. Е., Савин Г. И., Серебряков В. А., Сотников А. Н.* Принципы построения и формирования электронной библиотеки "Научное наследие России" // Программные продукты, системы и алгоритмы. 2012. Т. 4. № 100. С. 30–40. <http://www.swsysweb.ru> (дата обращения 14.11.2018).
5. *Калёнов Н. Е., Соболевская И. Н., Сотников А. Н.* Цифровые музейные коллекции и представление объектов естественно-научного музейного хранения в электронной библиотеке "Научное наследие России" // Научно-техническая информация. Сер. 1. 2016. № 10. С. 33–38.
6. *Калёнов Н. Е., Сотников А. Н., Ильина И. Н.* Архивная информация в электронной библиотеке "Научное наследие России" // Фундаментальная наука: проблемы изучения, сохранения и реставрации документального наследия: Материалы Международной научной конференции / Отв. ред. В. Ю. Афиани. М.: Архив РАН, 2013. С. 25–35.
7. Компания STN International. [http://www.stninternational.de/stn\\_content.html](http://www.stninternational.de/stn_content.html)
8. Русскоязычный список ресурсов STN International [http://www.nioch.nsc.ru/sibstn/onlin\\_db.htm](http://www.nioch.nsc.ru/sibstn/onlin_db.htm)
9. Компания EBSCO. <https://www.ebsco.com/>
10. Международный регистр баз данных по неорганической химии. <http://iric.imet-db.ru/DB.asp?ido=0&idc=1&nameDB=&ABR=&pgn=1>
11. Естественные науки в Интернет (раздел сайта БЕН РАН). [http://www.benran.ru/st\\_point.html](http://www.benran.ru/st_point.html)
12. Ресурсы в области физико-химической биологии на сайте библиотеки Пушкинского научного центра. <http://cbp.iteb.psn.ru/library/>
13. Информационная система по геологии и горному делу. <http://scirus.benran.ru/higeo/>
14. *Ивановский А. А., Калёнов Н. Е.* Технология формирования и поддержки проблемно-ориентированных информационных систем // Информация и инновации. 2018. № 4. С. 19–25.

## COMMON DIGITAL SPACE OF SCIENTIFIC KNOWLEDGE

© 2019 А.В. Antopolskii<sup>1\*</sup>, N.E. Kalenov<sup>2\*\*</sup>, V.A. Serebryakov<sup>3\*\*\*</sup>, A.N. Sotnikov<sup>4\*\*\*\*</sup><sup>1</sup>*Institute of Scientific Information on Social Sciences of the Russian Academy of Sciences, Moscow, Russia*<sup>2</sup>*Library for Natural Sciences of the Russian Academy of Sciences, Moscow, Russia*<sup>3</sup>*Dorodnicyn Computing Centre of the Federal Research Centre "Informatics and Control"  
of the Russian Academy of Sciences, Moscow, Russia*<sup>4</sup>*Joint Supercomputer Centre of the Russian Academy of Sciences – Branch of Federal State Institution "Scientific Research Institute  
for System Analysis of the Russian Academy of Sciences", Moscow, Russia*

\*E-mail: ale5695@yandex.ru, \*\*E-mail: nek@benran.ru,

\*\*\*E-mail: serebr@ultimeta.ru, \*\*\*\*E-mail: asotnikov@jscs.ru

Received 19.11.2018

Revised version received 21.11.2018

Accepted 26.11.2018

The article attempts a comprehensive analysis of the concept and phenomenon of the common digital space of scientific knowledge (CDSSK) as a narrower wording of the concept of the common Russian electronic space of knowledge, which is proposed in a number of government documents. Definitions of basic concepts are proposed in relation to this phenomenon. Goals, objectives, composition, and structure are considered to be the boundaries and the principles of creating CDSSK. A brief overview of the existing scientific information resources and services that have the features of the elements of space of scientific knowledge follows. It is concluded that the existing information resources can be sources of CDSSK but not its components. The paper proposes a structure for CDSSK, including static and dynamic components (base and add-in), various levels of knowledge representation, and core and periphery. This should include formalized knowledge over which a set of operations and the periphery – the sources of this knowledge (including documents) – are selected on the basis of quality. The logical structure of CDSSK must submit its ontology, formed as it ought to be on the basis of information languages and metadata of existing scientific information resources. A number of steps are proposed to organize the work of creating CDSSK, as well as its content and support.

*Keywords:* information resources, digital space, scientific electronic libraries, service and search systems, ontology of scientific knowledge.