

РОССИЙСКАЯ АКАДЕМИЯ НАУК

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ИНСТИТУТ НАУЧНОЙ ИНФОРМАЦИИ
ПО ОБЩЕСТВЕННЫМ НАУКАМ**

А.Б. Антопольский

**НАУЧНАЯ ИНФОРМАЦИЯ
И ЭЛЕКТРОННОЕ ПРОСТРАНСТВО
ЗНАНИЙ**

Под научной редакцией Д.В. Ефременко

**МОСКВА
2020**

УДК [002:004]:316.774

ББК 32.97

А 72

Рецензенты:

Лопатина Н.Н., д-р пед. наук, профессор, заведующая кафедрой библиотечно-информационной науки Московского государственного института культуры

Авдонин В.С., д-р полит. наук, ведущий научный сотрудник ИНИОН РАН

Антопольский, Александр Борисович.

А 72 Научная информация и электронное пространство знаний : монография / А.Б. Антопольский ; ИНИОН РАН, Фундам. б-ка ; науч. ред. Д.В. Ефременко. – Москва : ИНИОН РАН, 2020. – 252 с.

ISBN 978-5-248-00964-0

Монография содержит описание тенденций развития информационного пространства науки и видения будущего научных коммуникаций со стороны мирового научного сообщества. Рассматриваются состояние системы научной информации в России и основные решения органов государственной власти, связанные с ее развитием. Описываются информационные ресурсы академических институтов по социальным и гуманитарным наукам. В рамках предлагаемой концепции построения единого электронного (цифрового) пространства научных знаний особое внимание уделяется мониторингу и селекции научных информационных ресурсов, их интеграции, агрегации и формированию на основе качественных ресурсов ядра пространства знаний, включающего онтологию и базу знаний. Разработаны модель формирования пространства знаний для сферы языкознания и методика оценки качества лингвистических ресурсов.

Для специалистов по проблемам научной информации и цифровизации науки, научных библиотек, преподавателей вузов, студентов и аспирантов.

УДК [002:004]:316.774

ББК 32.97

DOI: 10.31249/spaknow/2020.00.00

ISBN 978-5-248-00964-0

© ФГБУН «Институт научной информации
по общественным наукам РАН»

СОДЕРЖАНИЕ

Введение.....	9
Глава 1. Тенденции развития и образ будущего научных коммуникаций	13
1.1. Тенденции развития научных коммуникаций.....	13
<i>Цифровизация</i>	<i>13</i>
<i>Открытая наука</i>	<i>14</i>
<i>Система оценки научных результатов</i>	<i>15</i>
<i>Коллаборации</i>	<i>16</i>
<i>Будущее научных библиотек и информационных центров</i>	<i>17</i>
<i>Тематические информационные центры</i>	<i>19</i>
<i>Публикации и данные</i>	<i>20</i>
<i>Перспективы развития информационных языков</i>	<i>21</i>
<i>Экономика научных коммуникаций</i>	<i>21</i>
<i>Модификация законодательства</i>	<i>23</i>
<i>Единое российское электронное пространство знаний (ЕРЭПЗ).....</i>	<i>23</i>
1.2. Инициативы мировой научной общественности по развитию научных коммуникаций в условиях открытой науки	24
<i>Исходные документы по открытой науке</i>	<i>24</i>
<i>Открытый доступ и оценка научных исследований. Декларация DORA</i>	<i>25</i>
<i>Лейденский манифест</i>	<i>28</i>
<i>Гаагская декларация по открытию знаний в цифровую эпоху</i>	<i>32</i>
<i>Конкордат по открытым научным данным.....</i>	<i>32</i>
<i>Амстердамский призыв</i>	<i>34</i>
<i>Призыв Жюсье</i>	<i>35</i>

<i>План S и рекомендации по его реализации</i>	35
<i>Будущее научных коммуникаций и изданий</i>	38
1.3. Выводы и рекомендации для России	44
<i>Общие выводы</i>	44
<i>Выводы в отношении категорий участников</i>	44
<i>Выводы относительно оценки научных результатов</i>	45
Глава 2. Научно-техническая информация	
в России сегодня	46
2.1. Введение	46
2.2. Краткий обзор мирового опыта	47
2.3. Общая характеристика отечественных ресурсов НТИ.....	51
2.4. Федеральные органы НТИ	53
2.5. Исследование отечественных ресурсов по управлению научкой	59
2.6. Инвентаризация владельцев ресурсов НТИ	60
2.7. Инвентаризация ресурсов	64
2.8. Выводы.....	67
Глава 3. Мониторинг информационных ресурсов	
социально-гуманитарных наук	68
3.1. Цели и задачи мониторинга информационных ресурсов.....	68
3.2. Структура научной инфосферы	69
3.3. Определение границ сферы мониторинга	73
3.4. Тематические классификации социальных и гуманитарных наук.....	73
3.5. Тематические границы мониторинга	77
3.6. Функциональная классификация информационных ресурсов	78
3.7. Методические вопросы инвентаризации информационных ресурсов	84
3.8. Сбор информации в ходе мониторинга	87
3.9. Результаты мониторинга и состав информационных ресурсов	89
<i>Архивные фонды</i>	89
<i>Фонды традиционных библиотек</i>	90
<i>Музейные фонды</i>	91
<i>Периодические, продолжающиеся и сериальные издания (ППСИ)</i>	92
<i>Вторичные ресурсы</i>	92
<i>Первичные документальные ресурсы</i>	93

<i>Информационные системы</i>	94
<i>Информация о научных событиях</i>	95
<i>Персональные ресурсы</i>	95
<i>Медиаресурсы</i>	95
<i>Прочие ресурсы Интернета</i>	96
3.10. Навигатор информационных ресурсов по общественным наукам.....	96
<i>Логическая структура НИРОН</i>	97
<i>Перспективы развития НИРОН</i>	104
3.11. О путях оптимизации академических социогуманитарных информационных ресурсов.....	105
3.12. Основные выводы по результатам мониторинга	117
Глава 4. Государственные решения в области научной информации (2018–2019)	120
4.1. О государственных решениях.....	120
4.2. Поручения Президента РФ	121
4.2. Законопроект «О научной и научно-технической деятельности в Российской Федерации»	121
4.3. Цифровые платформы Минобрнауки.....	127
4.4. Национальная электронная библиотека.....	132
4.5. Национальный энциклопедический портал.....	133
4.6. Другие правительственные действия в сфере научной информации	134
Глава 5. Единое российское электронное пространство знаний – постановка задачи	137
5.1. Единое российское электронное пространство знаний в «Положении о Национальной электронной библиотеке» ...	137
5.2. Некоторые замечания о понятии ЕРЭПЗ	139
5.3. Общие принципы создания ЕРЭПЗ	142
5.4. Цели и задачи ЕРЭПЗ	144
5.5. Этапы формирования ЕРЭПЗ.....	145
<i>Первый этап – мониторинг инфосферы</i>	145
<i>Второй этап – оценка и отбор научных ресурсов</i>	146
<i>Третий этап – формирование ядра ЕРЭПЗ</i>	149
<i>Четвертый этап – выявление новых знаний</i>	151
5.6. Функциональный состав участников ЕРЭПЗ.....	151
<i>Головная организация</i>	152
<i>Ведущие организации по отраслям науки</i>	152

<i>Операторы основных научно-образовательных электронных ресурсов</i>	153
<i>Службы формирования ядра ЕРЭПЗ и выявления нового знания</i>	154
5.7. Технологическая платформа ЕРЭПЗ.....	154
5.8. Организационные принципы функционирования ЕРЭПЗ	154
5.9. Экономическая модель ЕРЭПЗ.....	156
5.10. Правовые основы ЕРЭПЗ	156
5.11. Международное сотрудничество при создании и развитии ЕРЭПЗ.....	157
5.12. ГСНТИ и ЕРЭПЗ	158
Глава 6. Интеграция лингвистических информационных ресурсов.....	160
6.1. Введение	160
6.2. Анализ зарубежного опыта	161
<i>Общая характеристика</i>	161
<i>ELRA – Европейская ассоциация языковых ресурсов</i>	162
<i>OLAC – Сообщество открытых лингвистических архивов..</i>	164
<i>Стандарты</i>	165
<i>Рекомендации</i>	166
<i>Методики и проекты</i>	166
<i>TELRI – Трансъевропейская инфраструктура лингвистических ресурсов</i>	166
<i>CLARIN – Общие языковые ресурсы и технологическая инфраструктура.....</i>	167
<i>LDC – Консорциум лингвистических данных</i>	172
<i>ISCA – Международная ассоциация речевого общения</i>	173
<i>SIL – Летний институт лингвистики</i>	174
<i>Национальные центры ЛИП</i>	174
<i>ORTOLANG – Открытые ресурсы и инструменты для обработки языка</i>	175
<i>Терминологические веб-сайты и банки данных</i>	177
<i>Каталоги ЛИП.....</i>	178
6.3. Концепция создания Центра лингвистических ресурсов РАН	178
<i>Постановка задачи</i>	178
<i>Определения.....</i>	179
<i>Состояние создания ЛИП в РАН.....</i>	179
<i>Типология лингвистических ресурсов.....</i>	181

<i>Активные формы – алгоритмы, модели, программы, базы знаний</i>	181
<i>Задачи Центра лингвистических ресурсов</i>	181
<i>Этапность создания и развития Центра лингвистических ресурсов</i>	182
<i>Целесообразность интеграции и агрегации ЛИР по отдельным типам</i>	183
<i>Корпуса</i>	184
<i>Организационные аспекты создания Центра лингвистических ресурсов</i>	186
<i>Технологические аспекты функционирования Центра лингвистических ресурсов</i>	187
<i>Экономические аспекты создания и функционирования Центра лингвистических ресурсов</i>	187
<i>Правовые аспекты создания и функционирования центра</i> ...	188
6.4. Навигатор информационных ресурсов	
<i>по языкознанию (НИРЯЗ)</i>	189
<i>Структура и функциональность НИРЯЗ</i>	189
<i>Содержание информации, включенной в НИРЯЗ</i>	189
<i>Рубрикатор НИРЯЗ</i>	191
<i>Языковой фасет НИРЯЗ</i>	192
6.5. Методика оценки качества лингвистических электронных ресурсов	193
<i>Методы оценки ресурсов</i>	193
<i>Оценка качества сайта</i>	194
<i>Параметры оценки сайтов</i>	194
<i>Характеристики оценки качества сайта</i>	196
<i>Методика оценки качества сайта</i>	201
<i>Примеры оценки сайтов</i>	204
6.6. Онтология поисковых терминов по лингвистике	207
<i>Участие в создании онтологии по комплексному гранту</i>	207
<i>Онтология поисковых терминов по лингвистике (ОПТЕЛ)</i> ..	208
Заключение	210
Литература и электронные ресурсы	213
Приложения	223
Приложение 1. Общественные науки в ГРНТИ	223
Приложение 2. Фрагмент номенклатуры ВАК по общественным наукам	223
Приложение 3. Рубрикатор РГНФ	224
Приложение 4. Рубрикатор РНФ	225

Приложение 5. Рубрикатор FOS (ОЭСР) по социальным и гуманитарным наукам	225
Приложение 6. Рубрикатор WOS по социальным наукам	226
Приложение 7. Рубрикатор WOS по искусству и гуманитарным наукам	228
Приложение 8. Фрагмент рубрикатора Scopus	228
Приложение 9. Функциональная классификация типов ресурсов по результатам мониторинга	230
Приложение 10. Упрощенная функциональная классификация, используемая в Навигаторе НИРОН	235
Приложение 11. Карта LRE [электронная карта лингвистических ресурсов].....	236
Приложение 12. Участники CLARIN.....	239
Приложение 13. Индивидуальные веб-сервисы членов CLARIN.....	240
Приложение 14. Топ-20 терминологических служб и банков данных.....	240
Приложение 15. Мировые и зарубежные каталоги ЛИР.....	242
Приложение 16. Российские каталоги ЛИР.....	243
Приложение 17. Рубрикатор по языкознанию НИРЯЗ.....	244
Приложение 18. Таблица перекодировки кодов ГРНТИ в Рубрикатор НИР.....	245
Приложение 19. Указатель языков и языковых групп в НИРЯЗ.....	246
Приложение 20. Техническое задание на формирование. Онтологии поисковых терминов по лингвистике.....	249

Введение

Исследования и разработки последних 20–25 лет в области научной коммуникации и научной информации свидетельствуют об отсутствии у российского научно-информационного сообщества общего видения будущего научной коммуникации, которое определяется тенденциями ее развития.

Под научной коммуникацией здесь понимаются все формы информационного взаимодействия в научной сфере – как традиционные (издательская, библиотечная, научно-информационная, архивная), так и современные, реализуемые в цифровой среде, прежде всего в Интернете (сайты, порталы, электронные библиотеки, базы и банки данных, репозитории, социальные сети и др.).

В мировой практике регулярно делаются попытки сформулировать коллективное мнение по образу будущего научных коммуникаций. Последним примером может служить доклад группы европейских экспертов «Будущее научных изданий и научных коммуникаций» [127], опубликованный в январе 2019 г. и подробно рассмотренный в первой главе книги.

Подчеркнем, что авторы этого документа, как и многих других коллективных концептуальных документов в области научных коммуникаций, в центр проблемы оптимизации научных коммуникаций ставят задачу адекватной оценки научных результатов. В частности, признается, что существующие методы оценки на основе импакт-факторов журналов и вообще библиометрические методы неудовлетворительны. Но изменение этих методов требует существенных социальных инноваций в области организации науки.

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

- Вся ли научная коммуникация через 10–15–20 лет будет цифровой?

- Является ли переход к открытой науке безусловным?

- Как будет развиваться система оценки научных исследований?
- Какую роль в научной коммуникации должны играть научные и вузовские библиотеки? Каково будущее научных библиотек?
- Нужны ли в будущем специализированные центры научной информации? Какими будут их функции? Есть ли будущее у библиографических и реферативных служб?
- Как будут модернизироваться научные издательства в условиях открытой науки и в единой цифровой среде?
- Электронные научные журналы и открытые репозитории: каково их соотношение в обозримом будущем?
- Какова перспектива существующих российских ресурсов научной информации?
- Как будет меняться соотношение централизованных (универсальных) и отраслевых (тематических) ресурсов и систем научной информации? Тот же вопрос относится к международным и национальным ресурсам и системам научной информации.
- Есть ли реальная перспектива у Единого российского электронного пространства знаний (ЕРЭПЗ)? Имеет ли смысл говорить о воссоздании Государственной системы научно-технической информации (ГСНТИ)?
- Публикации и данные: будет ли меняться их соотношение?
- Как будут развиваться информационные языки, онтологии, поисковые машины?
- Какая экономическая модель будет определять научную коммуникацию?

Удастся ли модернизировать систему копирайта?

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

В то же время анализ государственных решений в области научной информации показывает, что у органов государственной власти и руководства отечественной наукой отсутствует сколько-нибудь четкий план развития научной инфосферы, хотя проблемам цифровизации науки, как и экономики в целом, в правительственных документах уделяется большое внимание.

Особый интерес вызывает сформулированная в ряде нормативных документов идея создания Единого российского электронного пространства знаний. Если данная идея действительно может быть реализована, то представляется очевидным, что планы цифровизации науки и образования, как и развития научной инфраструк-

туры в целом, должны быть полностью ориентированы на достижение этой цели. Однако в реальности это не так. Хотя, по мнению многих авторитетных специалистов, эта идея не имеет шансов на реализацию при современном состоянии научно-информационного сообщества, автор считает, что обсуждение путей создания единого пространства знаний является хорошим способом организации дискуссии по сформулированным выше проблемам.

В настоящей книге делается попытка описать ситуацию в контексте поставленных вопросов и сформулировать свою позицию по вопросу создания единого электронного пространства знаний. В основном автор опирается на исследования научных коммуникаций в сфере социальных и гуманитарных наук. В частности, предлагается модель создания пространства знаний на примере лингвистики, лидирующей в России среди гуманитарных наук в области создания информационных ресурсов и систем (цифровой гуманитаристики). В то же время в российской компьютерной лингвистике особенно очевидно отсутствие координации и общего плана развития в этой сфере по сравнению с ситуацией за рубежом.

Монография имеет следующую структуру.

В первой главе рассмотрены основные тенденции развития научных коммуникаций и информационных ресурсов, отражение этих тенденций в коллективных декларациях и докладах зарубежных ученых.

Во второй главе рассмотрено общее состояние научно-технической информации в России. Описываются органы и организации, владеющие информационными ресурсами, отражающими результаты научных исследований, а также ресурсы, предназначенные для управления наукой и науковедческих исследований. Приводятся результаты каталогизации ресурсов по науковедению и управлению наукой.

В третьей главе описывается состояние информационных ресурсов академических учреждений России по социальным и гуманитарным наукам, в основном на базе мониторинга, проведенного автором в ИНИОН РАН в 2017–2018 гг.

Четвертая глава содержит обзор и анализ правительственных решений последних лет, связанных с развитием и регулированием научной инфосферы. Констатируется несогласованность этих решений.

Пятая глава посвящена центральной проблеме развития научного информационного пространства в обозримом будущем. Обсуждается возникшая в последние годы и зафиксированная в

ряде официальных документов идея создания Единого российского электронного пространства знаний.

В заключительной шестой главе представлены результаты исследования возможной модели интеграции научных информационных ресурсов на примере языкознания, выполненного в рамках проекта, поддержанного Российским фондом фундаментальных исследований.

Глава 1.

ТЕНДЕНЦИИ РАЗВИТИЯ И ОБРАЗ БУДУЩЕГО НАУЧНЫХ КОММУНИКАЦИЙ

1.1. Тенденции развития научных коммуникаций

Цифровизация

Наиболее очевидной тенденцией изменения системы научных коммуникаций является цифровизация, т.е. переход многих видов научных коммуникаций в цифровую форму.

В начале XIX в. многим специалистам казалось, что развитие информационных технологий приведет к быстрому и полному переходу научных коммуникаций в цифровую форму. Можно вспомнить обещание руководства Google, сделанное в 2005 г., оцифровать к 2020 г. *все* книги, изданные в мире. По оценке той же компании, их в мире около 150 млн. Однако к настоящему времени оцифровано не более 25 млн, а доля потребления электронных книг составляет в России 5–6% [52].

В других формах научной коммуникации – периодике, библиотечных фондах, архивах – цифровизация также не стала полной. Более подробные данные о динамике цифровизации научной коммуникации в России представлены в монографии [22].

На основе этих данных можно сделать очевидный вывод: в обозримом будущем печатная книга и журнал не исчезнут, хотя их доля будет постоянно сокращаться. Но динамика сокращения востребованности печатных форм коммуникации вряд ли превысит 5–10% в год. Некоторые виды научных документов, впрочем, уже полностью переведены в электронную форму (диссертации, отчеты). В целом рост объема цифрового контента в Интернете составил в 2017 г. 12% [113].

Таким образом, в обозримом будущем цифровая и традиционная формы научных коммуникаций будут сосуществовать, с постепенным доминированием цифровой формы.

В то же время эта очевидная тенденция совершенно недостаточно учитывается в программах и концептуальных документах, определяющих научную информационную политику, а самое главное – в представлении акторов информационной отрасли. Это напрямую затрагивает издательскую отрасль – и в части научной периодики, и в части книжной политики. Но главным образом эта тенденция меняет информационно-библиотечную и архивную сферы. Однако у научной общественности России нет общего мнения о формах и способах организации этих отраслей в новых условиях цифровой эпохи.

Открытая наука

В последние годы мировая научная общественность, особенно в Европе и в США, активно обсуждает переход к открытой науке, и в связи с этим – будущее открытых научных коммуникаций в цифровой среде. В результате появился ряд деклараций, экспертных докладов и других коллективных документов на эту тему. Особый резонанс получило решение Европейской комиссии об открытии с 2020 г. всех публикаций по результатам исследований, финансируемых европейскими правительствами и фондами [134].

Для реализации этого решения в сентябре 2018 г. были приняты так называемый план S, разработанный коалицией научных организаций 13 стран Евросоюза (коалицией S) [82], и рекомендации по его применению [128].

По вопросу о присоединении России к плану S имеются разные мнения. С одной стороны, высказывается полная поддержка этого плана [56; 59], к которой присоединяется и автор этой книги. С другой стороны, на заседании Президиума РАН в марте 2019 г., когда представитель Евросоюза призвал Россию присоединиться к плану S, вице-президент РАН А.Р. Хохлов не поддержал это предложение. Действительно, в планах руководящих органов науки России нет программы перехода к открытому доступу, хотя имеется поручение Президента РФ подготовить доклад о степени открытости научных организаций [81]. Таким образом, вопрос о динамике движения российских научных коммуникаций к открытому доступу остается нерешенным.

Система оценки научных результатов

Авторы основных коллективных документов по развитию научных коммуникаций приходят к выводу, что основная проблема их развития связана с принятыми сегодня в управлении наукой методами оценки научных результатов.

Главный вывод декларации DORA гласит:

«Не используйте такие количественные журнальные показатели, как импакт-фактор, в качестве показателя качества отдельных научных статей при оценке вклада конкретных ученых, при приеме на работу, продвижении по службе или при принятии решений о финансировании» [141].

Аналогичный вывод сделан в коллективном призыве к библиоразнообразию (призыв Жюсье):

«Система оценки научных исследований должна быть тщательно реформирована и адаптирована к практике научной коммуникации» [88].

Действительно, во многих коллективных документах [118; 127] детально проанализированы негативные последствия для науки, к которым ведет использование библиометрических показателей, таких как импакт-фактор журнала, индекс Хирша, публикационная активность и др., в качестве основного метода оценки научного результата. Впрочем, это не удивительно, поскольку, согласно известному закону Гудхарта, «когда всякий показатель становится инструментом управления, прежние эмпирические закономерности, использующие данный показатель, перестают действовать».

С другой стороны, альтернативы для библиометрии пока не выработано, хотя большинство авторов сходятся во мнении, что основой для оценки научных результатов должна стать качественная экспертиза научного результата, учитывающая цели исследования.

Поскольку, как уже сказано, система оценки научных результатов существенно влияет на всю организацию научных коммуникаций, то общую структуру коммуникаций в будущем при условии изменения этой системы оценки представить сейчас достаточно трудно. Однако можно допустить, что существующие процессы аналитической и реферативной обработки научной информации могут быть модифицированы в сторону экспертизы и создания рекомендательных сервисов.

Коллаборации

Коллаборации в науке стали одной из важнейших тенденций по развитию науки в целом и научных коммуникаций в частности. Некоторые данные по состоянию коллабораций в российской науке содержатся в работе [30].

Центром нашего интереса являются коллаборации при создании научных информационных ресурсов. Конечно, прекрасным примером коллективного создания действительно массового информационного ресурса является Википедия. Хотя Википедия и продолжает подвергаться критике за спорность некоторых данных, она является наиболее посещаемым справочным информационным ресурсом как во всем мире, так и в России. Эффективными оказались и другие научные и образовательные ресурсы, использующие технологию вики, например Летописи.ру. Описание различных подходов к коллективным формам организации научной деятельности имеется, например, в работе [42].

Хорошим примером распределенной коллективной работы над научным информационным ресурсом является участие многих российских лингвистов в разметке текстов в ходе создания Национального корпуса русского языка [72].

Фактором, активно влияющим на переход к коллективным формам научной и научно-информационной деятельности, стал, конечно, переход к разным формам открытой науки, и особенно к открытым научным данным. Приведем цитату из Конкордата открытых научных данных:

«Во многих областях, таких как кристаллография, генетика, археология и лингвистика, существует ряд превосходных примеров использования открытых данных. Это обеспечивает прогресс исследований в данных областях, повышает эффективность использования ресурсов и, следовательно, улучшает финансирование науки.

В этих областях созданы практические механизмы обеспечения открытости исследований, которые повышают культуру транспарентности, защищают от мошенничества и непредвиденных ошибок, улучшают репутацию научных учреждений

Доступ к открытым данным во многих областях стимулирует развитие новых подходов в научных исследованиях, поскольку происходит объединение данных из различных источников. Это позволяет по-новому взглянуть на междисциплинарные проблемы в самых разных областях. Во многих случаях новые подходы

обеспечиваются объединением данных ряда государственных и коммерческих организаций с данными, полученными академическими исследователями.

Открытые данные могут лежать в основе инноваций, например, когда исследователи используют данные неожиданными способами или когда компании используют данные для разработки новых продуктов. Это может привести к существенным экономическим выгодам» [146].

Обзор зарубежного опыта по созданию открытых данных на примере лингвистических ресурсов приводится в главе 6.

В то же время переход к коллаборациям во многих случаях имеет негативные последствия, а также ставит новые проблемы. В частности, усиливается острота проблем, связанных с авторскими правами при использовании чужих данных, ответственностью и правомочиями при распоряжении коллективными ресурсами, распределением затрат и т.д.

Будущее научных библиотек и информационных центров

Исходя из описанных выше тенденций, можно представить себе будущее научных библиотек¹ следующим образом.

Объемы комплектования, хранения традиционных фондов и обслуживания по ним должны медленно сокращаться, впрочем, в обозримом будущем соответствующие службы ликвидироваться не должны. Одновременно должна возрасти роль служб, которые реализуют различные пользовательские сервисы на электронных ресурсах. Это учет, каталогизация и систематизация электронных ресурсов, улучшение доступа к ним, агрегация ресурсов, рекомендательные сервисы, развитие лингвистических средств поиска, персонализация поиска, создание и поддержание специализированных электронных библиотек².

Важнейшая функция научных библиотек, которые фактически должны стать научно-информационными центрами, обслужи-

¹ Понятно, что массовые библиотеки должны модифицироваться иначе, в сторону расширения их функций как культурно-досуговых центров для населения (с основной просветительской функцией).

² Специализированные электронные библиотеки в перспективе нужны будут тогда, когда они предоставляют сервисы или фонды, дополняющие универсальные электронные библиотеки.

вающими определенную тематику, регион, корпорацию, университет или научно-образовательный центр, – это экспертно-оценочная деятельность. Она должна включать анализ электронных ресурсов, их отбор и оценку, создание рекомендательных библиографий, избирательное распространение информации, другие сервисы, в которых заинтересован пользователь данной библиотеки, включая решение правовых вопросов, связанных с доступом к электронным ресурсам.

В ряде случаев научные библиотеки могли бы взять на себя некоторые издательские функции, например печать по требованию (print-on-demand). Интеграция библиотечных и издательских функций в цифровой среде, особенно в университетах, в целом представляется весьма перспективным направлением развития. Примером российской научной библиотеки, которая развивается в таком направлении, объединяя функции научно-информационного центра, вузовской библиотеки, а также многие другие, является Центральная научная медицинская библиотека Первого МГМУ им. И.М. Сеченова [110].

Библиографические продукты, как представляется автору, могут быть представлены в будущей цифровой среде в виде сводной базы данных национальной библиографии, вероятно, на основе баз данных Книжной палаты и библиографических агрегаторов, таких как ЛИБНЕТ, ЭКБСОН или АРБИКОН, а также каталога Национальной электронной библиотеки. Эта сводная база данных должна выполнять следующие функции:

- обеспечивать корпоративную каталогизацию и развитие электронных каталогов библиотек;
- обеспечивать навигационные функции – поиск местонахождения изданий;
- быть справочным аппаратом для национальной программы оцифровки;
- способствовать обеспечению сохранности книжного фонда и созданию Страхового книжного фонда.

Реферативные службы могут иметь будущее, только если они возьмут на себя функции оценки и экспертизы реферируемых публикаций, включая функции выявления плагиата. Иначе говоря, реферативные службы разумно переориентировать на обработку и оценку отечественной литературы. Такие функции наиболее естественно реализовывать в рамках репозитория научных публикаций.

Более подробное описание современного состояния российской научной библиотечно-информационной сферы представлено в главе 2.

Центральной задачей управления библиотечно-информационной сферой в России должна стать выработка оптимального соотношения между универсальными и частными (специализированными) ресурсами и сервисами научной информации, равно доступными в единой цифровой среде. Так, например, в случае электронных библиотек мы наблюдаем значительное дублирование при их создании, что означает неэффективное использование довольно ограниченных общественных ресурсов

Тематические информационные центры

Наряду с библиотеками будущего, превратившимися в научно-информационные центры, хорошие перспективы должны иметь специализированные структуры, обеспечивающие реализацию различных сервисов для определенной категории научно-информационных ресурсов.

Так, в настоящее время в рамках реализации исследовательского проекта РФФИ № 18-00-002/18 «Интеграция научно-информационных ресурсов учреждений РАН (на примере языкознания) как части единого цифрового пространства РАН» автор принимает участие в разработке концепции создания Центра лингвистических ресурсов РАН. Этот центр предполагается создать на базе служб поддержки Национального корпуса русского языка при Институте русского языка РАН. Более подробное описание принципов создания такого центра приводится в главе 6 настоящей монографии.

Возможно, в рамках социальных и гуманитарных наук должно возникнуть еще несколько тематических информационных центров (например, по социологическим или этнографическим данным, синологии, монголоведению и др.). Информационный ресурс по отечественной истории фактически уже создан на базе Президентской библиотеки им. Б.Н. Ельцина, хотя его функции в общей системе научной информации по социальным и гуманитарным наукам еще придется определить.

Публикации и данные

Основной массив научной информации, отражающей результаты научных исследований, представляет собой научные публикации, а также некоторые категории непубликуемых документов (отчеты, диссертации). Однако в современной научной коммуникации, особенно в условиях открытой науки, возрастает значение разнообразных результатов научных исследований, не прошедших традиционный путь оформления через публикации, отчеты и диссертации. Чаще всего такие результаты изначально появляются в цифровой форме (born digital). Эти результаты обобщенно стали называть «данными».

Так, в цитированном выше коллективном документе по управлению открытыми научными данными (Конкордате), который разработан консорциумом университетов Великобритании, следующим образом определяется понятие научных данных:

«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.» [146].

В Конкордате можно найти множество ссылок, описывающих опыт работы с научными данными.

Понятие научных данных дополняет понятие научных публикаций, которые ранее рассматривались как основная форма научной коммуникации.

Отметим, что работа с данными никогда не входила в сферу библиотечной деятельности, ориентированной на работу с публикациями, но входила в сферу интересов органов научной информации¹.

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

- хорошее управление данными имеет фундаментальное значение для всех этапов исследовательского процесса и должно быть обеспечено с самого начала;

¹ Опыт работы с данными (с фактографической информацией) в институте информации оборонного комплекса освещен в монографии [49].

- управление данными имеет жизненно важное значение для того, чтобы данные были полезными для других, а также для сохранения данных в долгосрочной перспективе;

- данные, на которые ссылается публикация, должны быть доступны к дате публикации и должны быть представлены в цитируемой форме;

- обучение работе с данными является сферой ответственности всех заинтересованных сторон; администрирование, архивирование, манипулирование и анализ данных требуют набора навыков, отличных от тех, которые используются для сбора, генерации или измерения данных.

Перспективы развития информационных языков

Параллельно с использованием традиционных информационных языков в последние годы создавались новые современные семантические и лексикографические ресурсы, в том числе в форме тезаурусов и онтологий, словарей на основе корпусов и др. Таких ресурсов только в РАН создано уже много десятков. Их описание содержится в Навигаторе информационных ресурсов по языкознанию, описанном в главе 6, а каталоги зарубежных и отечественных лингвистических ресурсов представлены в приложениях 14–16.

Задача состоит в организации сотрудничества для разработки и развития лингвистических средств как для научных целей, так и для прикладных задач, связанных с применением семантических технологий в научной коммуникации.

Экономика научных коммуникаций

Основное содержание социально-экономической составляющей научно-информационной деятельности сегодня – это конкуренция трех способов народно-хозяйственной деятельности: государственного, рыночного (коммерческого) и волонтерского (общественного). При этом и российский, и зарубежный опыт с очевидностью показывает, что доминирование любого из секторов приводит к негативным последствиям. Поэтому задача заключается в выстраивании баланса между экономическими моделями научных коммуникаций.

Однако механизм достижения этого баланса совершенно не ясен, тем более что ведется активная борьба владельцев научных ресурсов и средств коммуникаций за доходы в рыночных условиях. Владельцы научной информации, будь то научные организации, журналы или архивы, сплошь и рядом пытаются извлечь прямую финансовую выгоду из факта владения научной информацией, независимо от того, насколько это соответствует или противоречит общественным интересам.

Проблемы оптимизации экономических моделей научных коммуникаций составляют основное содержание дискуссий, связанных с переходом к открытой науке. Достаточно подробный анализ этих проблем содержится в докладе экспертов Европейского союза о будущем научных коммуникаций [127].

Специфика России заключается в том, что у нас вообще нормативно не определена роль основных акторов научной коммуникации, к которым относятся:

- государство в лице органов власти и управления;
- Российская академия наук;
- научные фонды;
- университеты и научно-исследовательские институты;
- издательства и редакции;
- коммерческие информационные сервисы;
- общественные информационные сервисы;
- библиотеки (традиционные и цифровые);
- архивы (традиционные и цифровые);
- авторы;
- потребители научной информации.

Особенно это удивительно на фоне грандиозных планов развития цифровой экономики. Признавая теоретически неизбежный переход к экономике нового типа, в том числе к экономике знания и экономике внимания, признавая также, что научная информация относится к общественным благам, руководство российской науки и образования тем не менее до сих пор не выработало адекватный взгляд на экономику научных коммуникаций.

Вместе с тем теоретических исследований этих проблем в России было проведено немало. Упомянем, в частности, монографии А.Б. Долгина [40; 41], а также фундаментальный отчет А.Н. Козырева и его коллег [94], в которых подробно рассматривается возможность применения различных экономических моделей к организации научной коммуникации.

Однако, насколько известно автору, результаты этих исследований не нашли отражения в экономической практике научных коммуникаций. Поэтому прогнозировать развитие научной коммуникации, не имея ясного представления о ее экономике, достаточно затруднительно. Впрочем, стратегия, направленная на огосударствление многих функций в общественной сфере, реализуемая в последние годы, дает некоторые возможности для прогноза, хотя ожидаемый результат не представляется оптимальным.

Модификация законодательства

Одним из важных препятствий для развития научных коммуникаций в цифровой среде является действующее законодательство в области авторского права. Правовые проблемы развития научной инфосферы достаточно подробно рассмотрены в цитированной выше монографии «Инфосфера общественных наук России» [22]. Там же изложены предложения по изменению российского законодательства в области авторского права, которое давно не соответствует современному уровню коммуникаций. Добавим, что законодательство многих стран, особенно европейских, существенно модифицировалось в направлении обеспечения свободы распространения научного знания.

Однако, несмотря на назревшую необходимость, в России пока не торопятся вносить изменения в законодательство, хотя законодательных актов, регулирующих деятельность в Интернете, принимается более чем достаточно. Остается надеяться, что стимулом для внесения в законодательство в области авторского права изменений, необходимых для развития научных коммуникаций, станет правительственная программа «Цифровая экономика», реализация которой потребует серьезной модернизации законодательной базы.

Единое российское электронное пространство знаний (ЕРЭПЗ)

Среди предложений, которые должны формировать образ будущего, особое место занимает идея Единого российского электронного пространства знаний (ЕРЭПЗ). Эта идея провозглашена в ряде нормативных документов и представлена как перспектива

развития Национальной электронной библиотеки. Проблемам создания ЕРЭПЗ и его соотношения с системой научной информации в России полностью посвящена глава 5 данной книги.

1.2. Инициативы мировой научной общественности по развитию научных коммуникаций в условиях открытой науки

В последние годы мировая научная общественность, особенно в странах Европейского союза, активно обсуждает проблемы перехода к открытой науке и в связи с этим – будущее научных коммуникаций в цифровой среде. В результате появился ряд деклараций, экспертных докладов и других коллективных документов на эту тему. Особый резонанс получило решение Агентства по науке Европейской комиссии об открытии с 2020 г. всех публикаций по результатам исследований, финансируемых европейскими правительствами и фондами [134].

В настоящем разделе представлен обзор основных коллективных документов, освещающих проблемы перехода к открытой науке и тесно связанную с этим переходом проблему оценки результатов научной деятельности. В этих документах возникает образ будущего научных коммуникаций в условиях их функционирования в открытой цифровой среде. Главная проблема, конечно, заключается в возможности применения в российских условиях идей и принципов, провозглашаемых в коллективных документах по открытой науке.

Исходные документы по открытой науке

Первое важное международное упоминание об открытом доступе содержится в материалах Будапештской конференции по открытому доступу, проходившей в феврале 2002 г. [38].

Вскоре за ней последовала Берлинская декларация об открытом доступе к научным и гуманитарным знаниям в октябре 2003 г. [29].

Далее последовало принятие ряда документов отдельными научными союзами и организациями, среди которых нужно отметить документы Организации экономического развития и сотрудничества (ОЭСР) и Фонда открытого знания. В 2004 г. ОЭСР (Organisation for Economic Cooperation and Development, OECD)

было подписано коммюнике, в котором сформулировано требование доступности для граждан архивов всех исследований, финансируемых за счет государственных средств [142].

В 2007 г. были опубликованы «Принципы и положения для обеспечения доступа к научным данным исследований, финансируемых государством» [132]. В них говорится о том, что в связи с высокой скоростью развития Интернета и технологий, связанных с передачей, хранением и обработкой информации, появились новые возможности по совместному использованию научных данных мировым сообществом. В качестве наглядного примера в них приводится международный проект «Геном человека».

В июле 2009 г. был опубликован первый вариант *Пантонских принципов*, которые представляют собой набор рекомендаций по опубликованию научных данных, доступных для их последующего использования всеми заинтересованными лицами. В 2010 г. Пантонские принципы были уточнены и доработаны при участии членов рабочей группы «Открытые данные в науке» Фонда открытого знания [138].

В Пантонских принципах, в частности, говорится о том, что научные данные должны стать частью общественного достояния и публиковаться с соответствующей лицензией таким образом, чтобы предоставлять возможность их повторного использования.

В 2009 г. Европейская организация по ядерным исследованиям (CERN) опубликовала книжный каталог открытых данных в рамках проекта «Открытая библиотека» [136]. ЦЕРН поддерживает инициативу открытого доступа, поэтому неудивительно, что она заключила соглашение с рядом организаций, по которому научные данные, полученные при помощи Большого адронного коллайдера, публикуются в открытом доступе под лицензией Creative Commons.

Открытый доступ и оценка научных исследований. Декларация DORA

Достаточно быстро научному сообществу стало очевидно, что проблема открытого доступа теснейшим образом связана с проблемой оценки результатов научной деятельности. В 2012 г. была опубликована *Сан-Францисская Декларация об оценке научного исследования*, получившая известность под акронимом DORA [141]. С этого момента все дальнейшие документы по открытому доступу ссылаются на принципы DORA и выражают им поддержку.

Ввиду важности Декларации DORA для дальнейшего развития представлений о развитии научных коммуникаций мы приводим ее рекомендации полностью¹.

Общая рекомендация

1. Не используйте такие количественные журнальные показатели, как импакт-фактор, в качестве косвенного показателя качества отдельных научных статей при оценке вклада конкретных ученых, при приеме на работу, продвижении по службе или при принятии решений о финансировании.

Для финансирующих организаций

2. Четко формулируйте критерии, используемые при оценке научной продуктивности соискателей грантов, и внятно объясняйте, особенно исследователям в начале работы, что научное содержание статьи гораздо важнее, чем публикационные показатели или тип журнала, в котором она была опубликована.

3. При оценке исследования рассматривайте ценность и влияние всех его результатов, включая, в дополнение к научным публикациям, массивы данных и программное обеспечение, а также рассматривайте широкий спектр вариантов использования результатов исследования, включая такие качественные индикаторы воздействия исследования, как его влияние на политику и практику.

Для институтов

4. Четко формулируйте критерии, используемые при принятии решений о приеме на работу и продвижении по службе, определении срока полномочий, внятно объясняя, особенно исследователям в начале работы, что научное содержание статьи гораздо важнее, чем публикационные показатели или тип журнала, в котором она была опубликована.

5. При оценке исследования рассматривайте ценность и влияние всех его результатов, включая, в дополнение к научным публикациям, массивы данных и программное обеспечение, а также рассматривайте широкий спектр вариантов использования результатов исследования, включая такие качественные индикаторы воздействия исследования, как его влияние на политику и практику.

Для издателей

6. Делайте гораздо меньший упор на импакт-фактор журнала в качестве инструмента продвижения, в идеале прекращая продвигать импакт-фактор или представляя систему показателей в кон-

¹ Перевод: В.И. Карнышев <https://docviewer.yandex.ru/view/15710660/>.
Источник: http://piohelp.ru/myfiles/dora/DORA_rus.html

тексте различных журнальных метрик, например пятилетний импакт-фактор, EigenFactor, SCImago, индекс Хирша, время на редакционную обработку и опубликование и т.д., которые дают более полноценное представление об эффективности журнала.

7. Внедряйте ряд показателей, касающихся именно статьи, поощряйте переход к оценке, основанной на научном содержании статьи, а не на публикационной метрике журнала, в котором она была напечатана.

8. Поощряйте практику ответственного авторства и предоставления информации о конкретном вкладе каждого из авторов.

9. Если журнал является изданием с открытым доступом или подписным, то удалите все ограничения на повторное использование списков литературы в научных статьях и сделайте их доступными в соответствии с актом передачи в общественное достояние.

10. Удалите или уменьшите ограничения на количество использованных источников в научных статьях и при необходимости потребуйте цитирования первоисточников вместо обзоров, для того чтобы отдать предпочтение группе, которая первой сообщила о полученном результате.

Для организаций, предоставляющих информацию о количественных показателях

11. Будьте открытыми и прозрачными, предоставляя данные и делая публичными методы, используемые для расчета всех показателей.

12. Предоставляйте данные по лицензии, которая дает возможность неограниченного использования, и обеспечивайте вычислительный доступ к данным, когда это возможно.

13. Откровенно говорите о том, что некорректные манипуляции метриками допускаться не будут; четко сообщайте о том, что собой представляют некорректные манипуляции и какие меры будут приняты для борьбы с ними.

14. Учитывайте разброс типов статей (например, обзоры в сравнении с научными статьями), а также различие предметных областей, когда используются объединенные или сопоставительные метрики.

Для исследователей

15. Участвуя в комиссиях, принимающих решения о выделении финансирования, приеме на работу, определении срока полномочий и продвижении по службе, делайте оценки на основе научного содержания, а не публикационных метрик.

16. В подходящих случаях цитируйте не обзоры, а первоисточники, в которых полученные результаты излагались впервые, для того чтобы указать на авторов, когда это нужно.

17. Используйте совокупность статейных метрик и указателей на персональную или вспомогательную информацию как доказательство влияния отдельных опубликованных статей и других результатов исследований.

18. Подвергайте сомнению устоявшийся порядок оценки научного исследования, который необоснованно опирается на импакт-фактор журналов, содействуйте и осваивайте передовой опыт, который учитывает в первую очередь значение и влияние конкретных результатов исследований.

Лейденский манифест

Дальнейшее развитие принципы DORA получили в *Лейденском манифесте для наукометрии* [118], в котором сформулированы 10 принципов оценки научной деятельности, которые приводятся ниже¹.

1. Количественная оценка должна дополнять качественную, экспертную оценку. Количественные измерения могут уравновесить возможное предубеждение перед экспертным рецензированием (peer review) и упростить обсуждение. Они должны усиливать экспертное рецензирование, поскольку трудно оценивать коллег, не владея спектром необходимых сведений. Тем не менее специалисты, проводящие оценку научной деятельности, не должны следовать соблазну переложить принятие решений на числа. Индикаторы – не замена информированному суждению. Каждый сохраняет ответственность за свою оценку.

2. Сопоставляйте научную деятельность с исследовательскими задачами организации, группы или ученого. Цели исследовательской программы должны быть описаны в начале работы, и индикаторы, используемые для оценки научной деятельности, должны четко соответствовать этим целям. При выборе индикаторов и путей их использования должен приниматься во внимание широкий социально-экономический и культурный контекст. У ученых разные научные задачи. Исследование, сдвигающее границы научного знания, отличается от исследования, сосредото-

¹ Перевод А.А. Исэрова.

ченного на поиске решений общественных проблем. Экспертная оценка может быть основана не только на академических идеях о научных достижениях, но и принимать во внимание достоинства, важные для политических решений, промышленности или общества. Ни одна модель оценки не применима ко всем контекстам.

3. Отстаивайте научное качество в исследованиях, важных для того или иного региона. Во многих частях мира высокое качество научного исследования приравнено к публикациям на английском языке. Испанское законодательство, к примеру, утверждает желательность публикаций испанских ученых в журналах с высоким импакт-фактором. Импакт-фактор рассчитывается по созданной в США и все еще в основном англоязычной базе Web of Science. Такого рода предубеждение создает особые проблемы в общественных и гуманитарных науках, где исследования в большей степени регионально и национально обусловлены. Многие другие дисциплины также имеют национальное или региональное измерение – например, эпидемиология ВИЧ в Африке южнее Сахары.

Этот плюрализм и общественная значимость могут приноситься в жертву при написании текстов, которые представляли бы интерес для «сторожей» высокого импакт-фактора – англоязычных журналов. В Web of Science широко цитируют тех испанских социологов, которые работают на абстрактных моделях или изучают данные по США. Теряются характерные черты работы тех социологов, чьи испаноязычные статьи имеют высокий импакт-фактор, с такими темами, как местное рабочее законодательство, здравоохранение для пожилых семей или занятость иммигрантов. Наукометрия, основанная на высококачественной неанглоязычной литературе, поможет определить и вознаградить высокий научный уровень в исследованиях, значимых для конкретных регионов.

4. Сохраняйте процесс сбор данных и аналитические процессы открытыми, прозрачными и простыми. Создание баз данных, требуемых для оценки, должно четко следовать определенным правилам, установленным до завершения оцениваемого исследования. В последние десятилетия это было обычной практикой среди академических и коммерческих групп, создававших методологию библиометрической оценки. Эти группы публиковали свои протоколы в рецензируемой литературе. Такая прозрачность делала возможной тщательную проверку. Например, в 2010 г. общественное обсуждение технических качеств одного важного индикатора, используемого одной из этих групп (Центром изучения науки и технологии в Лейденском университете в Нидерландах),

привело к пересмотру расчета этого индикатора. Коммерческие организации, которые недавно начали заниматься подобной деятельностью, должны отвечать тем же стандартам; никто не должен соглашаться с существованием «машины для оценки» в черном ящике.

Простота – достоинство для индикатора, поскольку увеличивает прозрачность. Но грубо упрощенная наукометрия может искажать результаты (см. принцип 7). Эксперты, занимающиеся оценкой, должны стремиться к балансу – простым индикаторам, соответствующим сложности исследовательского процесса.

5. Позволяйте оцениваемым исследователям проверять данные и анализ. Чтобы обеспечить качество данных, все исследователи, включенные в библиометрические подсчеты, должны получить возможность проверки верного определения их научных результатов. Каждый, кто руководит процессами оценки или занимается ими, должен обеспечивать правильность данных через самопроверку или проверку третьей стороной. Университеты могут реализовывать этот принцип в своих информационно-исследовательских системах, и именно он должен стать основным в отборе поставщиков этих систем. Чтобы получить точные, высококачественные данные, нужны время и деньги. Отведите на это средства.

6. Дисциплины отличаются друг от друга по практике публикаций и цитирования. Лучше всего составлять ряд возможных индикаторов и позволять дисциплинам выбирать среди них. Несколько лет назад группа европейских историков получила относительно низкий рейтинг в оценке, поскольку они писали больше книг, а не статей для журналов, индексируемых Web of Science. Этим историкам не повезло – они работали в департаменте психологии. Историкам и обществоведам важно, чтобы в подсчет их публикаций входили книги и литература на национальных языках; специалистам по информатике нужно, чтобы учитывались доклады на конференциях.

Количество цитат зависит от дисциплины: математические журналы с наивысшим рейтингом имеют импакт-фактор около 3, а такие же журналы по клеточной биологии – около 30. Требуются нормализованные индикаторы, и наиболее убедительный метод нормализации основан на процентилях: каждый текст оценивается на основе процентиля, к которому он принадлежит в распределении цитат в своей дисциплине (к примеру, верхние 1%, 10%, 20%). Одна высокочитируемая публикация несколько улучшает положение университета в рейтинге, основанном на процентильных индикато-

рах, но может вознести университет с середины на самый верх рейтинга, разработанного на средних показателях цитирования.

7. Основывайте оценку отдельных исследователей на качественной оценке их резюме. Чем вы старше, тем выше ваш h-индекс, даже если вы больше ничего не публикуете. H-индекс отличается по дисциплинам: максимум у ученых в области наук о жизни составляет около 200, у физиков – 100, обществоведов – 20–30. Он зависит от базы данных: есть исследователи, чей h-индекс составляет около 10 в Web of Science, но 20–30 в Google Scholar [117]. Чтение и оценка работы исследователя куда важнее, чем опора только на один показатель. Даже в сопоставлении больших групп ученых наилучшим подходом будет тот, где принимается во внимание больше сведений об уровне знаний, опыте, деятельности и влиянии отдельного исследователя.

8. Избегайте неуместной конкретности и ложной точности. Научно-технологические индикаторы грешат концептуальной двусмысленностью и неопределенностью, так что требуются четкие постулаты, с которыми не все соглашаются. К примеру, долго обсуждалось значение подсчетов цитат. Итак, лучше всего использовать разнообразные индикаторы, чтобы обеспечить более убедительную и многомерную картину. Если погрешности и ошибки можно квантифицировать, например, через показатель величины ошибки («усы»), то такие сведения должны сопровождать публикацию показателей индикаторов. Если это невозможно, то те, кто подсчитывает индикаторы, должны по крайней мере избегать ложной точности. К примеру, журнальный импакт-фактор публикуется с тремя десятичными знаками, чтобы избежать совпадения показателя. Тем не менее, если принять во внимание концептуальную двойственность и случайную изменчивость подсчетов цитат, нет смысла различать журналы на основе очень маленькой разницы в импакт-факторе. Избегайте ложной точности: гарантирован только один десятичный знак.

9. Признавайте системное воздействие оценки и индикаторов. Индикаторы меняют систему через те стимулы, которые они устанавливают. Это воздействие нужно предугадывать. Это означает, что набор индикаторов всегда предпочтителен – использование лишь одного создаст дух азартной игры и сместит цель (целью станет показатель). Например, в 1990-е годы Австралия финансировала университетские исследования на основе формулы, в основном построенной на количестве материалов, опубликованных организацией. Университеты могут подсчитать «стоимость»

материала в рецензируемом журнале; в 2000 г. она составляла 800 австралийских долларов (около 480 долл. США в том году) в финансировании исследований. Предсказуемо, что число материалов, публикуемых австралийскими учеными, выросло, но размещаться они стали в менее цитируемых журналах, что позволяет сделать предположение о падении качества статей.

10. Регулярно подвергайте индикаторы тщательной проверке и пересмотру. Исследовательские задачи и цели оценки меняются, и с ними развивается и исследовательская система. Когда-то полезная наукометрия становится неадекватной, но возникает новая. Системы индикаторов нужно пересматривать и порой менять. Поняв последствия упрощенности своей формулы, Австралия в 2010 г. ввела более сложный показатель – инициативу «Исследовательское превосходство для Австралии», которая делает акцент на качестве.

Гаагская декларация по открытию знаний в цифровую эпоху

Одно из важных направлений научной коммуникации в цифровую эпоху – это возможность и целесообразность обеспечения доступа не только к публикациям, но и к различным первичным данным, послужившим для получения научного результата.

Существенный вклад в развитие идеологии открытого доступа к научным данным сыграла Гаагская декларация по открытию знаний в цифровую эпоху, подготовленная Ассоциацией европейских научных библиотек (LIBER) и открытая к подписанию в 2015 г. [35]. В этой декларации провозглашается право на автоматическое извлечение фактов, идей и данных из легально опубликованного контента, формулируются принципы свободной обработки текстов и данных и предлагается дорожная карта для осуществления этих принципов.

Конкордат по открытым научным данным

Дальнейшим развитием подхода, представленного Гаагской декларацией, было опубликование в 2016 г. Конкордата по открытым научным данным [146]. В нем приводится определение открытых научных данных:

«Открытые научные данные – это научные данные, к которым можно получить свободный доступ, использовать, изменять и

совместно использовать; доступные для тестирования и проверки другими исследователями.

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.».

В конкордате сформулированы принципы управления научными данными, обобщающие высказанные различными авторами идеи по данной тематике, которые мы приводим в сокращенном виде:

- открытый доступ к данным исследований обеспечивает высокое качество исследований, содействует инновациям и способствует распространению лучших исследовательских практик;

- открытость исследовательских данных может быть ограничена, но любые ограничения должны быть обоснованными и оправданными;

- открытый доступ к данным исследований сопряжен со значительными затратами, которые должны учитываться всеми сторонами;

- использование чужих данных всегда должно соответствовать правовым, этическим и нормативным требованиям, включая соответствующие ссылки;

- должно быть установлено право создателей научных данных на их разумное использование;

- хорошее управление данными имеет фундаментальное значение для всех этапов исследовательского процесса и должно быть обеспечено с самого начала;

- управление данными имеет жизненно важное значение для того, чтобы данные были полезными для других, а также для сохранения данных в долгосрочной перспективе;

- данные, на которые ссылается публикация, должны быть доступны к дате публикации и должны быть представлены в цитируемой форме;

- обучение работе с данными является сферой ответственности всех заинтересованных сторон; администрирование, архивирование, манипулирование и анализ данных требуют набора навыков, отличных от тех, которые используются для сбора, генерации или измерения данных;

- должны готовиться регулярные обзоры прогресса в области открытых научных данных.

Амстердамский призыв

В последние годы появились некоторые новые коллективные обращения ученых по переходу к открытой науке, среди которых мы отметим Амстердамский призыв [2].

Амстердамский призыв содержит систематическое изложение основных проблем перехода к открытой науке и развитию научных коммуникаций. Эти проблемы сформулированы следующим образом.

Устранение барьеров для открытой науки

1. Изменение коммуникаций, оценка изменений и системы вознаграждения в науке.

2. Интеллектуальный анализ содержания текста и данных.

3. Трактовка прав интеллектуальной собственности и защиты личных данных в цифровой среде.

4. Прозрачность затрат и условий академических коммуникаций.

Развитие исследовательской инфраструктуры

5. Внедрение справедливых и безопасных принципов коммуникации.

6. Организация общей инфраструктуры электронных ресурсов.

Стимулирование и создание стимулов для открытой науки

7. Принятие принципов открытого доступа.

8. Стимулирование новых моделей публикаций для передачи знаний.

9. Стимулирование научно обоснованных инноваций по открытой науке.

Актуализация и дальнейшее поощрение открытой научной политики

10. Разработка, реализация и контроль планов по открытому доступу.

11. Привлечение исследователей и пользователей к открытой науке.

12. Поощрение заинтересованных сторон к обмену опытом и информацией по открытой науке.

Каждая проблема описана в стандартной форме:

- описание проблемы;
- пути решения проблемы;
- план конкретных действий по решению проблемы;
- положительный эффект от решения проблемы.

Призыв Жюсье

В Призыве Жюсье [88], принятом группой исследователей, связанных с Европейской научной цифровой библиотекой, говорится о поддержке Амстердамского призыва и формулируется цель:

«Наша цель заключается в разработке и внедрении альтернативных бизнес-моделей научной коммуникации, соответствующих целям открытой науки, путем поддержки инноваций для полного обновления издательских функций, провозглашенных Ассоциацией европейских исследовательских библиотек (LIBER) и Международным советом по науке (ICSU)».

Основные принципы Призыва Жюсье:

- открытый доступ должен дополняться поддержкой разнообразия в научных коммуникациях (*библиоразнообразие*), чтобы положить конец доминированию монополистов;

- разработка инновационных моделей научных публикаций должна быть приоритетом, поскольку они отвечают реальным потребностям исследователей в наш цифровой век;

- следует поощрять эксперименты в области публикаций (публикация сопутствующих данных), рецензирования (открытое рецензирование), редакторских услуг (веб-публикация за пределами pdf) и дополнительных услуг (интеллектуальный анализ текста и данных);

- система оценки научных исследований должна быть существенно реформирована и адаптирована к практике научной коммуникации.

План S и рекомендации по его реализации

Важным событием, связанным с развитием открытой науки, стало решение Европейской комиссии об открытии с 2020 г. всех публикаций по результатам исследований, финансируемых европейскими правительствами и фондами (OA2020 Initiative) [134].

Для реализации этого решения в сентябре 2018 г. был принят так называемый План S, разработанный коалицией научных организаций 13 стран Евросоюза (коалицией S). С одной стороны, о поддержке Плана S заявили многие организации различных стран, в том числе США и Китая. С другой стороны, ряд научных организаций высказали свои замечания, особенно по практическим рекомендациям коалиции S по выполнению этого плана. Одноре-

менно научные организации Европы ведут активные переговоры с ведущими научными издательствами по практическим шагам по переходу к открытому доступу, что зачастую связано с конфликтами между участниками научных коммуникаций.

Проф. В.М. Московкин в «Троицком варианте» публикует 10 принципов Плана S:

«План S существенно проясняет механизм OA2020 Initiative, благодаря сформулированным в нем 10 принципам.

1. Авторы сохраняют права на свои научные труды без каких-либо ограничений. Эти труды должны публиковаться на основе открытых лицензий (предпочтительно под Creative Commons Attribution Licence CCBY). Во всех случаях лицензия должна удовлетворять требованиям, определенным Берлинской декларацией.

2. Фонды будут сообща обеспечивать введение четких критериев и требований для служб, которые должны поддерживать соответствующее высокое качество ОА-журналов и ОА-платформ.

3. В случае отсутствия высококачественных журналов и платформ открытого доступа фонды будут координировать свои действия, чтобы обеспечить такое качество, а также будут содействовать созданию инфраструктуры открытого доступа там, где это необходимо.

4. Там, где это приемлемо, стоимость публикаций открытого доступа будет покрываться фондами или университетами, но не самими исследователями. Признаётся, что все ученые должны иметь возможность публиковать свои работы в открытом доступе, даже если их организации ограничены в средствах.

5. Финансирование публикаций открытого доступа в странах ЕС стандартизируется и реализуется единообразным способом.

6. Фонды будут требовать, чтобы университеты, исследовательские организации и библиотеки выстроили свои ОА-политики и стратегии надлежащим образом, особенно по части прозрачности.

7. Вышеуказанные принципы должны быть применимы для всех типов научных публикаций, но для монографий и книг срок запуска Плана S может быть смещен на более позднее время.

8. Важность ОА-репозитория и архивов для размещения научных результатов подтверждается из-за их функции долговременного хранения и потенциала для редакционно-издательских инноваций.

9. Гибридная модель публикаций считается не соответствующей вышеуказанным принципам.

10. Фонды будут контролировать соответствие этим принципам и применять санкции при их нарушении» [57].

В ноябре 2018 г. на сайте Science Europe (Брюссель) было опубликовано «Руководство по выполнению Плана S» [128]. Содержание этого важного документа подробно изложено и прокомментировано еще в одной публикации В.М. Московкина [59]. Здесь мы публикуем только содержание этого документа, включающего такие разделы:

- проверка соответствия Плану S;
- определение расходов на публикации;
- поддержка журналов и платформ открытого доступа;
- сроки реализации, контроль и санкции;
- лицензирование и права;
- обязательные общие критерии;
- критерии качества и дополнительные критерии для журналов и платформ ОД, совместимых с Планом S;
- размещение научного контента в репозиториях открытого доступа: требования к репозиториям, авторам и издателям;
- соглашения о переходе к открытому доступу.

Первое публичное обсуждение упомянутого руководства состоялось в декабре 2018 г. на XIV Берлинской конференции по открытому доступу.

Официальное заявление этой конференции гласит:

- мы все привержены сохранению авторских прав авторов;
- мы все стремимся к полному и немедленному открытому доступу;
- мы все привержены ускорению прогресса в обеспечении открытого доступа на основе трансформационных соглашений, которые носят временный и переходный характер, с переходом к полному открытому доступу в течение нескольких лет. Эти соглашения должны быть нейтральными, с расчетом на то, что экономика будет корректироваться в соответствии с развитием рынка;
- ожидается, что издатели будут работать со всеми членами глобального исследовательского сообщества для обеспечения полного и немедленного открытого доступа в соответствии с этим заявлением.

Подробное содержание дискуссий, происходивших на этой конференции, представлено в статье В.М. Московкина, которая завершается следующим выводом:

«Таким образом, дискуссии по поводу новой бизнес-модели на рынке научной периодики продолжаются, и вряд ли в скором

времени мы получим единодушное и безоговорочное признание всех положений Open Access 2020 Initiative. Будущее покажет. Однако предполагается, что российские научные организации и фонды вслед за китайскими более активно подключатся к данному движению. К этому есть предпосылки, так как после XIV Берлинской конференции статус Национальной контактной точки (National Contact Point / NCP) получил НЭИКОН и ожидается подписание OA2020 Initiative со стороны СПбГУ. К сожалению, до сих пор нет реакции на эту инициативу со стороны Министерства науки и высшего образования РФ и РАН, а без их поддержки российское научное сообщество не сможет внедрить прогрессивный План S Евросоюза. Необходимо формирование и развитие государственных проектов по переводу статей в открытый доступ» [90].

После публикации «Руководства по выполнению Плана S» началось оживленное обсуждение этого документа. Можно указать, например, на многочисленные материалы по проблемам выполнения Плана S, опубликованные на сайте Конфедерации репозиторий открытого доступа (COAR) [120], или рекомендации Альянса за честный свободный доступ [125].

Всего в адрес разработчиков плана S поступили сотни предложений и замечаний от различных организаций. Обсуждение продолжалось на Берлинской встрече по открытому доступу в ноябре 2019 г. Создан специальный сайт – форум для обсуждения проблем справедливого открытого доступа [139].

Будущее научных коммуникаций и изданий

Особый интерес для российских специалистов в области научных коммуникаций представляет опубликованный в январе 2019 г. доклад экспертов Евросоюза, который называется «Будущее научных изданий и научных коммуникаций» [127]. Дело в том, что в российской научной литературе практически не обсуждается образ будущего научных коммуникаций, который определяется тенденциями состояния научной инфосферы в настоящий момент. Соответственно, при отсутствии образа будущего и ясной дорожной карты для достижения образа будущего действия всех участников, начиная с финансирующих органов, носят противоречивый, зачастую хаотический характер. Добавим, что многочисленные концептуальные и программные документы, определяющие научно-технологическое развитие страны, практически полностью

обходят проблему формирования и модернизации научной инфосферы.

Европейские эксперты – авторы этого доклада, представляющие различных участников научной коммуникации, – как раз сосредоточили свое внимание на нынешнем состоянии инфосферы и процессах, происходящих в ней, с тем чтобы сформулировать рекомендации для всех участников научной коммуникации.

Авторы выделяют ключевые функции научной коммуникации, которые должны быть инвариантны к техническим, организационным и экономическим условиям:

- *регистрация*, чтобы установить, что работа была проведена отдельными лицами или группами исследователей в определенное время, и таким образом фиксировать их претензии на приоритет;

- *сертификация* (экспертиза), чтобы установить действительность выводов;

- *распространение*, чтобы сделать научные работы и их результаты доступными и видимыми;

- *сохранение*, чтобы гарантировать, что «записи науки» сохраняются и остаются доступными в течение длительного времени.

В то же время очевидны тенденции изменений в научной коммуникации, которые авторы сформулировали следующим образом:

- цифровизация коммуникаций;

- переход к открытой науке, включая разные гибридные модели, возможное сближение «зеленых» и «золотых» путей открытого доступа;

- дезагрегация многих коммуникативных функций, в том числе издательских;

- библиоразнообразие, рост роли первичных научных данных;

- преобразования в процессе рецензирования;

- появление жидких (текучих) публикаций, версияльность;

- трансформации публикаций в различные наборы услуг;

- активные инновации, реализуемые всеми участниками коммуникаций;

- важная роль методов оценки научных результатов для научной коммуникации; неудовлетворительность применяемых методов.

В докладе формулируются принципы, которым должны отвечать современные модели научной коммуникации:

- максимальная доступность;

- максимальное удобство использования;

- развитие форм результатов исследования;

- распределенная, открытая инфраструктура;
- равенство, разнообразие и знание как общественное благо;
- роль научных общественных структур;
- содействие качественным исследованиям и их целостности, экспертиза и сертификация;
- оценка научных результатов – разнообразная, количественная и качественная;
- содействие гибкости и инновациям;
- экономическая эффективность.

Каждый принцип достаточно подробно обсуждается, и описывается роль этих принципов для формирования будущей оптимальной системы коммуникации. Отдельная глава доклада посвящена недостаткам существующей системы научной коммуникации в отношении каждого из перечисленных принципов.

Авторы доклада выделяют категории участников научной коммуникации в зависимости от осуществляемых ими функций, интересов и мотивации:

- исследователи и их сообщества;
- университеты и институты;
- финансирующие организации и правительственные агентства;
- издатели и другие операторы коммуникации, поставщики услуг;
- другие потребители научной информации – практики, учащиеся, бизнес.

Для каждой из категорий участников коммуникации в докладе формулируются рекомендации, которые должны оптимизировать научную коммуникацию в будущем, в основном в цифровой среде. Приводим эти рекомендации.

«Исследователи и исследовательские сообщества должны:

1. При участии в оценке исследований, например при приеме на работу, продвижении по службе, при принятии решений о финансировании, сосредоточиваться на достоинствах и влиянии работы исследователя и воздерживаться от использования метрик – особенно метрик на основе журналов – в качестве прокси. В частности, используемые методы оценки должны учитывать рекомендации DORA и Лейденского манифеста.

2. Взять на себя ответственность за обеспечение того, чтобы все материалы исследований были общедоступными, т.е. доступными для обнаружения и многократного использования в соответствии с согласованными стандартами сообщества (включая справедливые принципы).

3. Повышать осведомленность авторов, рецензентов и членов групп, принимающих решения, и чувство ответственности за последствия выбора и действий.

4. Стремиться к сбалансированному и разнообразному представительству (с точки зрения пола, географии и карьеры) при поиске сотрудничества, организации конференций, созыве комитетов и назначении редакторов и рецензентов, а также создании сообществ, таких как научное общество.

5. Добиваться более широкого признания и признания экспертной оценки работы в качестве основных исследовательских задач. С этой целью поддерживать большую прозрачность, включая публикацию подписанных отчетов. Поддерживать лучшую подготовку экспертов и при экспертизе сосредоточить внимание на качестве исследования.

6. В случае сообществ исследователей, таких как научные общества, разрабатывать политику и практику, которые поддерживают способы научной коммуникации в соответствии с изложенным выше видением. Научным и другим исследовательским сообществам необходимо обучать своих членов и предупреждать о важности знаний и ответственности за их распространение, будь то формально, посредством публикации, или с помощью других средств.

Университеты и научно-исследовательские институты должны:

1. Разработать политику и практику для обеспечения того, чтобы все материалы исследований были общедоступными, т.е. доступными для обнаружения и многократного использования в соответствии с согласованными стандартами сообщества (включая справедливые принципы).

2. Поощрять и осуществлять рекомендации DORA и Лейденского манифеста для обеспечения того, чтобы при оценке исследований учитывался широкий круг научных результатов, включая исследовательские статьи, препринты, наборы данных, программное обеспечение, патенты и материалы (например, при принятии решений о найме и продвижении по службе).

3. При принятии решения о том, какую инфраструктуру использовать и поддерживать, следует выбирать платформы со свободным или открытым программным обеспечением, предлагать открытые данные через открытые лицензии и использовать открытые стандарты там, где это возможно. Действуя таким образом, мы

также укрепляем инициативы, направленные на облегчение научной коммуникации и издательской деятельности.

4. Стремиться к сбалансированному и разнообразному представительству (включая гендерные, географические и карьерные категории, но не ограничиваясь ими) при найме работников, поиске сотрудничества, при организации конференций, при созыве комитетов, а также при назначении редакторов и рецензентов и создании сообществ, например научных обществ.

5. В переговорах с поставщиками услуг отказываться от положений о неразглашении и включать положения, которые позволяют контролировать затраты и цены, а также соблюдение соглашений. Стремиться содействовать совместным действиям с другими учреждениями, например путем обмена данными о затратах и ценах через совместные инициативы (например, OpenAPC).

Финансирующим исследования и директивным органам следует:

1. Разработать политику – наряду с соответствующими механизмами финансирования – для обеспечения того, чтобы все финансы, выделяемые на научные исследования, были доступны для всех, везде, без каких-либо препятствий для доступа или ограничений на повторное использование результатов исследований.

2. При оценке научных работников обеспечить учет широкого круга результатов (научные публикации, а также данные, программное обеспечение, материалы и т.д.) и видов деятельности (наставничество, преподавание, подготовка обзоров и т.д.), а также соответствие процессов и критериев оценки исследовательской программе финансирующего органа и их прозрачность.

3. Разработать механизмы финансирования для поддержки развития открытых, взаимосвязанных и распределенных компонентов инфосферы научных публикаций и их поддержания в долгосрочной перспективе.

4. Контролировать влияние политики финансирования на разнообразие и качество исследований. В частности, органы управления должны гарантировать, что наблюдательные советы, комитеты, группы и т.д. представляют различные категории исследователей.

5. Согласовывать структуру затрат с другими субъектами научных коммуникаций, чтобы результаты исследования были доступными, открытыми и прозрачными.

Издатели и другие поставщики услуг должны:

1. Как можно скорее разработать и публично объявить о планах перехода к всеобъемлющему открытому доступу.

2. Разрабатывать, использовать и поддерживать совместимые инструменты (включая программное обеспечение с открытыми исходными кодами, где это возможно) и услуги не только для облегчения доступа к научным материалам и их повторного использования, но и для содействия новаторским инициативам новых участников.

3. Стремиться к сбалансированному разнообразию (включая гендерные, географические и квалификационные категории, но не ограничиваясь ими) среди авторов, рецензентов и редакторов, работающих с публикациями.

4. Повышать прозрачность экспертизы, например публикуя рецензии и ответы авторов вместе с опубликованными статьями.

5. Передавать гласности все сборы за публикацию (включая специальные цены и исключения) и предоставлять полное описание предоставляемых услуг, с тем чтобы создать прозрачный и экономически эффективный рынок, предназначенный для поддержки открытого обмена информацией и повторного использования всех научных материалов.

6. Экспериментировать с новыми подходами к оценке результатов исследований и их обмену и делиться результатами, с тем чтобы эти данные смогли помочь оптимизировать коммуникации в будущем.

Практикам, педагогам и другим социальным группам следует:

1. Организовывать и пропагандировать свободный доступ к результатам исследований, финансируемых государством, и право на их повторное использование.

2. Обратиться к финансирующим организациям, научно-исследовательским учреждениям и директивным органам с целью разработки новых каналов связи, новых форм совместного создания и совместного планирования исследований и новых форм финансирования в ответ на потребности, проблемы и вопросы, возникающие у населения в целом.

3. Искать возможности для внедрения результатов исследований, которые представляют интерес для социальных групп.

4. Продвигать темы / вопросы исследований, которые неправильно или недостаточно представлены (например, путем контактов с соответствующими исследователями, привлечения внимания других субъектов научной системы или мобилизации действий в организованных группах интересов)».

1.3. Выводы и рекомендации для России

На основе всех рассмотренных документов и инициатив, посвященных развитию научных коммуникаций в условиях открытой науки, можно сформулировать выводы, существенные для российских условий. Эти выводы для удобства мы разделим на три группы.

Общие выводы

- Ведущая роль в оптимизации научных коммуникаций принадлежит финансирующим организациям. В условиях России основные финансы для обеспечения научной инфосферы находятся у государства.

- Открытость – магистральный путь научной коммуникации, однако формы открытости, в частности экономика коммуникаций, должны быть исследованы дополнительно.

- Знание должно быть признано общественным благом.

- Законы об интеллектуальной собственности должны быть адаптированы к современным способам научной коммуникации.

- Результатом научной деятельности являются не только публикации, но и первичные научные данные в разных формах, прежде всего born digital.

- Инфраструктура научной коммуникации должна быть некоммерческой с исключением монополии; в российских условиях некоммерческая инфраструктура может быть распределена.

- Главные проблемы научной коммуникации – правовые, организационные и экономические, а не технологические; в основе необходимых реформ лежат социальные инновации.

- Одним из показателей качества коммуникации должна быть экономическая эффективность.

Выводы в отношении категорий участников

- Баланс интересов всех участников должен быть условием и целью модернизации коммуникаций.

- Функции участников научных коммуникаций должны быть разделены и оцениваться по-разному.

- Ведущая роль должна принадлежать научным обществам (в частности, они могут быть представлены в социальных сетях).

- Основные издательские функции могут быть дезагрегированы и распределены между другими участниками научной коммуникации.

- Для дезагрегированных функций в научных коммуникациях становятся особенно важными способы и методы коллаборации участников.

- Поставщики услуг в сфере коммуникации должны обеспечивать разнообразие услуг, активно развивать инновации в услугах.

Выводы относительно оценки научных результатов

- Оценка научных результатов лежит в центре организации научных коммуникаций.

- Для модернизации систем оценки следует опираться на рекомендации DORA и Лейденского манифеста.

- При оценке научных результатов важны: научность, прозрачность, новизна результатов, справедливость, отсутствие предвзятости и конфликта интересов при экспертизе.

- Рецензенты и эксперты должны быть включены в научный процесс.

- Импакт-фактор журнала не соответствует требованиям качественной оценки научных результатов.

- Для исследований, преследующих разные цели, должны использоваться различные критерии оценки научных результатов.

- Формальный показатель оценки не должен быть плановым (закон Гудхарта).

- Есть существенные различия в критериях оценки для разных наук, особенно для социальных и гуманитарных.

Конкретизация и уточнение этих выводов могут и должны проводиться на основе конкретных исследований процессов научной коммуникации. Часть этих исследований, относящихся к социальным и гуманитарным наукам, представлены в последующих главах настоящей книги.

Глава 2.

НАУЧНО-ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ В РОССИИ СЕГОДНЯ

2.1. Введение

До реформ 1990-х годов основную роль в организации научно-информационной деятельности в стране играла Государственная система научно-технической информации (ГСНТИ).

Создание ГСНТИ было, безусловно, одним из существенных достижений советской информатики. ГСНТИ отличалась практически полным общегосударственным охватом, включением всех существенных информационных потоков, разумным разделением труда между участниками и использованием современных на тот момент информационных технологий. Важную роль играла головная организация ГСНТИ – ВИНТИ РАН, которая организовывала многие общесистемные проекты. Основу ГСНТИ составляли общегосударственные центры научно-технической информации (НТИ), которые осуществляли обработку практически всего мирового потока информации и генерировали централизованные информационные продукты (прежде всего, реферативные издания и базы данных). Далее эти продукты распределялись по отраслям и регионам, пополняясь одновременно информационными материалами соответственно отраслевого и регионального происхождения.

В рамках ГСНТИ было реализовано достаточно много общих решений, некоторые из них, например создание Государственного рубрикатора научно-технической информации, комплекса стандартов СИБИД и др., сохранили свое значение до сих пор.

Однако с началом реформ начала 1990-х годов ГСНТИ как система стала очевидным образом деградировать, будучи несовместимой с той практикой управления наукой и народным хозяйством в целом, которая сложилась в ходе трансформаций последних 30 лет.

В настоящее время российская сфера НТИ представляет собой хаотическое множество информационных служб, лишенных

координации, и продуктов, дублирующих друг друга. В то же время российская сфера НТИ имеет множество лакун. К сожалению, даже Национальный проект «Наука», одно из направлений которого – развитие научной инфраструктуры, фактически не предлагает никакой разумной политики по оптимизации этой важной сферы научного обслуживания.

Острая дискуссия последних лет по поводу методов оценки эффективности научных учреждений, в том числе о роли наукометрии и экспертизы, практически не затронула проблему организации системы НТИ. В то же время, по мнению автора, именно на основе системы НТИ должна быть выстроена обратная связь между учеными и органами власти, реализованы способы оценки эффективности научных учреждений и, следовательно, сформирована адекватная система управления наукой.

В настоящей главе предлагается краткий обзор мирового опыта развития систем научной информации, а также делается попытка анализа современного состояния ресурсов НТИ России, по крайней мере той их части, которую можно использовать для изучения состояния науки в России, а также для решения практических задач управления научной сферы.

2.2. Краткий обзор мирового опыта

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 [22].

Существенное влияние на научную коммуникацию оказывают такие процессы, как цифровизация (дигитализация), переход к открытой науке, коллаборация и другие факторы, описанные нами выше.

Большое значение имеет появление специальных научных и научно-практических дисциплин, основным содержанием которых становится формирование и использование цифровых научных ресурсов. Примерами могут служить биоинформатика, геоинформатика, историческая информатика, компьютерная лингвистика, цифровая гуманитаристика и др.

Бурно развиваются сетевые информационные сервисы, основные виды которых перечислены ниже.

Прежде всего, это научные социальные сети, такие как Academia, ResearchGate, Social Science Research Network, Mendeley, CiteULike, myExperiment. Примерами российских научных социальных сетей являются «Соционет», «Ученые России», «Научная сеть», «Социальная научная сеть». Научные социальные сети предоставляют многочисленные сервисы для обмена разнообразной информацией, размещения результатов исследований, их рецензирования и комментирования, поиску вакансий и др.

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

- возможность быстрого доступа 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;

- создание 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.);

- обеспечение сохранности ценных научных коллекций для будущих поколений ученых и предотвращение утраты научной информации;

- обеспечение возможностей для 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), которая основана Национальным научным фондом (NSF) в 2000 г. с целью обеспечения доступа к ресурсам и сервисам высокого качества во многих областях науки, технологий и техники. NSDL агрегирует ресурсы из цифровых библиотек и проектов, поддерживаемых NSF, а также формирует собственные обзоры веб-сайтов 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.

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

Основным недостатком как зарубежных, так и отечественных электронных библиотек является примитивный поисковый интерфейс, построенный по аналогии с поисковыми машинами Интернета типа Google и не обеспечивающий приемлемую точность поиска.

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

Google Scholar (Академия Google) – поисковая система научной литературы. Используя единую форму запроса, сбор информации о свободно скачиваемых научных публикациях, Академия Google позволяет найти исследование, наиболее точно соответствующее запросу.

Scholar.ru – поисковая система научных публикаций. Проект был создан для упрощения поиска документов научной тематики на русском языке, в первую очередь – созданных в России. Основная цель проекта – 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 самих публикациях (аннотация, авторы и т.д.). Интерфейс на русском языке.

Science Research Portal – научная поисковая система, осуществляющая полнотекстовый поиск в журналах многих крупных научных издательств, таких как Elsevier, Highwire, IEEE, Nature и др.

CiteSeer Publications ResearchIndex – научная поисковая система, индексирующая статьи в PostScript и PDF-формате с научных веб-сайтов.

Ingenta – библиографическая база данных универсального профиля, включающая описания статей из англоязычных журналов и сборников. Объем базы составляет более 13 млн записей.

Особую нишу в информационном пространстве занимают мультидисциплинарные коммерческие базы данных цитирования Web of Sciences и Scopus, включающие миллионы записей из на-

учных журналов и сборников и предоставляющие пользователям достаточно мощный интерфейс, поддерживающий поиск и навигацию не только по метаданным собственно статей, но 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* – Международная информационная система по сельскохозяйственным наукам и технологиям;

- *GeoRef* – библиографическая база данных, охватывающая мировую литературу по наукам о Земле;

- *Chemical Abstracts Service* – химическая реферативная служба, издающая реферативный журнал *Chemical Abstracts* и поддерживающая крупнейшую химическую информационно-поисковую систему *SciFinder*.

Этот перечень можно продолжать, но приведенные примеры уже дают достаточное представление о составе и функциях международных информационных научных ресурсов.

За рубежом достаточно широкое распространение получили интеграторы научных информационных ресурсов – организации, приобретающие права на их распространение у владельцев. Такие организации разрабатывают единый поисковый интерфейс для доступа к различным ресурсам и предоставляют коммерческий доступ к ним через одну точку входа.

Среди основных игроков на этом поле можно назвать компанию *STN International* [144] (она обеспечивает доступ к почти 200 базам данных по научно-технической информации и к более чем 100 коллекциям научных журналов). Переведенный на русский язык список доступных через нее ресурсов представлен в [27]. Еще один заметный игрок – компания *EBSCO* [122], агрегирующая коллекции полнотекстовых электронных журналов и книг.

В качестве универсальных источников справочного материала, представленного в Интернете, наиболее популярны как за рубежом, так и в России такие ресурсы, как *Wikipedia* и *Wikidata*. Они, без сомнения, играют значительную роль в жизни современного информационного общества, но возможность их использования

как источника серьезных научных знаний вызывает активные дискуссии. Они особенно обострились в связи с проектом создания Национального энциклопедического портала, который должен в идеале заменить Википедию. Подробнее см. об этом в главах 4 и 5.

2.3. Общая характеристика отечественных ресурсов НТИ

Представляется очевидным, что модернизация и развитие научной коммуникации должны основываться на существующих информационных ресурсах, как традиционных, так и новых, поскольку именно в этих ресурсах зафиксирован опыт многолетней информационной деятельности по сбору, обработке научной информации и представлению знаний, извлеченных из этой информации. Поэтому мониторинг, учет и оценка качества отечественных информационных ресурсов являются актуальными задачами современной информатики. Следует добавить, что ресурсы НТИ стали играть важную роль в оценке эффективности научных и образовательных организаций (инфометрика, вебометрика, альтметрика).

В стране создается множество различных ресурсов научной информации, в том числе полнотекстовых ЭБ, энциклопедических, словарных, аналитических, фактографических БД, тематических, проблемных, видовых автоматизированных информационных систем (АИС). Так, каталог «Наука в Рунете» [51] содержит около 10 тыс. ссылок, каталог «Рамблер / Топ-100 / Наука» – 2,3 тыс. ссылок, Навигатор информационных ресурсов по общественным наукам [60] – свыше 4 тыс. ссылок.

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, ЦИТИС, Президентская библиотека им. Б.Н. Ельцина, Архив РАН, или таких популярных, как Википедия или КиберЛенинка. Однако много и небольших ресурсов, содержащих тем не менее уникальную научную информацию.

Число российских научных периодических изданий, индексируемых в РИНЦ, – 5,3 тыс., а общее количество российских журналов – свыше 14 тыс.

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

Быстрый рост количества научно-информационных продуктов и ресурсов потребовал 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 информацию в электронной форме. Перечислим некоторые из них:

- 1) Национальная электронная библиотека [68];
- 2) Национальный энциклопедический портал на основе Большой российской энциклопедии (см. главу 4);
- 3) Национальный портал открытого образования [69];
- 4) Электронная библиотека диссертаций [114];
- 5) Интегратор научной периодики: Научная электронная библиотека и Российский индекс научного цитирования [64];
- 6) Интеграторы библиографической информации: ЛИБНЕТ [71], ЭКБСОН [112], АРБИКОН [25];
- 7) Интеграторы материалов научных конференций, например [79];
- 8) Интегратор научных ресурсов открытого доступа Кибер-Ленинка [65];
- 9) Репозитории научно-образовательной информации: НОРА [70], СОЦИОНЕТ [102], Федеральная резервная система банков знаний [107];
- 10) Государственный каталог Музейного фонда РФ [37];
- 11) Центральный фондовый каталог Архивного фонда РФ [111];
- 12) Электронная библиотека «Научное наследие России» [66].

Следует, однако, отметить, что многие проекты, предполагавшие интеграцию, систематизацию и инвентаризацию научно-образовательных информационных ресурсов, не были реализованы или откровенно провалились, не будучи вписаны в реальное информационное пространство. Среди них наиболее известен проект «Карта российской науки». Также можно вспомнить такие проекты, как Единая информационно-образовательная среда, Центральная библиотека образовательных ресурсов, Единая система информационных ресурсов РАН и др. К сожалению, публичных выводов из неудачи этих проектов не было сделано.

Более подробные сведения о состоянии научных информационных ресурсов по социогуманитарной тематике представлены в главе 3 настоящей монографии.

2.4. Федеральные органы НТИ

Один из парадоксов современной ситуации в сфере НТИ состоит в том, что ее нормативная база – формально действующее Постановление Правительства РФ от 24 июля 1997 г. № 950 «Об утверждении Положения о государственной системе научно-технической информации» [87] – фактически описывает реалии, оставшиеся в прошлом. Созданная в советское время Государственная система научно-технической информации (ГСНТИ) в начале XXI в., к сожалению, утратила свою системную целостность и – в значительной степени – функциональность. В задачи настоящей книги не входит обсуждение причин, объективных и субъективных факторов, повлиявших на системную деградацию ГСНТИ. Подробнее позиция автора по этому вопросу изложена в работах [7; 22].

При подготовке данной книги специально были рассмотрены современное состояние и информационные ресурсы тех организаций, которые Постановлением Правительства 1997 г. № 950 были отнесены к федеральным органам научно-технической информации (далее – федеральные органы НТИ, ФОНТИ). Приведем перечень ФОНТИ с указанием их функций с учетом изменений в этом постановлении, сделанных в 2010, 2014 и 2018 гг. [125].

1. Центр информационных технологий и систем органов исполнительной власти (ЦИТИС) – по ведущимся в стране и законченным открытым научно-исследовательским и опытно-конструкторским работам, защищенным диссертациям на соискание ученых степеней, алгоритмам и программам.

2. Федеральное государственное унитарное предприятие «Научно-технический центр оборонного комплекса “Компас”» – по научно-исследовательским, опытно-конструкторским работам и результатам научно-технической деятельности оборонного комплекса.

3. Федеральное государственное учреждение «Российское энергетическое агентство» (РЭА) Министерства энергетики Российской Федерации – по использованию результатов научно-технической деятельности предприятий и организаций, а также организации обмена этой информацией между регионами.

4. Всероссийский институт научной и технической информации (ВИНИТИ) Российской академии наук и Министерства науки и технологий Российской Федерации – по естественным и техническим наукам.

5. Государственная публичная научно-техническая библиотека Министерства науки и технологий Российской Федерации, (ГПНТБ России) – по естественным и техническим наукам.

6. Библиотека Российской академии наук (БАН) – по естественным и техническим наукам.

7. Библиотека по естественным наукам Российской академии наук (БЕН РАН) – по естественным и техническим наукам.

8. Государственная публичная научно-техническая библиотека Сибирского отделения Российской академии наук (ГПНТБ СО РАН) – по естественным и техническим наукам.

9. Государственная центральная научная медицинская библиотека Министерства здравоохранения Российской Федерации (ГЦНМБ) – по медицине и здравоохранению.

10. Всероссийский научно-исследовательский институт информации и технико-экономических исследований агропромышленного комплекса Министерства сельского хозяйства и продовольствия Российской Федерации (ВНИИТЭИСХ) – по сельскому хозяйству.

11. Центральная научная сельскохозяйственная библиотека Российской академии сельскохозяйственных наук (ЦНСХБ) – по сельскому хозяйству.

12. Всероссийский научно-исследовательский институт растениеводства имени Н.И. Вавилова (ВИР) Российской академии сельскохозяйственных наук – по национальному генофонду сельскохозяйственных растений.

13. Всероссийский научно-исследовательский институт племенного дела Министерства сельского хозяйства и продовольствия Российской Федерации (ВНИИплем) – по национальному генофонду сельскохозяйственных животных.

14. Государственный научно-исследовательский институт озерного и речного рыбного хозяйства (ГОСНИОРХ) Министерства сельского хозяйства и продовольствия Российской Федерации – по национальному генофонду рыб.

15. Всероссийский государственный научно-исследовательский институт контроля, стандартизации и сертификации ветеринарных препаратов (ВГНКИ) Министерства сельского хозяйства и продовольствия Российской Федерации – по национальному фонду ветеринарных препаратов, штаммов микроорганизмов.

16. Всероссийский научно-исследовательский институт научно-технического прогресса и информации в строительстве (ООО ВНИИТПИ) – по строительству и архитектуре.

17. Центральная научно-техническая библиотека по строительству и архитектуре (ЦБНТБ СИА) Государственного комитета Российской Федерации по жилищной и строительной политике – по строительству и архитектуре.

18. Институт научной информации по общественным наукам Российской академии наук (ИНИОН РАН) – по общественным наукам.

19. Российский федеральный геологический фонд (Росгеол-фонд) – по геологии и минерально-сырьевым ресурсам.

20. Всероссийский научно-исследовательский институт экономики минерального сырья и недропользования (ВИЭМС) – по геологии и минерально-сырьевым ресурсам.

21. Всероссийская геологическая библиотека (ВГБ) Министерства природных ресурсов Российской Федерации – по геологии и минерально-сырьевым ресурсам.

22. Всероссийский научно-исследовательский институт гидрометеорологической информации – Мировой центр данных (ВНИИГМИ-МЦД) Федеральной службы по гидрометеорологии и мониторингу окружающей среды – по метеорологии и другим смежным областям, включая климатологию, агрометеорологию, гидрологию, геофизику, океанологию, мониторинг состояния и загрязнения окружающей среды.

23. Федеральный фонд государственных стандартов, общероссийских классификаторов технико-экономической информации, международных (региональных) стандартов, правил, норм и рекомендаций по стандартизации, национальных стандартов зарубежных стран (Стандартинформ) Государственного комитета Российской Федерации по стандартизации, метрологии и сертификации – по нормативным документам по стандартизации, метрологии и сертификации.

24. Федеральный институт промышленной собственности (ФИПС) – по патентной документации, относящейся к изобретениям, полезным моделям, промышленным образцам, товарным знакам, знакам обслуживания и наименованиям мест происхождения товаров, а также по документации, относящейся к зарегистрированным программам для ЭВМ, базам данных и топологиям интегральных микросхем.

25. Информационно-издательский центр (ИНИЦ ПАТЕНТ) Российского агентства по патентам и товарным знакам – по патентной документации, относящейся к изобретениям, полезным моделям, промышленным образцам, товарным знакам, знакам

обслуживания и наименованиям мест происхождения товаров, а также по документации, относящейся к зарегистрированным программам для ЭВМ, базам данных и топологиям интегральных микросхем.

26. Федеральное государственное унитарное предприятие «Информационное телеграфное агентство России» (ИТАР-ТАСС) – по опубликованным в Российской Федерации произведениям печати и государственной библиографии.

27. Всероссийский научно-исследовательский институт проблем машиностроения (РоснииПМ) Министерства науки и технологий Российской Федерации – по отечественным и иностранным каталогам на промышленную продукцию.

28. Всероссийский научно-исследовательский институт стандартизации (ВНИИС) Государственного комитета Российской Федерации по стандартизации, метрологии и сертификации – по отечественным и иностранным каталогам на промышленную продукцию.

29. Всероссийский научно-исследовательский центр стандартизации, информации и сертификации сырья, материалов и веществ (ВНИЦСМВ) Государственного комитета Российской Федерации по стандартизации, метрологии и сертификации – по сведениям о материалах и веществах.

30. Российский государственный архив научно-технической документации (РГА НТД) Федеральной архивной службы России – по документам научно-исследовательских, проектных, конструкторских, технологических организаций и предприятий федерального значения, переданных на постоянное хранение.

31. Научно-технический центр «Информрегистр» – по электронным изданиям.

32. Институт промышленного развития (Информэлектро) – по технико-экономической информации и результатам технико-экономических исследований в гражданских отраслях промышленности.

Заметим, что в данное постановление (вплоть до 2018 г.) трижды вносились изменения, связанные с реорганизацией отдельных организаций (Росинформресурс, Всероссийская книжная палата, Всероссийский институт межотраслевой информации). В то же время произошедшие за этот период радикальные реорганизации Минобрнауки, РАН, отраслевых академий, других учреждений в постановлении не отражены.

Однако более важно, что данное постановление фактически не выполняется в своих наиболее существенных положениях.

Прежде всего, исчезло централизованное управление государственными ресурсами и системой НТИ. Фактическое состояние государственных ресурсов НТИ не отслеживается, статистики по ним не ведется. В Минобрнауки, которое должно выполнять функции координации ГСНТИ, давно ликвидировано подразделение, которое курировало ГСНТИ. Не функционирует и предусмотренный Постановлением № 950 совещательный орган ГСНТИ. Отсутствуют общесистемные проекты, направленные на поддержку и развитие ресурсов НТИ. В Академии наук также проблема НТИ практически не обсуждается¹.

Следует при этом учитывать, что вопрос о статусе и роли государственных ресурсов НТИ, который является центральным для указанного постановления, в существующей системе научного законодательства остается нерешенным. Представляется очень важным, чтобы нормы, регулирующие создание, использование и сохранность государственных информационных ресурсов НТИ, вошли в новый Закон «О науке», который в настоящее время разрабатывается.

Современное состояние Федеральных органов НТИ (ФОНТИ) можно представить в виде таблицы (табл. 1).

В итоге в 32 ФОНТИ, перечисленных в Постановлении № 950, произошли следующие изменения статуса:

- ликвидированы – 4;
- акционированы – 3;
- изменили подчинение – 10;
- объединены с другой организацией – 6;
- реорганизованы с внесением изменений в постановление – 3;
- не изменили организационный статус – 6.

С точки зрения функционирования в качестве ФОНТИ, т.е. генерации общедоступных ресурсов НТИ:

- функционируют – 17;
- существуют, но не выполняют функций ФОНТИ – 10;
- не существуют – 4;
- выполняют функции частично – 1;
- нет сведений – 2.

¹ Исключение составляют проблемы научной периодики.

Таблица 1

Изменение статуса и функций ФОНТИ

Краткое наименование ФОНТИ	Изменение статуса	Функции ФОНТИ
ЦИТИС	Новое подчинение	Функции выполняются
Компас	Ликвидирован ВИМИ, отражено в постановлении	Функции ФОНТИ не выполняются
РЭА	Росинформресурс переподчинен и реорганизован, это отражено в постановлении	Функции выполняются частично – в некоторых ЦНТИ, входивших в Росинформресурс
ВИНИТИ	Новое подчинение	Функции выполняются
ГПНТБ России	Не изменился	Функции выполняются
БАН	Новое подчинение	Функции выполняются
БЕН РАН	Новое подчинение	Функции выполняются
ГПНТБ СО РАН	Новое подчинение	Функции выполняются
ГЦНМБ	Включена в состав МГМУ	Функции выполняются
ВНИИТЭИСХ	Ликвидирован	
ЦНСХБ	Объединена с ВНИИТЭИСХ	Функции выполняются
ВИР	Новое подчинение	Функции не выполняются
ВНИИПлем	Не изменился	Функции ФОНТИ не выполняются
ГОСНИОРХ	Не изменился	Функции ФОНТИ не выполняются
ВГНКИ	Не изменился	Функции ФОНТИ не выполняются
ВНИИТПИ	Акционирован	Функции ФОНТИ не выполняются
ЦБНТБ СИА	Новое подчинение	Функции ФОНТИ не выполняются
ИНИОН РАН	Новое подчинение	Функции выполняются
Росгеолфонд	Не изменился	Функции выполняются
ВИЭМС	Ликвидирован	
ВГБ	Включена в состав ВСЕГЕИ	Функции выполняются
ВНИИГМИ-МЦД	Не изменился	Функции выполняются
Стандартинформ	Объединен с издательством, переименован	Функции выполняются
ФИПС	Объединен с ВПТБ	Функции выполняются
ИНИЦ ПАТЕНТ	Акционирован	Функции ФОНТИ не выполняются
ИТАР-ТАСС	ВКП вошла в состав ИТАР-ТАСС, отражено в постановлении	Функции выполняются
РОСНИИ ПМ	Ликвидирован	
ВНИИС	Ликвидирован	
ВНИЦСМВ	Не изменился	Сведений нет
РГА НТД	Не изменился	Функции ФОНТИ не выполняются
Информрегистр	Новое подчинение	Функции выполняются
Информэлектро	Акционирован	Сведений нет

Если говорить о содержательной экспертной оценке деятельности ФОНТИ, то среди 17 действующих ФОНТИ можно выделить несколько наиболее продвинутых, в которых происходит реальное наращивание количества, качества и разнообразия предоставляемых пользователям ресурсов и сервисов. Среди библио-

тек можно отметить ГЦНМБ, ГПНТБ России, ГПНТБ СО РАН, ЦНСХБ. Среди информационных центров устойчиво работают ФИПС, ЦИТИС, ВНИИГМИ-МЦД.

Беспорным лидером среди отечественных ФОНТИ, соответствующим современному уровню развития инфосферы, является комплекс геологических информационных служб. Этот комплекс включает Росгеолфонд с сетью территориальных геологических фондов и ВСЕГЕИ, с входящими в его состав Всероссийской геологической библиотекой и Центральным научно-исследовательским геологоразведочным музеем имени академика Ф.Н. Чернышева. Всего Росгеолфонд и ВСЕГЕИ поддерживают 32 информационные системы, а также ресурсы, вошедшие в созданный нами каталог [61]. Кроме того, для информационного обслуживания отрасли действует большое количество других информационных систем, поддерживаемых органом власти отрасли (Роснедра) и предприятиями ведомства, а также учреждениями Минобрнауки и университетами.

Справедливости ради стоит отметить, что геологическая отрасль располагала самой развитой информационной системой в стране еще в советское время.

При оценке деятельности ФОНТИ, отраженных в упомянутом постановлении, следует, конечно, учитывать, что с 1997 г. возникло множество других информационных систем, служб и ресурсов, обслуживающих науку и управление наукой, в организациях различных форм собственности, не отраженных в постановлении.

Таким образом, состав и функции органов НТИ, определенных Постановлением № 950, существенно изменились. Соответственно, указанное постановление требует радикального пересмотра. Конечно, речь должна идти не только о пересмотре номенклатуры ФОНТИ. Должна быть определена структура современной научной инфраструктуры и определены роли, функции и статус всех организаций, работающих в сфере НТИ.

2.5. Исследование отечественных ресурсов по управлению наукой

В 2018 г. автором было проведено исследование отечественных информационных ресурсов, которые используются или могут быть использованы для управления наукой.

Исходной базой для проведения данного исследования послужили результаты мониторинга информационных ресурсов Россий-

ской академии наук в области общественных наук, проведенного автором ранее и подробно описанного в главе 3 и в ряде публикаций. Результатом данного мониторинга стало создание информационной системы, которая получила название Навигатор информационных ресурсов по общественным наукам (НИРОН) [60] (см. главу 3).

Соответственно, исходным материалом для проведения мониторинга и инвентаризации российских информационных ресурсов по управлению наукой послужили описания ресурсов, относящихся к этой проблеме, имевшиеся в учреждениях РАН. Критерием для отнесения описания ресурсов к этой проблеме был выбран тематический принцип: к ресурсам по управлению наукой были отнесены все ресурсы, имевшие индекс ГРНТИ *12 Науковедение*, или подчиненные индексы 2-го и 3-го уровней.

Начальная версия каталога была создана автоматически. Он получил название Науковед. Далее полученный каталог был существенно пополнен и расширен за счет включения в него ресурсов НТИ различных институций: органов НТИ, органов власти, университетов, корпораций, а также других организаций – владельцев профильных информационных ресурсов. В настоящее время Науковед включает свыше 1 тыс. ресурсов. Этот каталог размещен в Интернете [61].

Для создания нового каталога были использованы логическая структура НИРОН и разработанная типология информационных ресурсов, описанная в главе 3. Впрочем, в логическую структуру нового каталога, а также в типологию информационных ресурсов были внесены существенные изменения.

Таким образом, был создан специализированный каталог информационных ресурсов по науковедению и управлению наукой, включая сферу НТИ.

2.6. Инвентаризация владельцев ресурсов НТИ

Основной задачей исследования было выявление институций (органов власти, учреждений, структурных подразделений, советов и комиссий), которые владеют информационными ресурсами, связанными с управлением наукой.

Вначале приведем некоторую статистику по числу научных организаций страны (данные в основном за 2017 г.) [78].

Всего организаций, выполняющих НИР в России, – 3944. В том числе:

- 1) по типам организаций:

- научно-исследовательские организации – 1577;
- образовательные организации высшего образования – 970;
- 2) по секторам:
 - государственный сектор – 1493;
 - предпринимательский сектор – 1292;
 - сектор высшего образования – 1038;
 - сектор некоммерческих организаций – 121;
- 3) организации, подведомственные Минобрнауки;
 - научные учреждения – 466, в том числе:
 - бывш. РАН – 318;
 - бывш. РАМН – 31;
 - бывш. РАСХН – 97;
 - бывш. РАО – 7;
 - бывш. Минобрнауки – 13;
 - университеты – 252.

Разумеется, тотальная проверка всех организаций, занимающихся наукой в России, является слишком объемной, но не очень осмысленной работой. Очевидно, что основная доля ресурсов для управления наукой должна была находиться не столько в научных учреждениях, сколько в управляющих структурах. Поэтому при проведении инвентаризации было принято решение выполнить сплошную проверку управляющих структур и специализированных информационных организаций и выборочную – научных и образовательных учреждений. Результаты проведения инвентаризации представлены в табл. 2.

Ниже следуют комментарии к результатам этой работы.

Под академическими управляющими структурами понимаются Президиум РАН, отраслевые и региональные отделения, региональные научные центры, а также республиканские академии. Все они ведут учет научных учреждений, ученых, результатов исследований, научных мероприятий и др. Такими же ресурсами располагают некоторые научные учреждения.

Среди федеральных органов исполнительной власти (ФОИВ), к которым относятся Правительство РФ, министерства, агентства, службы, были выделены те, которые имеют организационные структуры (департаменты, управления, советы), ответственные за управление наукой и / или инновациями. Всего таких федеральных органов власти было выделено 32, т.е. менее половины от общего числа ФОИВ. Непонятно, по каким причинам некоторые ФОИВ, имеющие подведомственные научные организации, т.е. финансирующие научную и / или инновационную деятель-

ность, не выделяют явным образом структуры, ответственные за управление наукой. По крайней мере, таких структур в составе этих ФОИВ обнаружить не удалось.

Возможно, для составления списка ФОИВ, осуществляющих управление наукой, более корректно было бы использовать другой критерий, а именно – наличие или отсутствие в бюджете ФОИВ статьи «Финансирование НИОКР». Однако такой метод не позволит выделить информационные ресурсы, связанные с управлением наукой в данном ФОИВ. Кроме того, затруднительно получить полный список бюджетополучателей по статье «Наука и инновации» из бюджетов всех уровней.

В каталог включены не все органы власти субъектов РФ, а только те, в составе которых имеются управляющие структуры данного профиля. Всего их оказалось 34 из 85 субъектов РФ. Заметим, что во многих субъектах имеются министерство или департамент образования и науки, но все ресурсы данного органа власти относятся к образованию. Такие органы власти в каталог не включались.

В каталог включены управляющие структуры всех наукоградов. Однако выяснилось, что почти у всех этих структур отсутствуют ресурсы, отражающие результаты научных исследований, проводимых в данном наукограде.

Из числа университетов анализировались только относящиеся к категориям «исследовательские» или «федеральные» (а также, конечно, МГУ и СПбГУ, имеющие особый статус). Все проанализированные университеты имеют управленческие структуры, ответственные за науку и инновации, и соответствующие страницы или разделы сайта.

Государственные корпорации включены в каталог, если они входят в перечень организаций, на которые распространяется государственная программа инноваций, и если в их структуре есть подразделения, ответственные за науку и инновации.

Заметим, что государственные научные центры (ГНЦ) делегировали информирование о результатах своей научной деятельности общественной организации – Ассоциации государственных научных центров «Наука», – поэтому в число владельцев информационных ресурсов по управлению наукой не были включены.

Информационные организации – смешанная категория владельцев, включающая информационные центры, библиотеки, архивы, издательства, коммерческие информационные организации, а также условные группы владельцев информационных ресурсов, например создатели каталогов научных БД. Заметим, что большая группа ин-

формационных организаций – региональные центры научно-технической информации, ранее входившие в объединение «Росинформресурс», – не включена в число владельцев, поскольку они теперь не являются юридическими лицами (кроме ЦНТИ Татарстана).

Эти центры представлены в каталоге как филиалы, т.е. подразделения головного ведомства – Российского энергетического агентства.

Федеральные органы НТИ, перечисленные в Постановлении № 950, выделены в отдельную категорию. Состояние этих учреждений описано выше.

Научные учреждения включены в каталог на основе экспертных сведений, подтверждающих, что они владеют интересующими нас информационными ресурсами. В частности, это касалось всех учреждений Минобрнауки, унаследованных от прежнего министерства.

Вообще дать исчерпывающий перечень организаций, которые могли бы в настоящее время быть отнесены к категории «органы НТИ», довольно сложно. Большинство тех организаций, которые входили в эту категорию до реформ 1990-х годов, сменили профиль или исчезли. Зато возникло очень много организаций различных форм собственности, владеющих теми или иными ресурсами НТИ.

Общее число включенных в каталог владельцев ресурсов, которые, по оценке составителя, можно использовать для управления наукой, – 440 организаций.

Таблица 2

**Владельцы ИР по управлению наукой,
включенные в каталог Наукоед**

Категория владельцев	Число владельцев
Академические управляющие структуры	45
Академические научные учреждения	80
Федеральные органы исполнительной власти	32
Органы власти субъектов РФ	34
Федеральные органы НТИ	32
Университеты	50
Корпорации	52
Информационные организации	54
Научно-исследовательские институты, не входящие в РАН	18
Фонды	17
Общественные организации	13
Другие категории владельцев	13
Всего в перечне владельцев	440

2.7. Инвентаризация ресурсов

Центральной проблемой инвентаризации ресурсов, которые могут быть использованы для управления наукой и науковедческих исследований в целом, является вопрос о том, какие типы ресурсов могут использоваться для этой роли. Какие именно объекты являются результатом научной деятельности, насколько эти результаты могут быть метризованы, т.е. выражены количественными наукометрическими показателями, – эти вопросы продолжают оставаться дискуссионными.

Хотя некоторые наукометрические показатели являются официальными, т.е. используются органами управления для оценки деятельности научных учреждений, научное сообщество не вполне удовлетворено показателями, используемыми в существующей системе управления наукой.

Например, публикации признаются результатами научно-технической деятельности (РНТД), только если они изданы в определенных изданиях, в основном индексируемых в мировых наукометрических базах данных. В то же время размещение тех же сведений в отечественных базах данных или в социальных сетях в качестве РНТД не учитывается. Такие важные показатели, как видимость, считываемость и цитирование на интернет-порталах (вебметрика, альтметрика), не используются. Сведения о публикациях, включенных в библиотечные или иные каталоги, также не учитываются.

Часть РНТД могут быть зарегистрированы в виде объектов интеллектуальной собственности, но также далеко не все. Многие типы научных продуктов, например электронные библиотеки, базы и банки данных, информационные системы, 3D-объекты, интерактивные продукты, также не учитываются как РНТД, если они не зарегистрированы в ФИПС как программные средства или базы данных.

В целом автор придерживается точки зрения, что номенклатура информационных объектов, которые содержат результаты научных исследований, должна быть существенно расширена, прежде всего за счет новых цифровых объектов, сведения о которых распространяются через Интернет. Такого же мнения придерживается мировая общественность. Обзор мировых научных документов, посвященных этой проблеме, содержится в главе 1.

Подготовка научных кадров, в том числе аспирантура и защита диссертаций, также учитывается отдельно от собственно научных результатов, что автору представляется неверным.

С учетом всех этих соображений к ресурсам, которые могут использоваться для управления наукой, автор предлагает отнести ресурсы, содержащие следующие сведения:

- объекты интеллектуальной собственности;
- инновационные предложения, проекты;
- электронные библиотеки научных публикаций, библиографические издания, каталоги, БД;
- труды конференций по науковедению и управлению наукой;
- диссертации, отчеты, другие неопубликованные материалы;
- статистика науки;
- статистика печати;
- статистика высшего профессионального образования (обучение в аспирантуре и защита диссертаций);
- географические системы (ГИС) по результатам НИОКР;
- справочники научных и инновационных институций;
- индексы цитирования;
- вебметрические и альтметрические ресурсы;
- антиплагиатные ресурсы;
- персональные ресурсы – указатели ученых (выборочно);
- перечни грантов, проектов, экспедиций;
- программы НИР;
- архивные фонды, путеводители, каталоги по истории науки (выборочно).

Особую проблему представляет периодика, которая пока является важнейшим видом научной коммуникации. Суть проблемы – в сложности определения качества периодических изданий, которые владельцы относят к научным, критерии этого качества. Решению этой проблемы посвящено множество трудов, конференций, различных форумов. Специально проблемой качества российской научной периодики занимается общественная организация «Ассоциация научных издателей и редакторов» [24], на сайте которой можно найти много материалов, относящихся к этой теме.

В данной книге автор ограничился задачей определения перечня российских периодических изданий, которые освещают проблемы, связанные с управлением наукой. Наиболее полный перечень российских научных периодических научных изданий имеется в базе данных Научной электронной библиотеки eLIBRARY. В данном исследовании к самостоятельным информационным ресурсам

для управления наукой были отнесены издания, которым в этой системе присвоена рубрика ГРНТИ 12. *Науковедение*.

Как было отмечено выше, в каталог были включены все ресурсы Федеральных органов НТИ, хотя некоторые из них не относились к вышеперечисленным категориям.

Сложным методическим вопросом был вопрос о включении в число ресурсов управления наукой ресурсов, содержащих данные по истории науки. Из общетеоретических соображений эти ресурсы были включены в каталог, поскольку они являются, бесспорно, ресурсами науковедения, хотя для задач управления наукой эти ресурсы не используются.

Как уже было отмечено, логическая структура и типология каталога Науковед была унаследована от типологии ресурсов, разработанной для каталога НИРОН. В табл. 3 представлен количественный состав информационных ресурсов, отраженных в каталоге, на основе этой типологии.

Таблица 3

Количество ресурсов, отраженных в каталоге Науковед

Тип ресурса	Количество ресурсов данного типа
Библиотеки	7
Архивы	22
Музеи	33
Каталоги	19
Электронные коллекции и библиотеки	53
Информационные системы	149
Справочники, энциклопедии	124
Персональные ресурсы	74
Лингвистические ресурсы	12
Периодика	59
Библиографии	30
Мероприятия	33
Неопубликованные материалы	173
Медиаресурсы	22
Прочие интернет-ресурсы	212
Всего	1012

2.8. Выводы

Хотя представленная в каталоге Наукоед и в данном разделе монографии картина российского научно-информационного ландшафта не является исчерпывающей, она позволяет оценить состояние и тенденции развития соответствующих информационных ресурсов и сделать некоторые выводы.

Информационные ресурсы, отражающие результаты научных исследований, отличаются разнообразием, но в то же время имеют значительные лакуны.

В пределах каждого типа ресурсов, особенно в новых цифровых, таких как электронные библиотеки, информационные системы, аналитические БД, веб-сайты, наблюдается существенное различие состава и структуры информационных объектов, что затрудняет их сравнение. Нужны стандарты, предлагающие эталонные решения для новых видов цифровых научных объектов.

Поскольку номенклатура РНТД, используемая в официальных наукометрических показателях, плохо учитывает цифровые объекты, используемые в научной коммуникации, анализ сетевых научных ресурсов позволяет более полно оценить ситуацию, чем официальная статистика науки.

Многие ресурсы откровенно дублируются. Во избежание дублирования целесообразно организовать ведение ресурсов, общих для научного сообщества и отражающих определенные категории РНТД, причем эти ресурсы должны пополняться и редактироваться на основе коллаборации. Это среди прочего может существенно сократить затраты ученых и научных учреждений на оформление разного рода отчетности.

Приемлемым решением была бы организация официальной системы мониторинга, учета и распространения результатов научных исследований, представленных в цифровой форме. Многие методические и технологические решения для создания такой системы содержатся в отечественной платформе Соционет [102].

Представляется, что правовые, организационные и технические разработки, направленные на оптимизацию системы НТИ в рамках современной научной инфраструктуры и обеспечивающие формирование необходимых данных для принятия обоснованных решений по управлению наукой, должны быть реализованы в рамках Национального проекта «Наука», где имеется соответствующий раздел.

Глава 3.

МОНИТОРИНГ ИНФОРМАЦИОННЫХ РЕСУРСОВ СОЦИАЛЬНО-ГУМАНИТАРНЫХ НАУК

3.1. Цели и задачи мониторинга информационных ресурсов

Тенденции развития научных коммуникаций, рассмотренные выше, определяют постановку задачи исследования состояния информационных ресурсов – основной формы представления научной информации. Существо исследования – мониторинг информационных ресурсов академических организаций социального и гуманитарного профиля. Предмет и границы исследования определялись аффилиацией автора с ИНИОН РАН и дискуссиями по дальнейшему развитию этого научного и информационного учреждения.

Мониторинг проводился в следующих целях:

- формирование политики по развитию информационного пространства социально-гуманитарных наук как неотъемлемой составляющей единого российского электронного пространства знаний;
- получение исходных данных для дальнейшего скоординированного развития информационных систем академических учреждений социально-гуманитарного профиля;
- организация эффективной навигации в информационном пространстве социально-гуманитарных наук академических учреждений;
- определение параметров перспективной программы оцифровки библиотечных, архивных и музейных фондов учреждений РАН;
- определение параметров перспективной программы обеспечения сохранности (архивного хранения) ценной научной электронной информации академических учреждений;
- разработка предложений по условиям и перспективам развития информационного пространства в сфере общественных и гуманитарных наук и определения роли ИНИОН РАН в координации развития информационного пространства.

В ходе проведения мониторинга решались следующие задачи;
●определение границ инфосферы социально-гуманитарных наук;

●сбор информации об информационных ресурсах академических организаций социально-гуманитарного профиля;

●описание, классификация и оценка информационных ресурсов и сервисов академических организаций социально-гуманитарного профиля;

●создание каталога информационных ресурсов и сервисов академических организаций социально-гуманитарного профиля и размещение его в Интернете.

3.2. Структура научной инфосферы

Очевидно, что границы мониторинга информационных ресурсов могут быть определены по различным основаниям (фасетам), которые в совокупности образуют структуру научного информационного пространства (инфосферы). Укажем, что инфосфера понимается как совокупность информационных ресурсов, сервисов и институций, участвующих в научной коммуникации [22]. Предлагается следующий состав из 16 фасетов, образующих, по мнению автора, структуру инфосферы.

1. *Уровень организации* информационных ресурсов:

- международный;
- общегосударственный;
- ведомственный;
- региональный;
- уровень коллабораций;
- уровень учреждения;
- персональный.

2. *Виды деятельности* (назначение ресурсов и сервисов):

- управление наукой;
- научно-информационная;
- издательская и книготорговая;
- библиотечная;
- архивная;
- музейная;
- образовательная;
- просветительская;
- массовая информация.

3. Тематика, определяемая используемой классификацией наук и отраслей деятельности (подробнее см. ниже).

4. Видовая структура ресурсов (наиболее общий уровень):

- первичные традиционные;
- первичные электронные;
- вторичные, т.е. содержащие ссылку на первичный документ;
- специальные (информационные системы, использующие специальные программные средства или способы обработки информации).

5. Семиотическая (знаковая) природа ресурсов. Наиболее распространенную классификацию предлагает стандарт «Дублинское ядро» (фасет «Типы ресурсов»):

- collection / коллекция;
- dataset / набор данных;
- event / событие;
- image / изображение;
- interactive resource / интерактивный ресурс;
- moving image / движущееся изображение;
- physical object / физический объект;
- service / услуга;
- software / компьютерная программа;
- sound / аудио;
- text / текст.

6. Оригинальность. Один из важнейших для инфосферы параметров классификации ресурсов – это их разделение по оригинальности происхождения. Обычно выделяются:

- первичные документы;
- первичные данные;
- вторичные (каталоги, ссылки, метаданные);
- производные (антологии, дайджесты, обзоры, справочники).

7. Физическая форма. Согласно действующему ГОСТ 7.72-96 «СИБИД. Коды физической формы документов», ресурсы (документы) могут иметь следующие виды физической формы:

- 1 – бумажный носитель;
- 2 – микроноситель;
- 3 – носитель на киноплёнке;
- 4 – магнитный носитель;
- 5 – перфоноситель;
- 6 – оптический (лазерный) носитель;
- 7 – носитель на голографической основе;
- 9 – прочие.

8. Способ реализации сервиса (тип коммуникации):

- Интернет;
- локальные сети или компьютеры;
- эфирные каналы;
- переносимые электронные носители;
- книготорговля, библиотечное или архивное комплектование.

9. Социально-экономическая модель сервиса или институции. В наиболее общем виде можно выделить следующие модели функционирования:

- государственная;
- коммерческая;
- общественная;
- смешанные.

10. Открытость. С точки зрения правовых условий доступа к ресурсам выделяются:

- открытые:
 - без права копирования или с другими ограничениями использования;
 - доступные без регистрации;
 - бесплатные;
- конфиденциальные (включая персональные данные);
- секретные.

11. Интеллектуальная собственность. Отношение к интеллектуальной собственности особенно существенно для электронных библиотек и других собраний произведений (фотобанки, электронные музеи). По этому основанию можно выделить следующие категории ресурсов:

- государственные открытые данные;
- произведения, являющиеся общественным достоянием;
- произведения, охраняемые авторским правом:
 - доступные по обычным лицензиям;
 - доступные по открытым (свободным) лицензиям;
 - нелицензионные произведения;
 - сиротские произведения.

12. Широта охвата информационных ресурсов, сервисов и институций:

- универсальные;
- широкоотраслевые;
- узкотематические;
- проблемные (нетематические) ресурсы.

13. Сектора науки, которые обслуживает инфосфера:

- академическая;
- вузовская;
- прикладная;
- просветительская / любительская.

14. Государственная принадлежность ресурсов или институций:

- российские;
- иностранные;
- межгосударственные с участием России;
- неизвестной принадлежности.

Критерии отнесения к государственной принадлежности:

- по гражданству или юрисдикции владельца ресурса;
- по гражданству или юрисдикции создателя ресурса;
- по географическому размещению сервера;
- по предмету (содержанию) ресурса;
- по условиям распространения.

15. Форма собственности. Из российской правовой базы исключено понятие собственности на информационные ресурсы, однако существует собственность на информационные системы. При этом информационные системы (а также технические средства), в которых размещаются ресурсы или сервисы, могут принадлежать разным субъектам. В соответствии с Гражданским кодексом различаются формы собственности:

- государственная;
- муниципальная;
- акционерная;
- частная;
- собственность общественных организаций;
- собственность зарубежных владельцев;
- смешанные формы собственности.

16. Язык ресурса:

- русскоязычные;
- многоязычные (включая русский);
- языки народов России;
- другие языки.

3.3. Определение границ сферы мониторинга

Далее указываются те фасеты из числа перечисленных, по которым реально определялись границы инфосферы в ходе мониторинга академической социально-гуманитарной инфосферы. Если фасет не используется, он не упоминается. Тематические границы рассматриваются отдельно.

Уровень. В сферу мониторинга включены ресурсы ведомственные, уровня учреждений, а также персональные. Международные, коллаборативные, общегосударственные и региональные ресурсы включаются только тогда, когда в их создании участвуют учреждения РАН и региональных академий.

Способ реализации сервиса. Включаются все способы распространения, кроме эфирного.

Открытость. В сферу мониторинга включаются только открытые ресурсы.

Сектора науки. В сферу мониторинга включаются ресурсы, относящиеся к академическому сектору науки. Прочие типы ресурсов включаются, только если в их формировании и реализации ресурса участвуют учреждения РАН или сотрудники этих учреждений.

Государственная принадлежность. В сферу мониторинга включаются ресурсы российские и международные с российским участием. Принадлежность ресурса к российским определяется по государственной принадлежности владельца ресурса.

Форма собственности. В сферу мониторинга включаются все типы ресурсов по форме собственности (государственные, акционерные, общественные, частные), при условии, что в формировании и реализации ресурса участвуют учреждения РАН или сотрудники этих учреждений.

3.4. Тематические классификации социальных и гуманитарных наук

В настоящем разделе будет рассмотрена тематическая структура инфосферы общественных наук, как она представлена в большинстве научных классификаций, применяемых как для структуризации научных коммуникаций, так и для других целей управления наукой.

Основной классификацией для структуризации российской инфосферы является *Государственный рубрикатор научно-технической информации (ГРНТИ)*. В ГРНТИ тематика общественных наук отражена в первом разделе (рубрики 00/26). Заметим, что в данном рубрикаторе не различаются социальные и гуманитарные науки. Такое разделение стало широко применяться в последние годы, во время создания ГРНТИ в 1970-е годы оно не было принято.

ГРНТИ используется в ВИНИТИ, ИНИОН, РИНЦ, ЦИТИС, на портале «Российское образование» и в других известных ресурсах научной информации.

Тематика общественных (т.е. социальных и гуманитарных) наук в терминах рубрик ГРНТИ 1-го уровня представлена в приложении 1. Далее кратко опишем тематическую структуру социальных и гуманитарных наук, как они представлены в некоторых других распространенных российских и международных классификациях.

Классификация ВАК (номенклатура специальностей научных работников) [74] применяется для определения научных руководителей, оппонентов, тематики ученых советов и, конечно, диссертаций, в том числе в Электронной библиотеке диссертаций в РГБ. В приложении 2 приводится верхний уровень классификации ВАК в части социально-гуманитарных наук.

Рубрикатор Российского гуманитарного научного фонда (РГНФ) используется для определения тематики заявок на гранты и оформления отчетов. Правда, сейчас произошло объединение РГНФ и РФФИ в единый фонд, но пока нет информации о том, что рубрикатор объединенного фонда в части социально-гуманитарных наук как-то изменился. В приложении 3 приводится 1-й и частично 2-й уровень рубрикатора РГНФ.

Рубрикатор Российского научного фонда (РНФ) частично повторяет рубрикатор РГНФ. Он используется для тех же целей, но в меньших масштабах. В приложении 4 приводится фрагмент Рубрикатора РНФ (1-й и 2-й уровни).

Рубрикатор Организации экономического сотрудничества и развития (ОЭСР, OECD), известный также как классификатор FOS (Fields of Science). Это наиболее распространенный на сегодня в мире официальный классификатор областей науки, разработанный экспертами ОЭСР. В России он используется в конкурсе мегагрантов, его брали за основу при разработке нового перечня укрупненных групп специальностей бакалавриата, магистратуры и аспирантуры. FOS широко применяется в международных системах

статистики науки и для решения других научно-управленческих задач. Его используют большинство международных организаций и межгосударственных объединений: ООН, ЮНЕСКО, Европейский союз и др. Фрагмент FOS, относящийся к общественным и гуманитарным наукам, приводится в приложении 5. Официальную версию рубрикатора ОЭСР можно найти по адресу [126]. Чтобы исключить спорные толкования, мы приводим все международные классификации на английском языке, как они и используются на практике.

Рубрикаторы Web of Science. Из многочисленных информационных и наукометрических международных систем наиболее известен и авторитетен индекс компании Clarivate Analytics, широко известный по обозначению WOS (Web of Science). Этот индекс состоит из трех частей: общенаучный указатель, указатель по социальным наукам и указатель по искусствам и гуманитарным наукам. Рубрикаторы этих индексов представляют собой алфавитные перечни рубрик с пояснениями. В приложениях 6 и 7 представлены перечни рубрик по социальным наукам и по искусствам и гуманитарным наукам (без пояснений) по версии 2012 г., размещенной на сайте компании [149].

Рубрикатор Scopus. Другой почти столь же известный международный наукометрический инструмент, который создан компанией Эльзевир, получил название Scopus. Его рубрикатор представляет собой двухуровневую алфавитно-предметную классификацию [103]. Выборка, представленная в приложении 8, отражает, по мнению автора, основные направления социальных и гуманитарных наук, хотя, конечно, здесь возможны дискуссии.

Сравнение этих классификаций дает богатую пищу для размышлений о тематической структуре социальных и гуманитарных наук.

Во-первых, как уже было сказано, само это разделение не является общепринятым и однозначно понимаемым, что мы видим на приведенных примерах.

Так, например, *история* в классификациях Scopus и WOS относится как к социальным, так и к гуманитарным дисциплинам. Хотя в комментариях делается попытка различать аспекты рассмотрения исторических документов при отнесении их к социальным и гуманитарным наукам, эта попытка представляется неудачной.

Неясно, относятся ли к социальным или гуманитарным регионаловедческие науки (например, востоковедение, африканистика, финно-угорика).

Во-вторых, сам состав дисциплин, включаемых в эти категории, значительно различается в разных классификациях, что влечет за собой различие в управленческих решениях, вытекающих из соответствующего подхода.

Приведем некоторые примеры. В классификации WOS к социальным относятся такие дисциплины, как антропология, геронтология, уход за младенцами, клиническая психология, политика здравоохранения, которые в российской традиции относятся к медицине. Другой пример: гостиничное хозяйство, спорт и туризм в ГРНТИ отнесены к отраслям народного хозяйства. То же относится к управлению, организации бизнеса как в целом, так и (особенно) в отраслях, например в транспорте. Неясен вопрос с социальной и экономической географией, которая в отечественных классификациях включается в состав географии, хотя в ГРНТИ в разделе общественные науки есть рубрика «Комплексное изучение стран и регионов».

В рубриках РГНФ и РНФ отсутствует множество дисциплин, которые есть в ГРНТИ и в зарубежных классификациях, например средства массовой коммуникации и документалистика.

В классификации FOS четко разделяются социальные и технические аспекты некоторых дисциплин, например транспорта. В отечественных классификациях такого выделения нет.

Особый вопрос – это информатика. В отечественной литературе давно идет спор о сущности и отраслевой принадлежности данной науки (или группы наук). Некоторые авторы полагают, что эта наука междисциплинарная, другие – что в основном техническая (применение компьютеров). Автор вслед за составителями классификации FOS полагает, что существуют две информатики – компьютерная, место которой среди точных наук, и социальная, изучающая информационные процессы в обществе. Последняя должна включать и традиционные науки о коммуникации – библиотековедение, теорию СМИ и др. Но эту точку зрения разделяют далеко не все составители классификаций. Исследованию этой проблемы был посвящен специальный доклад автора [15].

Таким образом, по многим аспектам тематической структуры социальных и гуманитарных наук между разными отечественными и международными классификациями существуют за-

метные расхождения. Это следует иметь в виду при любых действиях по управлению инфосферой.

3.5. Тематические границы мониторинга

Для практических задач мониторинга, проведенного ИНИОН РАН, было принято решение использовать ГРНТИ, как общепринятый для сферы научной информации в России. При этом, соответственно, социальные и гуманитарные науки не разделялись.

В сферу мониторинга включаются все рубрики ГРНТИ 00/26, кроме следующих:

14 Народное образование. Педагогика;

18 Искусство. Искусствоведение;

19 Массовая коммуникация. Журналистика. Средства массовой информации.

Это объясняется тем, что в традиционную сферу социальных и гуманитарных исследований, проводимых в учреждениях РАН, эти дисциплины не входят.

Некоторые рубрики могут в сферу мониторинга включаться частично, в той степени, в которой эта тематика представлена в исследованиях и публикациях учреждений РАН социально-гуманитарного профиля:

13 Культура. Культурология;

15 Психология;

20 Информатика.

Это означает, что информационные ресурсы и сервисы по соответствующим дисциплинам будут отражаться неполно.

Предлагаемый вариант границ мониторинга определяется перспективными задачами создания академической информационной системы, программ оцифровки и пр. В то же время очевидно, что при решении любой задачи, связанной с управлением и развитием инфосферы, перед исполнителями встанут аналогичные задачи определения границ. Автор надеется, что предлагаемая модель инфосферы может оказаться полезной во многих других случаях.

3.6. Функциональная классификация информационных ресурсов

Далее рассматривается классификация информационных ресурсов, выявленных в ходе мониторинга информационного пространства академических организаций социогуманитарного профиля. Принципы и методика проведения мониторинга описаны в публикации [14].

Мониторингом были охвачены свыше 200 институций РАН и республиканских академий наук (научных учреждений, самостоятельных подразделений научных учреждений и организационных структур РАН). Всего было выявлено свыше 16 тыс. информационных ресурсов (массивов) разного рода, включая традиционные и электронные, первичные и вторичные, библиотечные, архивные, музейные, сайты и самостоятельные страницы, базы данных, аккаунты в социальных сетях и др.

Заметим что недавно в рамках концепции создания Единой системы информационного обеспечения науки (ЕСИОН) автором совместно с Н.Е. Каленовым и В.А. Серебряковым [3] была предложена классификация информационных ресурсов, в основу которой были положены четыре категории ресурсов:

- первичные традиционные;
- первичные электронные;
- вторичные, т.е. содержащие ссылку на первичный документ;
- специальные (информационные системы, использующие специальные программные средства или способы обработки информации).

Смысл этой классификации заключался в том, что для каждой из перечисленных категорий ресурсов должен быть предложен свой набор действий в рамках создания ЕСИОН. Для первичных традиционных ресурсов должна быть предложена программа оцифровки, для первичных электронных – программа сохранности, для вторичных – программа интеграции. Специальные ресурсы на первом этапе создания ЕСИОН предлагалось только учитывать.

Перед проведением мониторинга на основе этого подхода была разработана функциональная трехуровневая классификация, представленная в приложении 9.

Центральная дихотомия этой классификации, которая вытекает из основного тренда развития инфосферы в наше время, – это различие традиционных и электронных форм информационных ресурсов. Рассмотрим эту дихотомию подробнее.

Большинство видов научной информационной продукции находятся в промежуточном или переходном положении между двумя формами. Приведем примеры.

Наиболее авторитетный каталог периодических изданий Ulrich's Periodicals Directory [148] – регистрационная база данных, включающая в настоящее время описания более 220 тыс. текущих периодических (regular) и продолжающихся (irregular) изданий на 200 языках от 90 тыс. издающих организаций. В этой базе данных на каждый формат издания – печатный (print), электронный (online), CD-ROM, микроформа – дается отдельное описание. Издание может иметь несколько описаний в зависимости от числа имеющихся форматов. Оригинальная и переводная версии журналов имеют самостоятельные записи.

Статистика показывает, что большинство российских научных изданий доступны в двух версиях – печатной и электронной. При этом если печатная версия очевидным образом имеет единственную форму, то электронная версия может быть представлена в различных вариантах, в том числе на сайте издателя или в разных электронных библиотеках – полными текстами, аннотациями, содержанием выпусков, сводными указателями, избранными полными текстами. Встречаются комбинированные варианты, когда одни выпуски представлены полными текстами, а другие – аннотациями или содержанием.

Периодические издания как в электронной, так и в печатной форме в различных местах хранения существенно различаются глубиной ретроспективы и полнотой представления.

Таким образом, одно и то же периодическое издание может быть представлено в информационном пространстве в разных формах и с разной полнотой.

Похожая ситуация с трудами конференций. Можно найти примеры распространения трудов конференций как в традиционной, так и в электронной форме, через электронные библиотеки, сайты издателей и организаторов конференций, полными текстами или только в виде программ (наименования докладов).

Непростым также является вопрос об учете и мониторинге научных монографий в разных формах. С одной стороны, большинство научных книг пока выходят в печатной форме. С другой стороны, ведется активная оцифровка книг, как легально, так и нелегально. Общее количество книг на русском языке, доступных в Интернете, превышает 2 млн. Подавляющее большинство академических учреждений размещают на своих сайтах либо полные

тексты изданий этих институтов, либо обложки и аннотации, с возможностью заказать книгу в печатной или электронной форме.

С января 2017 г. в РГБ и в РКП начал поступать обязательный электронный экземпляр книг, однако вопрос о режиме доступа к этим книгам пока не решен.

Еще один пример – это диссертации. Как известно, согласно действующим правилам, диссертационный совет принимает к защите диссертации в двух формах – печатной и электронной, причем электронная форма размещается на сайте диссертационного совета. То же касается авторефератов, отзывов официальных и неофициальных оппонентов и других документов, сопровождающих защиту. Фактически весь комплект диссертационных документов дублируется в двух формах. Также теперь создается и иногда размещается на сайте видеозапись защиты. Можно добавить, что диссертации продолжают поступать в РГБ в печатном виде, и там вновь происходит их оцифровка, а затем они хранятся в двух формах.

При этом диссертации доступны на сайте диссертационного совета бесплатно, а через многочисленные коммерческие электронные библиотеки – за плату. Это пример того, что разделение по классификационному признаку – платности доступа к ресурсам – также не является однозначным.

Ярким примером смешанной формы представления информации являются библиотечные каталоги. В академических библиотеках можно наблюдать полный спектр вариантов представления библиографической информации: встречаются как ситуации, когда завершен переход к электронному каталогу в виде библиографической БД, включая ретроконверсию, так и ситуации, когда создание электронного каталога практически еще не начато и вся библиография реализована в виде одного или многих карточных каталогов. Конечно, можно наблюдать и множество промежуточных форм. Примером может служить научная библиотека Якутского научного центра РАН, в которой имеются 10 традиционных каталогов в карточной форме и девять библиографических баз данных.

Всего в обследованном информационном пространстве обнаружено:

- традиционных библиографических каталогов и картотек – 185;
- электронных каталогов – 88;
- библиографических БД – 67.

Очевидно при этом, что деление на электронные каталоги и библиографические БД достаточно условно.

Другие виды библиографической информации – различные указатели и перечни – также можно обнаружить как в электронной, так и в печатной форме. Это же относится к реферативной информации. Хотя тиражи печатных реферативных журналов и сборников все последние годы сокращаются, их создатели не торопятся отказаться от печатных форм, сохраняя тем самым обе формы.

Таким образом, наиболее важные виды научной информационной продукции находятся на переходном этапе от традиционной к электронной форме распространения, причем этот переходный этап может затянуться надолго.

Можно, однако, указать на примеры, когда переход от традиционных к электронным формам практически завершился. Это касается научных отчетов, а также статистических данных, которые теперь создаются и передаются в вышестоящие организации только в электронной форме.

Еще одна проблема, затрудняющая разделение научных ресурсов на традиционные и электронные, – это наличие ресурсов на аналоговых технических носителях – магнитных лентах и тому подобных устройствах, – а также на микроформах. Это касается прежде всего аудио- и видеоинформации. Очевидным решением представляется выделение этих ресурсов в отдельную категорию, что реально и происходит в больших универсальных информационных системах, таких, например, как Europeana.

В то же время существует много видов информационных ресурсов, четко идентифицируемых как традиционные или как электронные (цифровые). Это, с одной стороны, традиционные библиотечные, архивные и музейные фонды, с другой – информационные системы и ресурсы, не имеющие печатного аналога, например ГИСы, 3D-модели, инфометрические или аналитические базы данных.

Из всего этого следует, что разделение ресурсов на традиционные и электронные (цифровые и аналоговые), по крайней мере в некоторых случаях, не может быть базовым основанием для классификации ресурсов, а должно выступать в качестве фасета.

Следующая важная дихотомия – это разделение ресурсов на первичные ресурсы (документы и массивы документов, а также первичные данные) и вторичные, т.е. содержащие ссылки на первичные документы. Первичные документы разделены на традиционные и электронные. Массивы первичных традиционных документов – это, прежде всего, библиотечные, архивные и музейные

фонды, а также массивы фото-, аудио-, кинодокументов, а массивы первичных электронных документов – электронные коллекции, библиотеки и электронные архивы.

Данное разделение также не является абсолютным. Так, не очевидно отнесение к этим категориям некоторых видов документов, например дайджестов или презентаций. Последние могут быть как оригинальными, отражающими содержание научных докладов, отсутствующих в полном виде, так и вторичными.

Особую проблему с точки зрения разделения ресурсов на первичные и вторичные представляет классификация периодики, представленной в электронном виде. Периодическое издание может существовать в различных формах: полные электронные копии, аннотации статей, содержание выпусков и др. Встречаются случаи смешанного представления: одних статей – в виде полных текстов, других – в виде аннотаций или только библиографических описаний. Могут быть представлены в разных формах выпуски разных лет. Нередки случаи, когда периодическое издание представлено в разных формах на разных сайтах или в разных электронных библиотеках. В предлагаемой классификации все формы электронной периодики для удобства представлены вместе в разделе «Первичные электронные ресурсы».

Другие виды ресурсов, которые в этой классификации называются специальными, представляют собой информационные системы, включающие дополнительные программные средства, сервисы и возможности по сравнению с документальными электронными системами – библиотеками и архивами.

Отметим, что в отдельный тип (в силу большого количества и разнообразия) выделены лингвистические электронные ресурсы. Вообще, если говорить о тематическом разрезе, филология (и особенно языкознание) является лидером по числу и разнообразию выявленных электронных ресурсов среди всей социогуманитарной сферы. Нужно также учитывать, что, несмотря на то что лингвистические ресурсы (словари, грамматики) продолжают создаваться и в традиционной форме, практически все виды этих ресурсов представлены в электронном виде. Подробнее анализ лингвистических ресурсов представлен в главе 6.

К категории специальных были первоначально отнесены также средства представления ресурсов в Интернете: сайты, аккаунты в социальных сетях и некоторые другие.

Следует специально указать, что показателем для мониторинга и учета электронных ресурсов учреждения, представленных

в Интернете, является его «видимость», т.е. наличие и количество ссылок на ресурсы учреждения (прежде всего на его основной сайт) в различных каталогах или при использовании специальных инструментов поисковых машин Сети. Таким инструментом, например, является Тематический индекс цитируемости (ТИЦ) Яндекса. «Видимость» является одним из важнейших показателей при вебометрических измерениях информационных ресурсов. С недавнего времени стали использоваться более тонкие показатели, например упоминаемость ресурса в социальных сетях. Подробнее о способах измерения интернет-ресурсов см. работу [11]. Однако исследования видимости ресурсов на данном этапе мониторинга не проводились.

В предлагаемой классификации не отражены ресурсы, которые не являются специфическими для научных учреждений, например ресурсы систем документооборота, бухгалтерских, кадровых и других управленческих сервисов.

Предварительная функциональная классификация представлена в приложении 9. В некоторых частях она соответствует традиционной структуре, в других предлагается новая типология. Всего в функциональной классификации на трех уровнях имеется 215 рубрик.

После завершения мониторинга, в ходе формирования Навигатора информационных ресурсов, стало очевидно, что предлагаемая классификация является слишком детальной и сложной и навигация по ней будет затруднительной. К тому же многие функциональные типы ресурсов оказались представленными одним-двумя экземплярами, поэтому выделение для них специальных функциональных классов было признано избыточным.

В результате было принято решение упростить функциональную классификацию, сделать ее двухуровневой, многие рубрики объединить. Полученную упрощенную классификацию, которая используется в действующем Навигаторе, можно назвать навигационной типологией информационных ресурсов. Она представлена в табл. 6 в разделе 3.10 и в приложении 10. Эта классификация включает 15 рубрик первого уровня (типов), причем некоторые типы разделены на виды.

Всего классификация НИРОН включает на двух уровнях 69 рубрик.

3.7. Методические вопросы инвентаризации информационных ресурсов

Как уже отмечалось, важной особенностью современного этапа развития инфосферы является активный переход от традиционных форм научной коммуникации к цифровым.

Известно, что многие виды научной продукции (например, диссертации, отчеты, научная статистика и др.) создаются и распространяются главным образом в цифровой форме. В наиболее массовых формах научной коммуникации (книги, журналы) налицо конкуренция цифровых и традиционных способов и частичное их дублирование. Также активно осуществляется оцифровка библиотечных, архивных и музейных фондов. Электронные библиотеки стали одним из самых посещаемых научных ресурсов, и в этом они значительно обогнали библиотеки традиционные.

Параллельно возник ряд новых, полностью цифровых форм коммуникации (сайты, социальные сети, ресурсы коллективного творчества), для которых отсутствуют установленные правила учета, инвентаризации, а также официальная статистика.

В этих условиях императивом проводимой инвентаризации является комплексный подход к анализу ресурсов: в сферу инвентаризации должны быть включены как традиционные ресурсы – библиотечные, архивные и музейные фонды, издательская продукция, – так и новые формы цифровых научных информационных продуктов и средств коммуникации.

В ходе проведения инвентаризации решались многочисленные методические вопросы, связанные с определением границ инфосферы общественных и гуманитарных наук.

Прежде всего это касалось определения организационной сферы исследования. Очевидно, что информационные ресурсы создаются и хранятся:

- в специализированных информационных учреждениях – библиотеках, архивах, музеях;
- в собственно научных учреждениях;
- в органах управления наукой;
- в структурах (институциях), созданных научными учреждениями (примерами таких институций могут быть журналы, проекты, некоммерческие организации, советы, комиссии и др.).

Сфера инвентаризации была ограничена академическими учреждениями. В результате инвентаризацией были охвачены следующие институции:

- научные учреждения, входящие в отделения РАН социогуманитарного профиля (Отделение общественных наук, Отделение глобальных проблем и международных отношений и Отделение историко-филологических наук), а также учреждения, находящиеся под их научно-методическим руководством;

- библиотеки, архивы и музеи РАН, большая часть которых являются структурными подразделениями учреждений;

- учреждения республиканских академий наук социогуманитарного профиля;

- управляющие структуры РАН, включая Президиум РАН, отраслевые отделения РАН социогуманитарного профиля, региональные отделения РАН, региональные научные центры и республиканские академии;

- научные учреждения РАН или их подразделения, не являющиеся по основному профилю гуманитарными, но обладающие ресурсами социогуманитарного или универсального профиля. Пример – Межведомственный суперкомпьютерный центр РАН, поддерживающий Электронную библиотеку «Научное наследие России»;

- институты, созданные научными учреждениями РАН и обладающие собственными ресурсами социогуманитарного профиля. Примеры – Комиссия по истории Великой Отечественной войны при Институте российской истории РАН, Музейный совет СО РАН.

Всего в сферу инвентаризации вошли 214 учреждений и других институций. Полученный список послужил исходным материалом для первого этапа инвентаризации, а именно – был использован для рассылки запросов в учреждения и в различные информационные системы.

Сложным вопросом, связанным с определением границ инфосферы, является вопрос об идентификации ресурсов по истории науки, прежде всего архивных и музейных фондов по естественным наукам, в том числе материалов мемориальных музеев ученых-естественников, личных архивных фондов и т.п. Очевидно, что, с одной стороны, это ресурсы по истории науки, т.е. относятся к социогуманитарному профилю, поскольку науковедение является общественной дисциплиной, а история науки (в том числе по отдельным отраслям) относится к науковедению. С другой стороны, – эти ресурсы в большинстве своем относятся к истории конкретных естественных наук, например химии. Заметим, что в ГРНТИ для таких ресурсов имеются два варианта индексирования –

во-первых, через рубрику «12.09.09 История науки», а во-вторых – через рубрику «XX.01.09 История науки (по отраслям)» в перечне рубрик «Общие вопросы».

В результате было принято паллиативное решение: в сферу учета были включены все музеи и архивные фонды РАН, но описаны на самом верхнем уровне.

Еще один нетривиальный аспект – это установление принадлежности ресурса академической организации. Были выявлены несколько организационных форм создания и использования ресурсов, в отношении которых возникали неясности. Так, Институт философии РАН наряду с журналами, издаваемым самим институтом (т.е. бесспорно академическими), указывает на ряд журналов, «связанных с институтом», но издаваемых сторонними организациями.

Другой пример. Есть несколько ценных ресурсов, в частности лингвистических, созданных учеными, работающими в академических организациях, но эти ресурсы поддерживаются как личные, без указания на аффилиацию с РАН.

Третий пример. Одним из источников для сбора информации была база данных РГНФ, в которой были выбраны гранты, относящиеся к типу «в», – проекты создания информационного обеспечения, необходимого для выполнения научных проектов. Среди этих проектов к академическим были отнесены исследования, финансирование которых осуществлялось через организации РАН. Однако некоторое количество грантов, работу в рамках которых вели ученые РАН, было получено через специально созданные организации-посредники. Можно ли ресурсы, созданные таким образом, считать академическими, не вполне ясно.

Во всех этих случаях принадлежность ресурсов к академическим определялась с учетом мнения владельца ресурсов.

Однако есть и другие случаи. Ряд ресурсов, созданных при очевидном участии учреждений РАН при использовании грантов научных фондов после завершения проекта, на который выдан грант, не поддерживаются учреждениями РАН, но поддерживаются посторонними для РАН лицами и организациями. В частности, это относится к Институту русского языка РАН, руководство которого не считает такие ресурсы принадлежащими институту.

Очевидно, что все эти случаи свидетельствуют о необходимости наведения порядка в сфере владения и распоряжения информационными ресурсами, чему способствует проведенная инвентаризация.

3.8. Сбор информации в ходе мониторинга

В качестве источников информации использовались:

- результаты опроса академических учреждений, подведомственных ФАНО;
- данные из информационных систем и каталогов, статистической отчетности;
- данные с сайтов учреждений РАН и других сайтов.

Запрос о наличии информационных ресурсов был направлен в академические учреждения согласно списку, составленному с учетом критериев, описанных выше. Запрос был направлен всего в 160 организаций социально-гуманитарного профиля. Были запрошены следующие сведения для проведения мониторинга информационных ресурсов.

1. Сетевые ресурсы. Должен быть приведен перечень интернет-адресов сайтов и страниц, на которых размещаются научные результаты учреждения, включая официальный и неофициальные сайты учреждения, сайты подразделений, персональные страницы ученых, адреса баз данных и других информационных систем, адреса проектов, журналов, конференций и др.

2. Локальные ресурсы. Должно быть приведено описание в свободной форме имеющихся в учреждении ресурсов с указанием их объема на электронных и традиционных носителях, в том числе библиотечных каталогов, массивов текстовых документов, баз данных, геоданных, графических и фотоколлекций, аудио- и видеодокументов, картотек, материалов экспедиций, словарей, справочников, досье и других ресурсов.

3. Библиотечные фонды. Должна быть указана тематика фонда в свободной форме и приведены сведения об объеме фонда.

4. Архивные фонды. Должны быть приведены наименования фондов, их объем в единицах хранения и в листах, число описей, наличие описей в электронной форме. Должно быть указано число единиц хранения фото-, аудио- и видеодокументов.

При наличии таких сведений должна быть указана доля документов, созданных до 1940 г.

Допустимо предоставление копии отчета по формам Росархива, однако по архивным фондам, стоящим на государственном учете и описанным в ЦФК, предоставление данных необязательно, поскольку может быть получено из централизованных источников.

5. Музейные фонды. Приводится наименование коллекций, описание в свободной форме их содержания, указываются общее

количество музейных предметов по каждому типу, наличие и объем инвентарного и научного каталогов в традиционной и электронной формах.

Для каждой категории ресурсов указываются контактные данные администратора ресурса или другого ответственного лица.

Ответы на посланный запрос были получены из 30 организаций.

Также были проанализированы сведения об информационных ресурсах академических организаций, имеющиеся в следующих информационных системах:

- Научная электронная библиотека eLIBRARY (сведения о периодических изданиях, выпускаемых учреждениями РАН) [64];

- Единая государственная информационная система учета научно-исследовательских, опытно-конструкторских и технологических работ (ЕГИСУ НИОКТР) – сведения об отчетах, представленных учреждениями РАН [43];

- информационная система Российского фонда гуманитарных исследований – сведения о грантах типа «в» (проекты создания информационного обеспечения, необходимого для выполнения научных проектов, выданных организациям РАН)¹;

- информационная система Всероссийской аттестационной комиссии – сведения о диссертационных советах, имеющихся в организациях РАН [39];

- информационная система Mnemosyne – сведения об архивах, архивных фондах и описях, имеющихся в организациях РАН [131];

- сведения о музеях, имеющихся в организациях РАН, получены из нескольких источников, в том числе из ресурсов [84; 131]. Портал Музейного совета РАН в настоящее время не действует;

- статистические отчеты Информационно-библиотечного совета РАН – сведения об объемах и движении фондов библиотек, входящих в ЦБС РАН по состоянию на 2015 г. [45].

Однако основным источником информации для сбора сведений об информационных ресурсах послужили сайты академических учреждений. Большая часть сведений, представленных в НИРОН, получена именно из этого источника.

¹ В настоящее время эта система включена в общую информационную систему РФФИ, однако открытый доступ к электронному архиву РГНФ отсутствует.

3.9. Результаты мониторинга и состав информационных ресурсов

Результаты собственного мониторинга, как и следовало ожидать, отличаются от тех, которые предполагались в запросе. Это объясняется разнообразием информации, поступающей из различных источников, неполнотой ответов, а также неопределенностью некоторых базовых понятий. Прежде всего, это относится к понятию социогуманитарного профиля информационного ресурса, а также к типологии ресурсов.

1. Многие ресурсы, например библиотечные фонды или коллекции научных отчетов, а также массивы документов научных центров и региональных отделений РАН, могут быть универсальными по тематике, но включать гуманитарную часть.

2. Не очень ясно, к какому профилю можно отнести множество различных ресурсов по истории естественных и точных наук, прежде всего архивные и музейные фонды.

3. Многие ресурсы, очевидно социогуманитарные по тематике, например лингвистические, создаются и поддерживаются учреждениями РАН отделения информатики.

4. Типология ресурсов, реально имеющих в учреждениях и отраженных на их сайтах, не всегда однозначно соотносится с выделенными нами типами ресурсов.

5. Тематика ИР во многих случаях не может быть определена по имеющимся данным, поэтому на данном этапе она была исключена из сферы мониторинга.

6. Некоторые виды ИР, например музейные фонды, практически не описаны, поэтому мы ограничились краткими сведениями о музеях в целом. С учетом сказанного выше, в перечень включены все музеи РАН.

Результаты мониторинга по основным типам ресурсов кратко представлены ниже. В полном виде они размещены на сайте ИНИОН РАН в каталоге НИРОН [60].

Архивные фонды

Число организаций, в которых имеются архивные фонды, – 44. Всего фондов в этих организациях 5288, но по семи учреждениям нет сведений. Описей 7877, но по 12 учреждениям сведения отсутствуют. Имеются сведения о некотором количестве путеводителей

по архивным учреждениям и каталогов по отдельным фондам. Число архивных фондов, не состоящих на государственном учете, невелико (3–5).

Фонды традиционных библиотек

Всего мониторингом были охвачены 74 библиотеки социогуманитарного или универсального профиля. Их распределение по централизованным библиотечным сетям (ЦБС) представлено в табл. 4.

Таблица 4

Состав ЦБС в РАН

Головная организация ЦБС	Количество социогуманитарных библиотек в ЦБС	Количество универсальных библиотек в ЦБС
ИНИОН	21	
БАН	14	1
ГПНБ СО РАН	5	6
ЦНБ УРО РАН	3	4
ЦНБ ДВО РАН	1	3
Не входят в ЦБС	5	10
БЕН и ЦНСХБ		2
Всего	49	25

БЕН РАН и ЦНСХБ являются отраслевыми, но включают гуманитарную часть фонда, поэтому мы их считаем условно универсальными. Таким образом, в РАН имеется всего 25 универсальных библиотек (включая БЕН и ЦНСХБ). Всего библиотек социогуманитарного профиля выявлено 49.

Объемы книжных фондов библиотек РАН, входящих в ЦБС, представлены в табл. 5. Данные приведены по сведениям Информационно-библиотечного совета (ИБС) РАН по состоянию на 2015 г. Сведений о библиотечных фондах других библиотек РАН в статистике ИБС нет. Также нет сведений о библиотечных сетях бывших РАМН и РАСХН, других государственных и республиканских академий. После 2015 г. статистика по библиотечным фондам РАН отсутствует, а сам Информационно-библиотечный совет распущен.

Таблица 5

**Объемы книжных фондов библиотек
Российской академии наук в 2015 г.**

№ п/п	Сеть библиотек Российской академии наук на 01.01.2016 г.	Книжные фонды			
		Поступило литературы в 2015 г.		Состоит на 01.01.2016 г.	
		Всего	В том числе иностранной литературы	Всего	В том числе иностранной литературы
1.	БАН ее сеть – 26 б-к всего 27 б-к	219 725 15 296 235 021	39 724 6191 45 915	12 297 430 5 270 926 17 568 356	3 960 678 2 761 304 6 721 982
2.	БЕН РАН ее сеть – 93 б-ки всего 94 б-ки	8218 63 560 71 778	4777 23 900 28 677	1 095 898 10 655 929 11 751 827	847 745 4 026 835 4 874 580
3.	ГПНТБ СО РАН ее сеть – 69 б-к всего 70 б-к	122 017 39 696 161 713	3782 4319 8101	9 488 460 4 773 726 14 262 186	1 987 017 1 348 872 3 335 889
4.	ЦНБ ДВО РАН ее сеть – 19 б-к всего 20 б-к	6402 3355 9757	443 165 608	787 276 980 042 1 767 318	310 680 148 549 459 229
5.	ЦНБ УрО РАН ее сеть – 25 б-к всего 26 б-к	8829 8866 17 695	1584 364 1948	1 000 398 1 590 104 2 590 502	487 061 420 299 907 360
6.	ФБ ИНИОН РАН ее сеть – 21 б-ка всего 22 б-ки	45 586 15 622 61 208	6038 1741 7779	11 673 785 2 860 640 14 534 425	1 463 209 *** 879 812 2 343 021
	Библиотеки РАН – 6 б-к их сети – 253 всего – 259 б-к	410 777 146 395 557 172	56 348 36 680 93 028	36 343 247 26 131 367 62 474 614	9 056 390 9 585 671 18 642 061

Музейные фонды

В сферу мониторинга вошли свыше 100 музеев РАН, в том числе: естественно-исторические музеи, этнографические, исторические музеи, Литературный музей (Пушкинский дом), мемориальные музеи, а также музеи научных центров и региональных отделений РАН. Нужно сказать, что сведения о количестве музеев РАН разнятся в различных источниках, а официального учета не существует.

Описаний музейных фондов и данных об их объемах в открытом доступе практически нет. Также нет сведений о наличии инвентарных книг и каталогов в традиционной или электронной формах и о наличии оцифрованных копий музейных предметов.

Очевидно, состав и структура музейного фонда РАН нуждаются в систематической инвентаризации и анализе.

Периодические, продолжающиеся и сериальные издания (ППСИ)

Этот вид коммуникации остается самым важным для общественных наук, и в нем одновременно реализуется наибольшее разнообразие форм и способов хранения и представления доступа.

Всего в сферу мониторинга вошло свыше 400 названий ППСИ социогуманитарного профиля, сведения о которых приводятся на сайтах учреждений РАН, в базах данных Научной электронной библиотеки и многочисленных каталогах периодики. Впрочем, связь издания с учреждениями РАН неоднозначна. Пример с Институтом философии РАН приведен выше. Ряд ППСИ издаются совместно академическими и другими организациями.

Кроме того, в фондах академических библиотек хранится значительное количество архивных массивов ППСИ, изданных как учреждениями РАН, так и другими организациями.

Форма представления ППСИ также сильно различается: встречаются архивы полных текстов всего издания, тексты отдельных статей, содержание или аннотированные содержания выпусков и пр. Также имеется большой разброс по полноте представления издания, глубине ретроспективы. В целом можно сказать, что полными текстами представлена существенная доля академических ППСИ (до 30%), особенно изданных 2–10 лет назад.

Значительная доля академических ППСИ (не менее 160 наименований) социогуманитарного профиля имеется в Научной электронной библиотеке и в других электронных библиотеках. Однако в них имеется большой разброс по полноте и глубине ретроспективы. Полные комплекты с начала издания ППСИ весьма редки.

Вторичные ресурсы

К ним мы отнесли библиотечные традиционные и электронные каталоги, библиографические, аннотированные и реферативные БД, указатели, перечни, списки и картотеки.

Всего в обследованных учреждениях было обнаружено не менее 300 библиотечных каталогов в традиционной или электронной форме, притом что собственно библиотек было выявлено 74. В большинстве учреждений сочетаются обе формы, причем электронные каталоги обычно ограничены годом начала создания. Полностью электронных каталогов немного (не более пяти библиотек). Всего обнаружено:

- традиционных библиографических каталогов и картотек – 185;
- электронных каталогов – 88;
- библиографических БД – 67;
- библиографических перечней трудов учреждения – 94;
- библиографических перечней по научным подразделениям

или по отдельным сотрудникам – более 100.

Реферативные БД создаются почти исключительно в институтах научной информации. Аннотированные перечни создаются гораздо чаще, в основном по изданиям данного учреждения, обычно включают также обложку и содержание издания, всего их обнаружено 73.

Библиографические, аннотированные и реферативные информационные продукты составляют большинство информационных ресурсов, всего их обнаружено не менее 500. Следует упомянуть о большом количестве смешанных ресурсов, когда в библиографических перечнях у части произведений имеются отсылки к полному тексту

Первичные документальные ресурсы

Наибольший интерес для пользователя представляют полнотекстовые коллекции документов, включая электронные библиотеки. Среди них можно отметить следующие.

Опубликованные документы:

• полнотекстовые коллекции публикаций учреждений и сотрудников – 119;

• электронные библиотеки – 131.

Неопубликованные документы

• отчеты – представлены на сайтах 61 учреждений;

• в ЕГИСУ ЦИТИС имеется свыше 1,5 тыс. отчетов 62 организаций, включая коллекции по гуманитарной и универсальной тематике;

• экспертные заключения – в 14 учреждениях;

- описания проектов, грантов, услуг, экспедиций – 135;
- очерки направлений деятельности и результатов учреждений и подразделений, исторические очерки – 116¹.

Отдельно нужно сказать о диссертациях. После известных постановлений ВАК все диссертационные советы обязаны размещать в Интернете весь пакет документов, включая видеозапись защиты. Это, правда, делают не все учреждения, но практически все размещают авторефераты и диссертации, начиная с 2014 г. Всего в социогуманитарном секторе РАН в 34 учреждениях имеются 78 диссертационных советов. При этом нужно иметь в виду, что диссертации на сайтах учреждений и диссертационных советов размещаются раньше, чем в Электронной библиотеке диссертаций в РГБ, и при этом доступны бесплатно.

В целом очевидно, что в Интернете доступно большое количество полных текстов научных произведений социогуманитарной тематики, размещенных академическими учреждениями. Общее количество обнаруженных коллекций, включая смешанные, достигает 700. Хотя в большинстве коллекций количество документов не указано, по экспертной оценке их суммарная величина составляет не менее 300 тыс. наименований.

Информационные системы

Мы сочли целесообразным выделить в отдельную категорию ИР разнообразные информационные системы, отличные от документальных систем – электронных библиотек и библиографических БД. Всего таких ресурсов в обследованном массиве было выявлено не менее 300. К ним относятся следующие виды АИС:

- фактографические, аналитические базы и банки данных;
- ГИС и атласы;
- электронные научные издания, в том числе электронные представления книжных и рукописных памятников;
- экспертные системы и базы знаний;
- инфометрические системы и индексы цитируемости;
- виртуальные модели и 3D-ресурсы.

¹ В 2020 г. указатель неопубликованных документов был существенно дополнен и включает в настоящее время около 1,5 тыс. описаний ресурсов.

Данная классификация является достаточно условной, и в отдельных случаях выделенные категории информационных ресурсов могут пересекаться.

Лингвистические ресурсы более подробно рассматриваются в главе 6.

Информация о научных событиях

Всего выявлено свыше 130 перечней конференций, семинаров, круглых столов и других научных событий. Следует указать, что в большинстве случаев представлены либо анонсы мероприятий с краткой информацией о них, либо программы с наименованиями докладов и составом участников. Полные тексты докладов приводятся не более чем в 10% случаев, совсем редко – стенограммы или видеозаписи научного события.

Персональные ресурсы

Всего выявлено свыше 100 перечней персональных ресурсов. Они включают биографические справки и перечни научных трудов, сведения о наградах, проектах, грантах членов академии, сотрудников учреждений, членстве ученых в различных советах, комиссиях, редакционных коллегиях и других научных институциях. Среди них следует упомянуть мемориальные ресурсы, посвященные отдельным ученым и сотрудникам научных учреждений.

Медиаресурсы

Всего выявлено 85 коллекций фото-, аудио-, кино- и видеоресурсов. Среди аудиоресурсов выделяется Национальный звуковой депозитарий, созданный при Пушкинском доме, который объединяет ряд аудиокolleкций фольклорного и этнографического характера.

Прочие ресурсы Интернета

Интернет-ресурсы академических учреждений, не отнесенные к описанным выше типам, могут быть разделены на такие виды:

- сайты научных учреждений;
- сайты подразделений;
- сайты-сателлиты;
- аккаунты в социальных сетях;
- размещение результатов научных исследований учреждений на общих информационных ресурсах;
- каталоги интернет-ссылок.

Всего сайтов институций отражено 202 (не обнаружено сайтов у 12 учреждений), сайтов подразделений научных учреждений – 100, сайтов или страниц журналов – 403 (значительная часть страниц журналов размещены на доменах третьего уровня сайтов учреждений-издателей).

Других сайтов, обычно называемых сателлитами, т.е. самостоятельных сайтов, связанных с сайтом основной институции, обнаружено 95. Это сайты проектов, аффилированных институций, персональные сайты сотрудников и др. В 41 случае указаны аккаунты институции в социальных сетях. Кроме того, выявлено 29 каталогов ссылок на интернет-ресурсы.

Таким образом, сеть академических сайтов по социогуманитарным наукам включает примерно 850 сайтов или отдельных страниц, которые должны быть отражены в навигационной системе по информационным ресурсам. Причем это число должно расти, потому что, например, стремление редакций журналов быть включенными в базу данных Scopus обязывает создавать самостоятельные сайты, поскольку наличие такого сайта является императивным требованием для Scopus.

3.10. Навигатор информационных ресурсов по общественным наукам

В настоящем разделе описывается информационная система, разработанная для того, чтобы сделать результаты мониторинга доступными общественности, а также обеспечить возможность постоянного поддержания актуальности полученного каталога. Описываемая информационная система получила название Нави-

гатор информационных ресурсов по общественным наукам (НИРОН) и размещена в Интернете [60].

Логическая структура НИРОН

Логическую структуру НИРОН составляют две сущности, называемые *владельцы* и *информационные ресурсы*.

Владельцы – это академические институты. Прежде всего, конечно, это научные учреждения. Также к владельцам могут относиться самостоятельные подразделения учреждений, органы управления, аффилированные институционализированные структуры, например научные общества, комиссии, советы, имеющие собственные ресурсы.

Информационные ресурсы – организованные совокупности документов или данных, частично институционализированные, но в большинстве случаев представляющие собой просто поименованный информационный массив (коллекцию, фонд, сайт или страницу сайта, материалы конференции и др.).

Эти сущности связаны отношениями *владения*, т.е. каждый *информационный ресурс* принадлежит *владельцу*. Для каждого ресурса в НИРОН указан его владелец, а для каждого владельца – множество принадлежащих ему ресурсов. В НИРОН используются также более привычные названия для владельцев: для мероприятий – организатор, для периодики – учредитель, для библиотек, архивов, музеев – вышестоящая организация.

Теоретически это разделение не вполне корректно, поскольку часть ресурсов – библиотеки, архивы музеев – институционализированы в виде структурных подразделений и даже юридических лиц, т.е. могут быть владельцами. Периодика также институционализирована через систему регистрации, а также благодаря наличию у периодического издания таких институций, как издатель, учредитель, редакционная коллегия и редакционный совет. Большинство же других видов ресурсов, например сайт, электронная библиотека или иная информационная система, не имеют собственных институций, а просто поименованы их создателями.

Автор вполне сознает логическую противоречивость предложенной конструкции, однако убежден, что она соответствует интуитивному представлению как создателей, так и пользователей информационных продуктов и услуг, а потому практически приемлема.

Навигация по организационной принадлежности. Владельцы образуют структуру в соответствии с существующей иерархией институций внутри РАН. Иерархическая связь указывается в реквизитах владельца – *вышестоящая организация и структура*. Учитывая двойную (территориально-отраслевую) структуру РАН, владелец может иметь две вышестоящие организации. Вершиной иерархии является Президиум РАН. Подчинение структурным подразделениям ФАНО, а затем Минобрнауки, в НИРОН не учитывалось¹. В некоторых случаях для удобства навигации вводились искусственные структуры, например *Музеи РАН по естественным наукам*. Всего в НИРОН в настоящий момент включено свыше 200 владельцев.

Организационная принадлежность ресурсов является первой иерархией НИРОН. Соответственно, первая очевидная навигационная возможность – найти ресурсы данного учреждения, ресурсы регионального или отраслевого отделения РАН и других организационных категорий учреждений РАН.

Функциональная типология информационных ресурсов. Основной иерархией информационных ресурсов в НИРОН является типология, разработанная по результатам мониторинга и в упрощенном виде использованная в НИРОН. Содержательные вопросы разработки этой типологии рассмотрены выше в разделе 3.6. Эта типология представлена в табл. 6.

При систематизации по данной типологии конкретных ресурсов, включенных в НИРОН, принято несколько условных решений, которые могут быть признаны спорными. Например, мультимедийные ресурсы отнесены к информационным системам, а коллекции медиадокументов выделены отдельно. В самостоятельный класс выделены лингвистические ресурсы, хотя во многих случаях они также могут быть определены как информационные системы (словарные БД) или как электронные коллекции (это относится к корпусам). Личные архивные фонды отнесены к персональным ресурсам, а архивы описаны в целом. Отдельные архивные фонды включены в НИРОН, если эти данные были представлены владельцами ресурсов. Электронные представления книжных и рукописных памятников, созданные с учетом правил кодикологии, отнесены к информационным системам, а не к электронным библиотекам. Можно обсуждать обоснованность этих и

¹ Это решение оказалось провидческим: при ликвидации ФАНО в мае 2018 г. принадлежность учреждений РАН департаментам ФАНО исчезла.

других типологических решений, однако следует иметь в виду, что главная задача предлагаемой типологии – предложить пользователям для навигации привычные, хотя и не всегда строго логичные типы и виды ресурсов.

Таблица 6

Типология информационных ресурсов в НИРОН

Типы ИР	Виды ИР
1	2
Библиотеки	
Архивы	
Музеи	
Каталоги	Традиционный Электронный Смешанный
Электронные коллекции и библиотеки	Публикации учреждений Электронные библиотеки Смешанные коллекции Репозитории ЭБ на переносимых коллекциях
Информационные системы	Информационно-аналитические БД ГИС Индексы цитирования Инфометрические ресурсы Электронные представления памятников Экспертные АИС и базы знаний Мультимедийные, 3D-, VR-системы Комплексные АИС
Справочники, энциклопедии	Энциклопедии Справочники Хронологии Фонды учреждений Архивные разряды
Персональные ресурсы	Персональные сайты и страницы Личные фонды Сотрудники учреждений Указатели лиц Биобиблиографии

Продолжение таблицы 6

1	2
Лингвистические ресурсы	Корпуса текстов Словарные БД и электронные картотеки Лингвистические процессоры Грамматические ресурсы Описания языков, реестры языков Лингвистические атласы Этно- и социолингвистические БД Комплексные лингвистические АИС (сайты) Информационные языки
Периодика	
Библиографии	Перечни трудов учреждений / сотрудников Аннотированные или реферативные указатели Предметные, тематические библиографии Библиографические БД
Мероприятия	
Неопубликованные документы	Диссертации Отчеты Экспертные заключения Гранты, проекты, экспедиции Очерки деятельности учреждения, подразделения
Медиаресурсы	Банки изображений и фото Аудиоресурсы Видеоресурсы Киноматериалы Смешанные ресурсы
Прочие интернет-ресурсы	Сайты подразделений Сайты-сателлиты Аккаунты в социальных сетях Ресурсы во внешних ЭБ и АИС Каталоги ссылок

Описание ресурсов. Все ресурсы при описании имеют стандартный набор из четырех реквизитов:

- владелец;
- название ресурса;
- адрес сайта;
- код ГРНТИ.

Ресурсы некоторых типов имеют дополнительные специфические реквизиты. Реквизиты библиотек используются в соответствии с отчетностью Информационно-библиотечного совета РАН. Данные по архивам приводились в соответствии со сведениями в ИСА РАН. Отчетность по музейным фондам РАН была недоступна.

Перечень специфических реквизитов отдельных типов ИР представлен в табл. 7.

Таблица 7

Специфические реквизиты типов ресурсов

Типы ИР	Специфические реквизиты
Библиотеки	Входит в ЦБС Описание фонда Объем фонда Книги всего Книги иностранные Периодика всего (вып.) Периодика иностранная (вып.) Редкий фонд
Архивы	Число фондов Число описей Указатели, путеводители
Музеи	Описание фонда
Периодика	Форма представления Период издания и хранения
Мероприятия	Сведения о мероприятии

Тематическая навигация. Кроме организационной и функциональной, в НИРОН организована тематическая навигация. Для этой цели используется Государственный рубрикатор научно-технической информации (ГРНТИ). Каждому ресурсу присвоено от одного до трех индексов (кодов) ГРНТИ глубиной от первого до третьего уровней. Однако тематическая организация НИРОН, так же как и применение ГРНТИ для описания тематики конкретных ресурсов, встречает ряд методических сложностей.

Во-первых, значительная часть ресурсов, в том числе ресурсов региональных научных центров и библиотек, не имеют явной тематической привязки. Им присваивался код *00.29 Информационная деятельность в общественных науках*.

Во-вторых, много важных для общественных наук направлений и категорий искусственно разделены в ГРНТИ, что требует постоянного использования комбинаций рубрик. Например, *Латинскую Америку* нужно индексировать комбинацией рубрик *23.73 Центральная Америка* и *23.75 Южная Америка*. Другой пример: в ГРНТИ отсутствует *Филология*, которую нужно заменять комбинацией *16. Языкознание* и *17. Литературоведение*.

Отсутствует много важных для отечественных общественных наук рубрик. Например, отсутствует важнейшая для отечественного языкознания рубрика *Русский язык*. Вообще в ГРНТИ отсутствует комплекс дисциплин, которые можно обозначить как *Народоведение*, например *Тюркология* или *Финно-угорика*.

С другой стороны, некоторые рубрики неоправданно детализированы. Для множества рубрик, особенно по разделам *06 Экономика и экономические науки* и *10 Государство и право. Юридические науки* адекватных ресурсов не нашлось.

В НИРОН вошло много универсальных по тематике и широкоотраслевых ресурсов, для которых критическим является принятое ограничение в три индекса. Поэтому мы были вынуждены часто использовать рубрику *00.29 Информационное обеспечение общественных наук*, которая заменяла разные варианты универсальной и широкоотраслевой тематики.

В целом проведенное индексирование академических ресурсов по ГРНТИ показало, что для предложенных целей ГРНТИ нужно существенным образом переработать. Также, вероятно, должна быть модернизирована и методология тематического индексирования. Эта задача решена применительно к лингвистическим ресурсам, о чем сказано в главе 6.

Функциональность НИРОН. Функции НИРОН можно разделить на первоочередные, реализованные в настоящее время, и перспективные, необходимость в которых должна быть установлена в ходе опытной эксплуатации. К первым относятся:

- предоставление права администрирования НИРОН;
- добавление и удаление владельцев;
- добавление и удаление ресурсов;
- поиск и визуализация ресурсов, принадлежащих данному владельцу;
- сортировка ресурсов по организационной принадлежности;
- поиск и визуализация ресурсов, относящихся к данному типу;
- сортировка ресурсов по видам;
- сортировка ресурсов по кодам ГРНТИ.

К перспективным можно отнести следующие функции:

- поиск ресурсов по лексике наименований с учетом морфологии;
- статистика ресурсов по различным категориям владельцев;
- предоставление права администрирования конкретного раздела НИРОН;
- добавление новых типов и видов ресурсов;
- добавление новых реквизитов ресурсов.

Ограничения контента НИРОН. Контент, включенный в НИРОН в настоящее время, имеет ряд ограничений, установленных отчасти из ресурсных соображений, отчасти из нецелесообразности дублирования НИРОН с другими информационными системами, отчасти из сомнений в полезности включения в НИРОН определенных категорий ресурсов. Перечислим принятые ограничения.

В НИРОН отсутствуют наименования архивных фондов и описей, в том числе представленных на портале Mnemosyne [131] Архива РАН, где имеются сведения о 5,3 тыс. фондов и 7,4 тыс. описей¹. В НИРОН мы ограничились описаниями академических архивов в целом. Это решение можно рассматривать как временное. В перспективе выбор способов включения архивных фондов в НИРОН является частным случаем более общей проблемы – насколько архивная система должна быть самостоятельной или должна быть интегрирована в общее пространство научной информации; и если должна быть интегрирована, то на каком уровне. Пока в НИРОН включены только те архивные фонды (как правило, личные), о которых заявили владельцы ресурсов при проведении опроса. Дублировать существующую систему Архива РАН на данном этапе признано нецелесообразным.

В НИРОН не включены коллекции диссертаций, размещенных на сайтах всех академических организаций, имеющих диссертационные советы, однако имеется информация о материалах диссертационных советов, в которых могут быть найдены коллекции диссертаций, а также сопутствующие документы. Кроме того, как известно, все защищенные в России диссертации представлены в Электронной библиотеке диссертаций в РГБ.

В НИРОН не включены ссылки на научные отчеты, представленные академическими организациями в ЦИТИС (ЕГИСУ). Это можно было сделать на уровне отсылок к конкретным документам, что противоречит идеологии НИРОН. В целом было установлено, что свои отчеты представили менее 70 учреждений, т.е. около 1/3 владельцев. Однако в НИРОН имеются описания коллекций отчетов, размещенные на сайтах академических организаций.

Ссылки на персональные ресурсы ограничены коллекциями, страницами сайта и базами данных; хотя многие академические

¹ На портале на момент написания книги отсутствовали сведения об архивах, имеющих в бывших РАСХН и РАМН.

организации включают персональные страницы для каждого из сотрудников.

Библиографические списки научных коллективов ограничены уровнем учреждений, в то время как они часто составляются на уровне научных лабораторий и отдельных ученых.

Описания мероприятий (конференций, конгрессов, семинаров и др.) представлены обобщенно, на уровне ссылок на сведения обо всех мероприятиях, проводимых учреждениями. Следует учесть, однако, что в России имеется несколько порталов, в которых сведения о научных мероприятиях представлены на индивидуальном уровне.

В НИРОН не включены ресурсы типа «СМИ о нас», ленты новостей и другие сообщения, характерные для СМИ и лишь косвенно связанные с научными результатами учреждений.

Перспективы развития НИРОН

Частично перспективы развития НИРОН уже намечены в предыдущих разделах. Конечно, возможное и желательное направление расширения НИРОН – это выход за пределы организаций РАН и создание общего каталога российских научных информационных ресурсов по общественным наукам, включая ресурсы университетов, прикладных и отраслевых научных учреждений, а также ресурсы общественных организаций. Полезность и востребованность такого каталога не вызывает сомнений. Однако такой широкий подход существенно выходит за рамки возможностей разработчиков, к тому же он требует новой методической проработки.

Таким образом, контент, включенный в НИРОН в настоящее время, может быть существенно расширен. В настоящее время НИРОН включает примерно 4 тыс. описаний ресурсов, в том числе основные сайты владельцев. Экспертная оценка показывает, что включение в него всех ресурсов РАН на предлагаемом уровне глубины может составить до 15–20 тыс. объектов. Если же говорить обо всех российских ресурсах по общественным наукам, то можно их число увеличить на порядок.

Как и для всех каталогов ресурсов Интернета, для НИРОН необходима регулярная проверка актуальности и достоверности данных. Возможно и другое развитие функциональности НИРОН.

Однако наиболее важные перспективы НИРОН связаны, на наш взгляд, с перспективами создания Единого российского электронного пространства знаний (ЕРЭПЗ). Эти перспективы будут подробно рассмотрены в пятой главе.

Очевидно, что учет информационных ресурсов как для проектирования ЕРЭПЗ, так и для других задач управления инфосферой должен осуществляться комплексно, что относится к различным типам ресурсов. Однако конкретные задачи развития НИРОН должны определяться контурами проекта ЕРЭПЗ или его аналога как на уровне ведомства, так и в масштабах страны.

3.11. О путях оптимизации академических социогуманитарных информационных ресурсов

Выше представлены основные результаты мониторинга информационных ресурсов, созданных в учреждениях РАН социогуманитарного профиля. Далее делается попытка обсудить возможные направления оптимизации существующего состояния информационных ресурсов академической инфосферы с учетом современных тенденций ее развития, рассмотренных в главе 1.

Рассмотрим направления и способы оптимизации каждого из типов ресурсов, выделенных в предложенной типологии.

Под способами оптимизации мы будем понимать разработку и реализацию централизованных сервисов, обеспечивающих координацию информационных процессов, исключение дублирования, интеграцию ресурсов, их агрегацию, навигацию в распределенном пространстве, и другие известные технологии.

Библиотеки. В этом традиционном секторе инфосферы обсуждение способов и направлений оптимизации сложившейся ситуации в библиотечной сфере в целом происходит весьма оживленно и на профессиональном уровне. Поэтому мы ограничимся выражением мнения, что реорганизация РАН, объединение РАН с другими государственными академиями и включение академических библиотек в состав Министерства науки высшего образования отнюдь не способствовали улучшению управления академическими библиотеками.

Укажем на следующие факты. Существовавший долгие годы Информационно-библиотечный совет РАН ликвидирован, а его аналог не создан. В РАН и Минобрнауки нет даже намека на какие бы то ни было планы или концепции развития академической биб-

библиотечной сети. Отсутствует система учета библиотечной деятельности и библиотечной статистики. До сих пор не решены правовые вопросы существования библиотек – филиалов в институтах, не определена судьба библиотек бывшей РАМН. Фактически полностью прекращено финансирование проектов по развитию и совершенствованию библиотек. Один из наиболее видимых факторов – это отсутствие управляющей структуры или головной организации в библиотечной сети Минобрнауки и РАН.

Некоторые полезные соображения о направлениях оптимизации сети академических библиотек содержатся в работах Н.Е. Каленова [50] и Т.С. Маркаровой [55], а также в предложениях по созданию Единой системы библиотек ФАНО [22].

Архивы. В развитии архивной сети РАН ситуация заметно лучше, что, видимо, связано с активностью головной организации архивной сети РАН – Архива РАН. Создана информационно-управляющая система ИСА РАН [131], которая дает достаточно полное представление о состоянии Архивного фонда РАН, описей и каталогов в отдельных архивах и предоставляет другую полезную информацию. Однако в архивной сети РАН также накопилось множество проблем, требующих своего решения, особенно в части размещения архивов и сохранности архивных фондов [63]. Примером нерешенной проблемы служит тот факт, что с 2013 г. не была проведена адаптация архивов бывших РАМН и РАСХН в общую систему архивных фондов РАН.

Музеи. В области организации музейного дела в РАН также следует констатировать значительную деградацию по сравнению с периодом, предшествовавшим реорганизации РАН в 2013 г.

Музейный совет РАН ликвидирован, отсутствует система учета музейных фондов, и даже сам перечень существующих музеев РАН существенно различается в разных источниках. Автору неизвестны никакие проекты по развитию музейного дела в РАН в целом. В частности, деятельность по созданию электронного каталога Музейного фонда РФ в части РАН то ли приостановлена, то ли прекращена. На этом печальном фоне положительно выделяется деятельность по организации музейного дела в Сибирском отделении РАН [84].

Реферативные журналы институтов научной информации (ВИНИТИ и ИНИОН). Реферативные журналы научной информации, бывшие в свое время бесспорными лидерами российских научных коммуникаций, в настоящее время по многим причинам утратили лидирующую роль. Хотя реферативные жур-

налы продолжают оставаться крупнейшими в РАН научно-информационными продуктами, их перспективы должны обсуждаться в контексте создания общей информационной системы РАН и, конечно, Единого российского электронного пространства знаний. Более подробно эти вопросы обсуждаются в главе 5.

Электронные библиотеки (ЭБ). Этот тип ресурсов является самым востребованным в социогуманитарной научной сфере. Практически все научные учреждения размещают на своих сайтах изданные в этих учреждениях научные произведения. Также активно создаются и комплексные ЭБ – такие как Соционет [102], «Научное наследие России» [66]. Следует особо отметить крупные электронные библиотеки в области филологии – Фундаментальную электронную библиотеку «Русская литература и фольклор» (разработка ИМЛИ РАН и НТЦ «Информрегистр») [109], АИС «Русская словесность» (разработка ИРЛИ РАН и компании Альтсофт) [46].

Всего в академических учреждениях социогуманитарного профиля создано свыше 200 ЭБ, включающих сотни тысяч произведений – книг, статей, словарей, диссертаций, библиографий и других видов документов. Эти ЭБ существенно отличаются между собой по качеству оцифровки, используемым метаданным, форматам представления, методике комплектования, структуре, используемому программному обеспечению и другим параметрам. Есть вопросы по легитимности многих ЭБ.

Поэтому о механической интеграции существующих ЭБ, в том числе в рамках проекта Национальной электронной библиотеки, не может быть и речи. Задача состоит в разработке централизованных сервисов, которые смогли бы, во-первых, мотивировать разработчиков на координацию и коллаборацию, а во-вторых, направить вектор развития академических ЭБ в направлении единого электронно-библиотечного пространства. Перспективным являются такие сервисы, как координация оцифровки (сервис разработан в рамках ЭБ «Научное наследие России»), позволяющие исключить или хотя бы сократить дублирование работ по оцифровке. Конечно, для такой деятельности необходимы стандарты по оцифровке, учитывающие особенности различных категорий документов. В этом отношении существенный задел создан Президентской библиотекой им. Б.Н. Ельцина, в которой разработан комплекс методических материалов по оцифровке.

Очевидным результатом деятельности по координации оцифровки должен быть единый каталог оцифрованных докумен-

тов, который должен стать центральным компонентом каталога-навигатора, о котором шла речь выше.

Из других возможных сервисов поддержки ЭБ можно рассмотреть предоставление разработчикам бесплатного лингвистического и программного обеспечения, правовую поддержку и др. Необходимо выработать систему количественных и качественных показателей ЭБ, позволяющих определить критерии для их поддержки и архивации.

Интересным направлением развития научных ЭБ является их превращение в репозитории, такие как Соционет, предоставляющие услуги по размещению научных произведений, их рецензированию, включению в инфометрические процессы, минуя традиционные издательские процессы.

Безусловно, целесообразно централизовать систему архивного хранения ЭБ.

Очевидно, что разработка централизованных сервисов для ЭБ в РАН должна учитывать практику интеграционных электронно-библиотечных проектов в России и за рубежом, таких как Национальная электронная библиотека, Президентская библиотека им. Б.Н. Ельцина, Europeana и др., и предусматривать взаимодействие с ними.

Возникает, однако, вопрос, о целесообразности создания самостоятельной системы развития академических ЭБ при наличии проектов национального и международного уровней.

Автор придерживается мнения, что, учитывая сложившуюся в России систему финансирования информационных проектов и управления ими, а также наличествующую структуру информационных учреждений РАН, прямая интеграция академических электронных библиотек в НЭБ нецелесообразна и определенную самостоятельность электронно-библиотечных сервисов в РАН следует сохранить. Таким образом, имеет смысл проектировать либо интегрированную электронную библиотеку РАН, либо электронные библиотеки РАН по пяти основным научным областям: естественные науки, общественные науки, технические науки, сельскохозяйственные науки, медицинские науки.

Особое место в системе академических ЭБ должна занять ЭБ «Научное наследие России» [66], организованная по персональному принципу наследия российских ученых, перешедшего в общественное достояние. Важно, что в этой ЭБ реально осуществлена интеграция библиотечных, архивных и музейных ресурсов, а также практически реализована коллаборация многих учреждений РАН.

Информационные системы. Большинство информационных систем, реализованных в учреждениях РАН социогуманитарного профиля, являются уникальными, поэтому прямая их агрегация и тем более интеграция нецелесообразны. Наверное, мог бы быть полезен централизованный академический сервис для размещения типовых АИС, таких как ГИС, банки изображений, стандартные учетные БД. Имеет смысл централизовать инфометрические и вебометрические сервисы.

Однако в любом случае необходимы учет таких систем и отражение их в качестве научных результатов при отчетности научных учреждений. Однако для этого нужно разработать некоторые показатели для оценки качества и объема таких АИС, вероятно, как функцию от трудоемкости затрат на разработку и сопровождение АИС.

Справочные и энциклопедические ресурсы. Данная категория информационных ресурсов предоставляет особые возможности для интеграции и сотрудничества академических учреждений и отдельных ученых.

С одной стороны, опыт Википедии показывает, какого эффекта можно достичь путем коллаборативного сбора справочно-энциклопедической информации на основе современных информационных технологий. В настоящее время Википедия, бесспорно, является наиболее посещаемым информационным ресурсом данного класса

С другой стороны, именно академические учреждения обладают наибольшим опытом и авторитетом в данной сфере информационной деятельности. Недаром основную роль в редакционных советах Большой советской, а затем Большой российской энциклопедии (БРЭ) играли и играют представители Российской академии наук. Как указал главный редактор и председатель научно-редакционного совета БРЭ акад. Ю.С. Осипов, в подготовке последней редакции БРЭ принимали участие 120 академиков РАН.

Действительно, проведенный мониторинг информационных ресурсов академических учреждений социогуманитарного профиля показывает значительную роль этих учреждений в создании справочно-энциклопедической информации. Обнаружено свыше 120 таких ресурсов, включая столь заметные, как *Лингвистический энциклопедический словарь (ЛЭС)*, *Энциклопедия «Языки мира»*, *Литературная энциклопедия терминов и понятий*, *Философская и Новая философская энциклопедии*, *Электронная энциклопедия «Дискурсология»*, *Татарская, Башкирская* и другие региональные

энциклопедии, а также множество справочников по различным направлениям научно-организационной деятельности. Особым типом энциклопедических ресурсов являются хронологии и летописи, такие как *Хронология главных событий отечественных гуманитарных наук с XVIII по XXI в.*, *Хроника российско-американских отношений*, *Летопись литературных событий в России конца XIX – начала XX в.*, *Летопись первой мировой войны* и мн. др.

Причем в число этих ресурсов не входит огромное количество персональных сайтов и страниц, биографических справок, биобиблиографий, различных указателей и словарей, которые также могут быть использованы при формировании единого российского справочно-энциклопедического пространства. Практически все академические учреждения социального и гуманитарного направления имеют соответствующие ресурсы и заделы по созданию таких ресурсов.

Как уже отмечалось, современные информационные технологии предоставляют прекрасные возможности для коллективной работы над энциклопедиями. Кроме упомянутой технологии *вики* разработано много технологий в рамках идеологии семантического веба, среди которых наиболее перспективной для указанных целей представляется технология связанных открытых данных (Link Open Data – LOD).

Важной инициативой для организации коллективной работы академических организаций по формированию электронной энциклопедии являются распоряжения Правительства РФ по созданию «Общенационального научно-образовательного интерактивного энциклопедического портала» на базе Большой российской энциклопедии с привлечением других российских научных энциклопедий [86; 95]. Подробнее об этом см. главу 4.

Очевидная и актуальная задача состоит в организации коллаборации академических учреждений и соответствующих специалистов для создания, поддержки и развития справочно-энциклопедических ресурсов.

Персональные (просопографические) ресурсы. Этот вид информационных ресурсов является, с одной стороны, разновидностью справочно-энциклопедических ресурсов, поскольку сведения о лицах (персонах) – неперемнная часть энциклопедических данных. С другой стороны, эти ресурсы обладают несомненной спецификой как с точки зрения организации данных, так и с правовой точки зрения, так как сведения о лицах подпадают под действие законодательства о персональных данных.

Центральный вид персональных ресурсов – это личные архивные фонды, которые составляют большинство среди архивных фондов РАН (свыше 4 тыс. из 5,3 тыс. фондов). Большую популярность получили в последние годы персональные сайты и страницы ученых, а также персональные аккаунты в социальных сетях.

Но, кроме того, учреждения РАН в заметных количествах создают различные указатели и справочники сведений персонального характера. Это справочники академиков, членов-корреспондентов РАН, отделений РАН и республиканских академий наук, профессоров РАН, докторов, а также сотрудников академических учреждений. На многих сайтах есть мемориальные страницы. Достаточно много биографических и биобиблиографических словарей и указателей, справочников «Кто есть кто», соответствующих баз данных и др. Примером может служить упомянутая выше электронная библиотека «Научное наследие России».

Другой пример – Personalia – база данных, созданная в рамках Фундаментальной электронной библиотеки «Русская литература и фольклор» и содержащая биобиблиографические материалы о жизни и творчестве русских ученых-гуманитариев [109].

Некоторые указатели и справочники носят уникальный характер, например «Картотека Резуна» персоналий сибирских первопроходцев XVII–XVIII вв., знаменитая «Картотека Б.Л. Модзалевского», Сводный каталог всех корреспондентов А.М. Горького или Каталог египетских личных имен Среднего царства.

Всего в результате мониторинга было обнаружено свыше 200 ресурсов персонального характера, созданных в учреждениях РАН, не считая архивных фондов, персональных сайтов и аккаунтов ученых в социальных сетях.

Кажется достаточно очевидным, что для этого множества ресурсов следует создать централизованную навигационную систему, внутри которой можно было бы искать во всех этих материалах сведения о любой персоне, с учетом различных написаний имен, хронологии, а также некоторых важнейших характеристик персон. Более отдаленной перспективой могло бы стать единое пространство сведений о персонах, учитывающее, конечно, требования законодательства о персональных данных и объединяющее различные достоверные источники.

Периодические, продолжающиеся, сериальные издания.

Ситуация с периодическими изданиями академии наук – пожалуй, единственная сфера информационной деятельности РАН, которая

находится в центре внимания как руководства наукой, так и научной общественности. Нужно отметить, что научные журналы пока еще остаются важнейшим и по количественным показателям, и по авторитетности каналом распространения научной информации. В исследуемом нами фрагменте научного информационного пространства зафиксировано не менее 400 периодических, продолжающихся и сериальных изданий, причем наблюдается активный рост числа журналов. Значительная доля научной периодики интегрирована в широко известных электронных библиотеках eLIBRARY [64] и КиберЛенинка [65].

Интерес к научным журналам в значительной мере связан с внедрением в управление наукой библиометрических показателей, прежде всего публикационной активности и индекса цитируемости. При этом российские управленцы преимущественно ориентированы на учет публикаций в периодических изданиях, индексируемых в известных наукометрических БД Web of Science и Scopus. Поэтому среди российских научных журналов началось движение за включение в эти БД.

Для содействия этому стремлению создана и активно действует некоммерческая организация Ассоциация научных редакторов и издателей (АНРИ), возглавляемая О.В. Кирилловой [24]. Деятельность этой ассоциации, как и в целом активность научно-журнального сообщества, безусловно, способствует повышению качества научных журналов, соблюдению мировых стандартов по оформлению публикаций и в целом расширению присутствия российских научных изданий в международном научно-информационном пространстве.

Однако это движение вряд ли можно назвать оптимизацией данного класса информационных ресурсов и магистральным путем развития научных коммуникаций. Дело в том, что погоня за публикационной активностью и цитируемостью в принципе отрицательно сказывается на качестве научной информации. Прежде всего, важно отметить опасность появления «хищнических» или «мусорных» журналов, готовых за плату печатать все что угодно. Еще одно негативное явление – «cutting salami» – стремление разделить научный результат на множество отдельных публикаций. Сюда же можно отнести безудержный рост числа соавторов научных публикаций, самоцитирование и перекрестное цитирование и т.п. В результате падает качество и снижается авторитет как самих научных журналов, так и наукометрических БД. Так, новая британская система оценки эффективности научной деятельности Research

Excellence Framework [140] уже отказывается от использования Web of Science, что напрямую связано с коммерциализацией этого сервиса. В общем, очевидным образом наблюдается действие известного закона Гудхарда¹.

Мнение мирового сообщества по поводу перспектив развития научной периодики в цифровой среде в условиях открытой науки и важная роль принятых показателей оценки эффективности научных учреждений достаточно подробно рассмотрены в главе 1.

Библиографии. Вероятно, наиболее распространенным видом академических информационных ресурсов, особенно в гуманитарных и социальных науках, являются библиографические перечни, указатели, БД и др. Причины этого очевидны: библиография – самый наглядный способ представить результаты отдельного ученого, научного коллектива или состояние данной науки, дисциплины, предмета исследований. В результате проведенного мониторинга выявлено до 400 ресурсов этого типа, причем во многих случаях мы объединяли в один библиографический ресурс множество библиографических перечней, сотрудников научного коллектива.

Можно не доказывать факт дублирования труда научного сообщества, затраченного на составление этих перечней. Очевиден также способ оптимизации библиографической деятельности – нужен централизованный сервис, позволяющий любому ученому или информационно-библиотечному работнику в своем личном кабинете отражать библиографию своих трудов, а затем экспортировать их для различных нужд, например для отчетности, заявок на гранты и др. Такие сервисы уже давно действуют, в том числе на упоминавшейся платформе Соционет. Задача заключается в том, чтобы мотивировать научное сообщество использовать централизованные сервисы, а не тратить силы и средства для поддержания собственных перечней и БД. Начальным этапом создания централизованного библиографического сервиса могла бы стать агрегация библиографий, размещенных на сайтах научных учреждений, с указанием аффилиации авторов, тематики и других полезных реквизитов сводной библиографической БД. Понятно, что такая БД могла бы стать важным компонентом единой информационной системы РАН, а в перспективе – ЕРЭПЗ.

¹ Закон (принцип) Гудхарта заключается в том, что когда экономический показатель становится целью для проведения экономической политики, прежние эмпирические закономерности, использующие данный показатель, перестают действовать.

Мероприятия (конференции, конгрессы, симпозиумы, выставки). Данный ресурс также является весьма массовым. По каждой гуманитарной и социальной науке ежегодно проводятся сотни научных мероприятий, в большинстве из которых принимают участие ученые академических учреждений. Достаточно много мероприятий организуют и сами учреждения РАН.

В последние годы число научных конференций сильно возросло за счет междисциплинарных мероприятий, как правило, удаленных и предназначенных прежде всего для молодых ученых, особенно магистров и аспирантов, которым необходимы выступления на конференциях для подтверждения квалификационных требований.

К сожалению, многие из этих конференций практически отказались от рецензирования для включения в сборники трудов докладов или тезисов. Соответственно, качество и оригинальность этих докладов никак не контролируются. Известно, что в настоящее время РИНЦ исключает из своих массивов много сборников трудов именно по этой причине.

С другой стороны, в Интернете уже реализовано множество сервисов, поддерживающих проведение научных конференций, содержащих каталоги этих конференций, а также осуществляющих публикацию трудов¹. Создание еще одного ресурса для конференций РАН может быть оправданным, только если будут предложены принципиально новые сервисы, например квалифицирующие не оригинальные и некачественные доклады.

Неопубликованные документы. Традиционно к неопубликованным видам научной информации относились диссертации и научные отчеты. Обработка этих видов документов давно регламентирована, и созданы соответствующие централизованные ресурсы, в том числе в РГБ [114] и в ЦИТИС [43]. Мы сейчас не будем обсуждать возможности улучшения этих ресурсов и облегчения доступа к соответствующим видам документов, поскольку это проблема не академическая, а общегосударственная.

Заметим, однако, что научные учреждения РАН размещают на своих сайтах много другой интересной и полезной неопубликованной научной информации, в частности свои планы и программы исследований, описания проектов, экспедиций, результаты исследований по грантам, экспертные заключения и другие документы.

¹ См., напр.: <http://konferencii.ru/>, <http://konferen.ru/>, www.konferenc.ru

Интересным жанром неопубликованной информации являются очерки деятельности учреждений и научных подразделений, которые имеются практически на каждом сайте. Агрегация этой информации и сопоставление этих очерков дали бы ясную картину российской академической науки, в том числе тематического дублирования и имеющихся лакун.

В последнее время ИНИОН РАН проявил инициативу по интеграции неопубликованных научных документов и предложил учреждениям РАН создать на основе коллаборации интегрированный репозиторий открытого доступа по общественным наукам (РОДОН). Содержание этих предложений излагается ниже.

РОДОН предлагается создавать как важный автономный фрагмент ЕРЭПЗ, взаимодействующий с Национальной электронной библиотекой. В составе РОДОН предполагается размещение следующих типов научных материалов, подготовленных научными учреждениями РАН:

- статьи, монографии, сборники статей открытого доступа;
- препринты и электронные публикации;
- программы конференций, симпозиумов, семинаров, вебинаров и других научных мероприятий, а также презентации и тезисы докладов;
- научные отчеты, в том числе по исследованиям, выполненным по грантам российских и зарубежных научных фондов;
- отчеты об экспедициях, командировках, другие документы, отражающие результаты деятельности научных учреждений;
- сведения об участии научных учреждений в международных проектах и программах.

Научные материалы, имеющие признаки служебных произведений, должны размещаться в РОДОН только с согласия научных учреждений РАН как правообладателей. Электронные документы могут передаваться в РОДОН в общепринятых форматах без дополнительной обработки.

ИНИОН РАН при размещении научных материалов в РОДОН будет производить следующие операции:

- присвоение уникального национального идентификатора (вместо зарубежного коммерческого проекта DOI);
- систематизацию документов в соответствии с Государственным рубрикатором научно-технической информации;
- формирование библиографической записи в соответствии с национальными стандартами;

- формирование поискового индекса для лексического поиска по тексту;
- форматирование в соответствии со стандартами открытого доступа;
- архивирование и обеспечение долговременного хранения электронных материалов;
- размещение материалов в Интернете для открытого доступа.

Размещение электронных научных материалов в РОДОН обеспечит для научных учреждений следующие преимущества и перспективы.

1. Существенное повышение востребованности и цитируемости научных материалов учреждения, в том числе за рубежом.

2. Сокращение затрат на обработку и хранение электронных документов в научных учреждениях.

3. Использование уникального идентификатора и архивное хранение электронного документа обеспечат правовую защиту электронных документов в случае правовых споров.

4. Участие в комплектовании Национальной электронной библиотеки в форме, приемлемой для научных учреждений, и с минимальными затратами.

В перспективе РОДОН позволит обеспечить более полный и комплексный учет результатов деятельности научных учреждений РАН при подготовке отчетности и в наукометрических исследованиях.

Лингвистические ресурсы. В качестве самостоятельного вида ресурсов имеет смысл рассматривать лингвистические ресурсы, активно создаваемые филологическими учреждениями РАН. Их анализ, а также предложения по дальнейшему развитию и координации деятельности академических учреждений по созданию лингвистических ресурсов представлены в главе 6.

Фото-, аудио-, видеоресурсы. Количество ресурсов этого специфического вида относительно невелико. Всего зафиксировано около 100 таких ресурсов. Для них, конечно, должен быть создан специальный централизованный каталог, который позволит исключить дублирование и выявить все накопленные в академических учреждениях документы данных категорий. Следует учесть, что фонд аудиодокументов при Пушкинском доме имеет статус Национального электронного звукового депозитария [107]. Этот фонд содержит свыше 150 тыс. единиц хранения и является, безусловно, крупнейшим в РАН. А основная коллекция фотодокументов по истории российской науки хранится в Институте истории естествознания и техники.

Интернет-ресурсы. Интернет-ресурсы являются, безусловно, наиболее перспективным и динамично развивающимся видом информационных ресурсов.

Представляется, что на данном этапе единственным разумным вариантом развития сервисов может быть только каталогизация Интернет-ресурсов и отслеживание актуальности гиперссылок при помощи специальной программы (спайдера). В перспективе могут быть предложены новые возможности с учетом развития информационных технологий.

Понятно, что в рамках настоящей главы невозможно обсуждение всех вопросов, связанных с оптимизацией информационных ресурсов. Автор считает полезным сосредоточиться на целесообразности координации и централизации деятельности по созданию, поддержанию, предоставлению в доступ информационных ресурсов, принадлежащих учреждениям РАН социогуманитарного профиля. Под оптимизацией здесь понимается достижение максимально высоких результатов в процессах создания и эксплуатации информационных ресурсов, таких как сокращение непроизводительных расходов, улучшение пользовательских характеристик, повышение качества самих ресурсов и др. при заданных ограничениях финансового и интеллектуального характера.

3.12. Основные выводы по результатам мониторинга

Представленные в данной главе результаты мониторинга информационных ресурсов носят предварительный и неполный характер. В частности, в НИРОН отсутствуют необходимые сведения об объеме большинства ресурсов, их востребованности, а также сведения правового и технологического характера.

Однако уже сейчас можно получить некоторое представление о составе и структуре феномена, который мы называем информационными ресурсами академических организаций социогуманитарного профиля. В настоящее время речь идет о приблизительно 16 тыс. информационных массивов (архивных фондов, описей, баз данных, сайтов, каталогов, электронных библиотек, информационных систем и др.), которые содержат 60–80 млн информационных объектов таких типов, как библиографическая запись, текстовый документ, изображение, аудиозапись, словарная, энциклопедическая статья или сведения о персоне.

Эти массивы принадлежат примерно 200 учреждениям, причем распределение между ними крайне неравномерное. По мнению автора, распределение ИР по академическим институциям подчиняется закону Парето, т.е. примерно 20% (40 институций) владеют 80% ресурсов. Ведущую роль играют, конечно, профильные организации: библиотеки, архивы и институты информации. Однако некоторые научные организации, особенно археологические, филологические, лингвистические и этнографические, также обладают весьма значительными ресурсами. По музейным фондам имеющейся информации совершенно недостаточно, их состояние требует дополнительного исследования.

Обсуждая принципы развития и модернизации информационного пространства академической науки, можно ориентироваться на указанные параметры.

По результатам мониторинга можно сделать еще ряд выводов.

1. Информационные ресурсы и сервисы учреждений РАН находятся на переходном этапе процессов цифровизации – многие виды ресурсов представлены в двух формах (электронной и традиционной), некоторые – только в традиционной, некоторые – только в электронной.

2. Отсутствует как государственный, так и ведомственный учет научных информационных ресурсов, хотя их создание часто является основным результатом научной деятельности учреждений. В то же время имеется государственный или ведомственный учет некоторых категорий научных ресурсов – архивных фондов, периодических изданий, диссертаций, отчетов археологических экспедиций. Учет и каталогизация некоторых видов информационных ресурсов – сайтов, электронных библиотек, лингвистических ресурсов – ведется на добровольной основе и на любительском уровне. Централизованный учет музейных фондов и музейных каталогов РАН вообще отсутствует.

3. Официальная статистика как создания, так и потребления (использования) информационных ресурсов в РАН в принципе отсутствует, хотя существует определенная статистика, собираемая любительскими и коммерческими структурами.

4. Система сохранности информационных ресурсов отчасти реализована для традиционных информационных ресурсов (библиотечных и архивных) и полностью отсутствует для цифровых ресурсов.

5. Большинство информационных ресурсов создаются по инициативе учреждений, а также отдельных лабораторий и ученых,

никакой координации по созданию ресурсов не существует. Отсюда значительное дублирование и бессмысленная трата и так очень небольших средств, которыми располагают учреждения РАН.

6. Современные технологии коллаборации при создании и поддержании информационных ресурсов, обеспечивающие значительную экономию, в подведомственных учреждениях Минобрнауки практически не применяются.

7. Общепринятой классификации ресурсов в ведомстве нет, приблизительно одинаковые ресурсы называются по-разному, что еще более затрудняет учет и статистику.

Что же касается других задач мониторинга, прежде всего связанных с включением выявленных информационных ресурсов в единое российское электронное пространство знаний, то для их решения необходимо продолжать исследования инфосферы. Наиболее актуальным направлением представляется организация экспертной и наукометрической оценки созданных ресурсов. И, конечно, необходимо, расширить сферу инвентаризации на другие сектора общественных наук (вузовский, прикладной, общественный).

Глава 4.

ГОСУДАРСТВЕННЫЕ РЕШЕНИЯ В ОБЛАСТИ НАУЧНОЙ ИНФОРМАЦИИ (2018–2019)

4.1. О государственных решениях

Современные цифровые технологии оказывают существенное влияние на формы организации научно-исследовательской деятельности, способы распространения и обеспечения сохранности ее результатов. Поистине драматические изменения на концептуальном и технологическом уровнях происходят и в научной коммуникации. Появление социальных сетей, новые требования к экспертной оценке качества научных исследований и публикаций, обретающие все большую поддержку в научно-образовательном сообществе движения за открытый доступ и открытую науку, – вот лишь малая толика тех новаций, которые характеризуют сегодня как науку в целом, так в особенности и сферу научно-технической информации. Все эти изменения способствуют тому, что и на уровне руководства страны возникает потребность в постановке задач, относящихся к развитию инфраструктуры научно-технической информации.

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 и действует. Подробнее ситуация с этим положением и позиция автора по этому вопросу изложены в главе 2 данной монографии.

В данной главе мы рассмотрим относящиеся к сфере научной информации решения институтов государственной власти за последние годы (2018–2019).

4.2. Поручения Президента РФ

Важным этапом формирования государственной политики в данной сфере стал перечень из 12 поручений по итогам заседания Совета при Президенте РФ по науке и образованию, состоявшегося 27.11.2018 г. [81].

Среди них имеется следующее.

«Правительство Российской Федерации совместно с Федеральным государственным бюджетным учреждением “Российская академия наук” должны представить предложения *“О создании современной цифровой инфраструктуры для хранения и анализа научно-технической информации, а также для обмена такой информацией”*».

Очень важным для развития системы НТИ представляется также и другое поручение, содержащееся в этом же перечне:

«Правительству Российской Федерации принять дополнительные меры, направленные на обеспечение открытости научных организаций и доступности научных данных, а также на популяризацию отечественной науки».

Данные поручения, на наш взгляд, свидетельствуют о том, что проблемы развития научной информации продолжают оставаться приоритетными для государства. А внимание к проблемам открытости подчеркивает важность соответствия основным тенденциям развития научных коммуникаций, о которых мы много говорили в главе 1.

Вместе с тем на сегодняшний день остаются не до конца проясненными как общий вектор развития сферы НТИ, так и ее инфраструктурные составляющие, позволяющие интегрировать науку как социальную подсистему в формирующееся цифровое общество. Выполнение президентских поручений должно, на наш взгляд, способствовать выстраиванию системы приоритетов модернизации сферы НТИ и формированию эффективной организационно-технической модели научно-информационной деятельности.

4.2. Законопроект «О научной и научно-технической деятельности в Российской Федерации»

Центральное место в законодательном регулировании научно-информационной деятельности должен занять Федеральный закон «О научной и научно-технической деятельности в Россий-

ской Федерации». Общественное обсуждение этого законопроекта Министерство науки и высшего образования инициировало в июне 2019 г. [77]. Ввиду важности законопроекта для рассматриваемой проблемы, мы приводим полностью главу, посвященную научной информации.

«Глава 6. ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ НАУЧНОЙ И НАУЧНО-ТЕХНИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ

Статья 41. Государственные информационные системы в сфере научной, научно-технической деятельности / Информационное обеспечение научной и научно-технической деятельности.

1. В целях сбора, хранения, обработки информации, необходимой для осуществления научной и научно-технической деятельности, предоставления услуг субъектам такой деятельности, создаются федеральные государственные информационные системы.

2. Федеральный орган исполнительной власти, осуществляющий функции по выработке государственной политики и нормативно-правовому регулированию в сфере науки и научно-технической деятельности, определяет перечень информации, содержащейся в государственных информационных системах, создает и обеспечивает функционирование этих систем в порядке, установленном Правительством Российской Федерации.

3. Сведения, содержащиеся в государственных информационных системах, создание которых предусмотрено указанной статьей Федерального закона, являются общедоступными, за исключением информации, доступ к которой ограничен в соответствии с законодательством Российской Федерации.

Статья 42. Федеральная информационная система государственной научной аттестации

1. Федеральная информационная система государственной научной аттестации (далее – единая информационная система) создается в целях информационного обеспечения проведения государственной научной аттестации.

2. Создание единой информационной системы и обеспечение ее функционирования осуществляются федеральным органом исполнительной власти, осуществляющим функции по выработке государственной политики и нормативно-правовому регулированию в сфере научной и научно-технической деятельности.

3. В единой информационной системе содержится информация:

1) о персональном составе Высшей аттестационной комиссии и ее деятельности;

2) о советах по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук;

3) о персональном составе советов по защите диссертаций на соискание ученых степеней и об их деятельности, в том числе о защищенных и представленных к защите диссертациях на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, об авторефератах защищенных и представленных к защите диссертаций на соискание ученых степеней;

4) о присуждении, лишении, восстановлении ученых степеней, присвоении, лишении, восстановлении ученых званий, признании иностранных ученых степеней, иностранных ученых званий;

5) об иной определенной Правительством Российской Федерации информации в части государственной научной аттестации.

4. Порядок создания и ведения единой информационной системы, категории ее пользователей, перечень информации, предоставляемой каждой категории пользователей, и условия ее предоставления устанавливаются Правительством Российской Федерации.

5. Указанная в пункте 3 настоящей статьи информация предоставляется пользователям единой информационной системы бесплатно.

6. Научные организации, образовательные организации высшего образования, организации дополнительного профессионального образования представляют информацию о государственной научной аттестации для включения в единую информационную систему, состав и порядок представления которой определяются федеральным органом исполнительной власти, осуществляющим функции по выработке государственной политики и нормативно-правовому регулированию в сфере научной и научно-технической деятельности».

То, что текст этой главы совершенно неудовлетворителен, доказывать, видимо, не требуется. Законопроект не решает ни одной из важных проблем организации научно-информационной деятельности, в том числе с учетом тенденций развития научных коммуникаций, рассмотренных нами в первой главе. При этом первоочередное значение имеет государственная политика по проблеме открытости научной информации. Принципы этой политики необходимо закрепить законодательно. В рамках общественного обсуждения законопроекта автор представил в Минобрнауки свои предложения по доработке законопроекта. Ниже излагаются эти предложения.

В ст. 2 «Основные понятия, используемые в настоящем Федеральном законе» следует добавить:

9) информационные ресурсы научной информации – электронные библиотеки, архивы, справочники и энциклопедии, базы и банки данных, массивы пространственных данных, иные систематизированные массивы научной информации, создаваемые и используемые в научной деятельности, прежде всего в цифровой форме.

Обоснование

Этот термин уже используется в ст. 2, п. 3.

В ст. 10 «Научная и научно-техническая деятельность. Структура системы научной и научно-технической деятельности» в п. 1. следует добавить:

4) информационное обеспечение научной деятельности.

Обоснование

Этот вид научной деятельности предусмотрен главе 6 настоящего закона.

В ст. 11 «Инфраструктура научной и научно-технической деятельности», п. 1 следует добавить:

6) информационные ресурсы научной информации.

Обоснование

Это соответствует ст. 2. п. 3, где уже указано, что информационные ресурсы входят в инфраструктуру научной и научно-технической деятельности. Возможна также формулировка:

«2. Научные коллекции, банки данных и другие информационные ресурсы научной информации».

Главу 6 «Информационное обеспечение научной и научно-технической деятельности» предлагается изложить следующим образом.

41.1. Государственное управление (координация) деятельностью по информационному обеспечению науки.

Государственное управление информационным обеспечением научной и научно-технической деятельности осуществляется федеральным органом исполнительной власти, осуществляющим функции по выработке государственной политики и нормативно-правовому регулированию в сфере науки и научно-технической деятельности, при помощи (на основе) постоянно действующей государственной программы по информационному обеспечению научной и научно-технической деятельности. Целями указанной программы являются:

1) комплексное информационное сопровождение научных программ и проектов;

2) целенаправленное формирование Единого российского электронного пространства знаний (в соответствии с ст. 16.1. п. 1 ФЗ «О библиотечном деле»);

3) организация коллаборации участников научно-информационной деятельности на российском и международном уровнях;

4) координация деятельности государственных информационных систем;

5) обеспечение сохранности информационных ресурсов научной информации;

6) оптимизация научных изданий и коммуникаций и популяризация науки.

Участниками государственной программы по информационному обеспечению научной и научно-технической деятельности могут быть организации всех форм собственности, как специализированные (библиотеки, институты информации, издательства, архивы), так и другие научные организации.

Для финансовой поддержки деятельности по информационному обеспечению научной деятельности создается специальный фонд (эндаумент) в соответствии со ст. 19 настоящего закона.

41.2. Государственные информационные системы в сфере научной и научно-технической деятельности (*ст. 41 законопроекта предлагается оставить без изменений*).

41.3. Информационные ресурсы научной информации.

Информационные ресурсы научной информации (электронные библиотеки, электронные справочники и энциклопедии, базы и банки данных, массивы пространственных данных, иные систематизированные массивы научной информации) могут создаваться всеми участниками научной деятельности. Создание информационных ресурсов научной информации является самостоятельной формой научной деятельности и учитывается при оценке деятельности научных организаций. Создание и функционирование информационных ресурсов научной информации поддерживается государственной научной программой по информационному обеспечению научной деятельности. Если создание, поддержка и хранение информационных ресурсов научной информации осуществляются различными юридическими и физическими лицами, правовые отношения между ними устанавливаются в соответствии с ст. 24 настоящего закона.

41.4. Мониторинг и экспертиза информационных ресурсов научной информации.

Федеральный орган исполнительной власти, осуществляющий функции по выработке государственной политики и нормативно-правовому регулированию в сфере науки и научно-технической деятельности, определяет порядок проведения мониторинга и экспертизы информационных ресурсов научной информации. Целью экспертизы является определение качества информационных ресурсов научной информации, целесообразность их государственной поддержки и включения в федеральную научную программу по информационному обеспечению научной деятельности. Результаты мониторинга размещаются в государственной информационной системе «Каталог (регистр) российских информационных ресурсов научной информации».

42.5. Открытость и доступность научной информации.

При заключении контрактов на выполнение НИОКР за счет средств государственного бюджета или с привлечением средств бюджета, включая средства государственных научных фондов (РНФ, РФФИ), федеральный орган исполнительной власти, осуществляющий функции по выработке государственной политики и нормативно-правовому регулированию в сфере науки и научно-технической деятельности, должен предусмотреть в контракте:

1) положение, обязывающее исполнителей контракта размещать результаты НИОКР, имеющих форму произведений науки, в Интернете под открытой лицензией в соответствии со ст. 1286¹ Гражданского кодекса РФ «Открытая лицензия на использование произведения науки, литературы или искусства» или под свободной международной лицензией Creative Commons;

2) положение, обязывающее исполнителей контракта размещать результаты НИОКР, выполненные за счет государственного бюджета и охраняемые как объекты смежных прав (базы данных), а также информацию о результатах НИОКР, не охраняемую авторским правом, в Интернете в форме открытых данных в соответствии со ст. 7 ФЗ Федерального закона от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации», с изменениями и дополнениями от: 27 июля 2010 г., 6 апреля, 21 июля 2011 г., 28 июля 2012 г., 5 апреля, 7 июня, 2 июля, 28 декабря 2013 г., 5 мая, 21 июля 2014 г.

Нужно заметить, что процитированный законопроект подвергся серьезной критике со стороны РАН и других участников общественных консультаций. Поэтому дальнейшая судьба законо-

проекта, так же как и предложений автора, на момент написания настоящих строк (декабрь 2019 г.) неизвестна.

4.3. Цифровые платформы Минобрнауки

Очевидно, что центральная роль в выработке и реализации государственной политики в области научной информации должна принадлежать Министерству науки и высшего образования РФ, которое, согласно Положению о Министерстве, «осуществляет функции по выработке государственной политики и нормативно-правовому регулированию в сфере науки и научно-технической деятельности». Однако реорганизации последних лет сильно задержали разработку необходимых документов.

В настоящее время разработаны и одобрены Советом по цифровому развитию науки три документа, которые, по идее, должны определять государственную политику, в том числе в области цифровых научных коммуникаций и научной информации. Это:

- Концепция создания Единой цифровой платформы науки и высшего образования (ЕЦПН) [53];

- Концепция цифровой автоматизированной системы предоставления сервисов научной инфраструктуры коллективного пользования (АС УСНИКП) [54];

- Концепция Единой цифровой платформы научного и научно-технического взаимодействия, организации и проведения совместных исследований в удаленном доступе, в том числе с участием зарубежных ученых (ЦПСИ) [44].

Во всех этих документах есть положения, прямо затрагивающие сферу НТИ и научных коммуникаций, однако формулировки этих положений дают основания полагать, что авторы указанных документов недостаточно хорошо знакомы с реалиями современной ситуации в НТИ. Так, в концепции ЕЦПН сформулирована задача (п. 4.1.):

«Обеспечить на основе общих стандартов участникам совместных исследований доступ к внешним распределенным системам хранения и обработки научно-технической и наукометрической информации, к следующим федеральным информационным фондам и организациям, образующим государственную систему научной и технической информации:

• к фонду депонированных рукописей и зарубежной реферативной информации Всероссийского института научной и технической информации Российской академии наук;

• к фонду опубликованной научно-технической литературы;

• к электронным фондам Российской государственной библиотеки;

• к цифровым коллекциям отчетов, выполненных по грантам Российского фонда фундаментальных исследований, Российского научного фонда, Фонда поддержки малого предпринимательства;

• к фонду патентов Федерального института промышленной собственности;

• к системам и фондам комплектования обязательного экземпляра документов;

• к единой государственной информационной системе учета научно-исследовательских, опытно-конструкторских и технологических работ гражданского назначения;

• к федеральной информационной системе государственной научной аттестации (в части фонда диссертаций);

• к федеральной системе мониторинга результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы».

Однако выше (глава 2) было показано, что современную систему научной и технической информации образует совсем другой перечень фондов и организаций. Цитированный перечень концепции ЕЦПН не соответствует ни действующему до сих пор Постановлению № 950, регламентирующему ГСНТИ, ни реальному положению вещей. В документе упомянуты несуществующие информационные системы, например «фонд опубликованной научно-технической литературы», «системы и фонды комплектования обязательного экземпляра документов», и в то же время не упомянуты многие федеральные системы НТИ, действующие на официальной нормативной базе (ИНИОН РАН, ГПНТБ России, ЦНМБ, ЦНСХБ, Росгеолфонд, Национальная электронная библиотека и мн. др.). Вообще не упомянуты наиболее востребованные на сегодня информационные системы и ресурсы: eLIBRARY, КиберЛенинка, другие популярные электронные библиотеки.

В Концепции АС УСНИКП также присутствует ряд функций, прямо предусматривающих использование ресурсов и фондов НТИ. Например: «АС УСНИКП будет:

• выполнять роль агрегатора информации объектов научной инфраструктуры коллективного пользования;

- содержать актуальные реестры ресурсов, необходимых для оказания услуг;

- автоматизировать подготовку отчетных материалов по результатам использования научной инфраструктуры коллективного пользования в унифицированном формате, пригодном для включения в типовой отчет о НИР и экспертизы РАН;

- обеспечивать возможность оценки вклада научной инфраструктуры и ее коллективного использования в рост публикационной и патентной активности».

Если признать, что цифровые коллекции и банки данных коллективного использования (в том числе электронные библиотеки, музеи и архивы) являются компонентом научной инфраструктуры (а это автору представляется очевидным), то нужно, по крайней мере, провести инвентаризацию этих компонентов, чего ни в России в целом, ни в РАН в частности до сих пор не сделано.

Далее в концепции АС УСНИКП сказано:

«В рамках разработки АС УСНИКП будут созданы интерфейсы для подключения цифровых коллекций и банков данных организаций, выполняющих научные исследования и разработки».

Очевидно, что этот тезис также не может быть реализован без проведения детальной инвентаризации существующих цифровых коллекций и банков данных, а их число весьма велико и составляет не менее нескольких десятков тысяч.

Еще одна цитата из концепции АС УСНИКП:

«В части мониторинга и анализа использования научной инфраструктуры должны обеспечиваться:

- сбор информации и возможность оценки состояния и ресурсного потенциала объектов научной инфраструктуры;

- сбор информации и возможность оценки результативности использования объектов научной инфраструктуры, в том числе для реализации приоритетов научно-технологического развития Российской Федерации (в комплексных научно-технических программах и проектах);

- сбор информации и возможность оценки достигнутого уровня выполнения программ развития объектов научной инфраструктуры;

- сбор информации и возможность оценки достигнутого уровня выполнения проектов создания новых объектов научной инфраструктуры».

В свете сказанного выше этот пункт также не может быть выполнен.

Нужно отметить, что сроки выполнения этапов, обозначенные в данной концепции, не выдерживаются. Так, до 30 апреля 2019 г. должна была быть «отобрана организация-оператор для обеспечения инфраструктуры разработки, последующей эксплуатации и сопровождения АС УСНИКП». Насколько известно автору, это до сих пор не сделано.

Концепция ЦПСИ также предусматривает разработку сервисов, важных для системы НТИ. Цитируем:

«ЦПСИ должна представлять собой совокупность инструментов и сервисов на основе цифровых технологий, которые:

- поддерживают эффективный обмен научно-технической и наукометрической информацией между участниками проектов;
- на основе общих стандартов обеспечивают доступ участников совместных исследований к внешним распределенным системам хранения и обработки научно-технической и наукометрической информации;
- обеспечивают поддержку эффективного обмена научно-технической и наукометрической информацией между участниками проектов».

И далее:

«ЦПСИ должна обеспечивать следующие функциональные возможности:

- интеграционное взаимодействие с внешними системами;
- идентификацию участников (в том числе зарубежных) научно-технического взаимодействия;
- работу участников в режиме удаленного доступа, в том числе с зарубежными партнерами;
- создание и работу виртуальных команд и виртуальной коллаборации;
- регламентированный доступ зарубежных ученых для организации и проведения совместных исследований, а также международной экспертизы;
- эффективный обмен научно-технической и наукометрической информацией между участниками проектов;
- мониторинг реализации проектов;
- автоматизированное формирование отчетов по унифицированным требованиям;
- доступ участников совместных исследований к внешним распределенным системам хранения и обработки научно-технической и наукометрической информации;

- автоматизацию совместной работы участников в научных и научно-технических проектах;

- автоматизированный интеллектуальный, в том числе с привлечением экспертов, поиск и подбор необходимых специалистов и формирование проектных команд, в том числе в виде виртуальных команд и виртуальной коллаборации при реализации комплексных научных, научно-технических проектов, на основании интеллектуального анализа профилей ученых, исследователей и научной среды в целом;

- автоматизированный интеллектуальный, в том числе с привлечением экспертов, подбор ресурсного обеспечения под проект;

- совместную работу с документами, информацией;

- ведение каталогов:

- научных исследований и сервисов исследований;

- планируемых исследований и возможностей принятия участия в них;

- кадров научно-исследовательского персонала и их компетенций;

- средства коммуникации участников;

- средства автоматизированных уведомлений;

- обеспечение проведения экспертизы;

- мониторинг научной, научно-технической деятельности пользователей ЦПСИ, деятельности по управлению правами на полученные результаты, позволяющий проводить оперативную оценку результативности и (или) эффективности их деятельности;

- подготовку документации с учетом имеющихся сведений».

Среди сервисов, разрабатываемых в рамках ЦПСИ, предусмотрены:

- «Сервисы по хранению, каталогизации и наукометрии:

- научная библиотека исследований;

- контекстный поиск по всем исследованиям и полученным охраноспособным результатам, включая результаты, охраняемые как изобретение, полезная модель или промышленный образец, включая доступ к другим базам;

- отображение объема использования (востребованности) результатов, в том числе индекса цитируемости, данных об использовании результатов в случае их лицензирования;

- интеграция с имеющимися справочниками».

Среди ожидаемых результатов концепция ЦПСИ указывает: «Качественное повышение уровня и интенсивности информационного обмена научно-технической и наукометрической информацией

о проектах, организациях, коллективах, исследователях и результатах интеллектуальной деятельности для повышения качества исследований и их результативности».

С учетом сказанного выше автор выражает глубокий скепсис в отношении реализуемости данных положений, по крайней мере в отношении существующей системы научной информации. Тем не менее автор считает, что специалисты должны внимательно следить за ходом реализации этих концепций и по возможности влиять на него, чтобы в максимальной степени учесть реальные потребности научно-информационной деятельности.

4.4. Национальная электронная библиотека

Проект Национальной электронной библиотеки (далее – НЭБ) реализуется уже достаточно много лет (первая концепция НЭБ была утверждена в 2004 г.), и ее описание содержится во многих работах, в том числе в монографии [22]. Однако в 2019 г. вышли ряд правительственных документов по развитию НЭБ, которые требуют обсуждения.

Прежде всего, это «Положение о НЭБ» [83], содержащее важные изменения в идеологии комплектования и формирования НЭБ, а также интерпретацию понятия *Единого российского электронного пространства знаний*. Это понятие будет специально рассмотрено в главе 5.

Кроме того, отдельным правительственным распоряжением был утвержден план развития НЭБ [98].

Важнейшей новацией в развитии НЭБ является переход от отбора приоритетных изданий к тотальному включению в НЭБ всех выходящих в России книг на основе обязательного электронного экземпляра. Для решения возникающих проблем копирайта предполагаются заключение лицензионных соглашений с правообладателем и отчисление правообладателю роялти либо за счет читателей, либо за счет участников проекта НЭБ. Права на определенную часть наиболее ценных книг предполагается выкупить за счет бюджета и передать эти книги в открытый доступ.

Высокий уровень утверждения Положения о НЭБ и плана развития НЭБ свидетельствует, на наш взгляд, о том, что НЭБ должна занять важное или даже центральное место в общем информационном пространстве науки и образования. Однако почему-

то этот факт не учитывается в рассмотренных выше концепциях развития платформ цифровизации науки.

Заметим также, что Межведомственный совет по НЭБ, созданный в 2015 г. [97], распущен Постановлением Правительства № 169, а новый состав этого совета не утвержден. Поэтому в 2019 г. совет не собирался и стратегия развития НЭБ не обсуждалась. Соответственно, не принимаются многие решения по НЭБ (например, создание экспертных советов), которые в соответствии с Положением о НЭБ может принять только Межведомственный совет.

4.5. Национальный энциклопедический портал

В целом представляется очевидным, что не существует общей государственной политики и плана по упорядочению и выстраиванию в систему научных и образовательных информационных ресурсов, в том числе по их агрегации и интеграции.

Приведем свежий пример. 26 ноября 2019 г. Правительство РФ издало распоряжение № 2804-р, которым утвержден устав Автономной некоммерческой организации Национальный научно-образовательный центр «Большая российская энциклопедия» [136].

Согласно ранее принятым правительственными документам по созданию Национального энциклопедического портала, именно эта организация должна играть центральную роль в создании и распространении в современной цифровой среде достоверных и систематизированных научных и образовательных знаний. В уставе организации говорится:

«10. Целью организации является оказание услуг по выполнению функций проектного офиса общенационального интерактивного энциклопедического портала, в том числе:

а) по обеспечению доступа граждан к знаниям, информации, культурным ценностям и благам;

б) по систематизации, актуализации и распространению современных научных знаний;

в) по обеспечению взаимодействия между компетентными источниками информации;

г) по содействию научного обеспечения системы образования Российской Федерации».

Однако в цитированных выше правительственных документах по созданию Национальной электронной библиотеки схожие функции возложены именно на Национальную электронную биб-

лиотеку (практически – на РГБ). Однако во всех этих документах нет ни слова о том, как эти организации, которые к тому же относятся к различным ведомствам, должны взаимодействовать между собой. Тем более – ни слова о том, как те и другие решения соотносятся с формально действующей Государственной системой научно-технической информации.

Таким образом, можно констатировать полное отсутствие государственной политики по формированию информационного пространства науки и образования.

Правительственному решению о создании Национального энциклопедического портала сопутствовала оживленная общественная дискуссия о соотношении этого проекта с популярнейшим энциклопедическим порталом Википедия, тем более что Президент РФ обронил неудачную реплику о том, что Национальный энциклопедический портал должен «заменить Википедию».

Конечно, на наш взгляд, ни о какой «замене» не может быть и речи – это совсем разные проекты с разной идеологией и технологией сбора, обработки и распространения информации, хотя и существенно пересекающиеся по контенту. Однако не праздным является другой вопрос: как в перспективе оба проекта должны сосуществовать между собой и с другими системами научной информации в некоем едином электронном пространстве знаний?

Как представляется автору, идеальным решением должен быть поиск форм сотрудничества – и в идеале коллаборации, – сочетающих достоинства этих и других важных проектов в области научной информации.

4.6. Другие правительственные действия в сфере научной информации

Основным объектом внимания органов власти в сфере научной информации в последние годы были российские научные периодические издания. При поддержке Минобрнауки начали реализовываться проекты по продвижению научных журналов в мировые наукометрические индексы¹. Для этого была организована библиометрическая и экспертная оценка российских научных журналов. В Министерстве была даже создана комиссия по оценке

¹ Подробности программ поддержки журналов см., напр., на странице: <https://xn--80aafbngaja7auoph0at.xn--p1ai/> (развитиежурналов.рф)

научных журналов. В результате было отобрано 70 журналов, для которых была принята специальная программа развития.

Этой программе соответствуют рекомендации круглого стола при Комитете Государственной думы по науке и образованию «Совершенствование системы научных публикаций» [101]. Ассоциацией научных редакторов и изданий (АНРИ) было проведено несколько конференций и других мероприятий с этими же целями.

Не секрет, что эти усилия были предприняты потому, что в качестве важнейшего критерия эффективности российских научных учреждений приняты показатели цитируемости публикаций российских исследователей в журналах с высоким импакт-фактором.

В то же время в главе 1 было показано, что ориентация на импакт-факторы журналов как основной критерий качества исследований является ошибочной и ведет к гонке за публикациями в престижных журналах вместо стремления к повышению качества научных исследований. Поэтому мировая научная общественность активно ищет более адекватные системы оценки научных исследований.

Еще одним направлением активности органов власти была координация деятельности по обеспечению доступа к зарубежной научной информации. В Минобрнауки создан специальный Межведомственный совет по организации предоставления доступа к информационным наукометрическим базам данных и полнотекстовым научным ресурсам [89]. Последнее его заседание было проведено 3 декабря 2019 г. В отчете об этом заседании говорится:

«В настоящее время Минобрнауки России реализует ряд проектов, направленных на повышение видимости публикаций российских ученых, повышение показателей и рейтинга российских журналов и их вхождения в международные наукометрические базы данных. Так, в 2019 г. министерство обеспечило доступ к 29 полнотекстовым ресурсам для 614 организаций (в 2018 г. – для 204 организаций) и к международным наукометрическим базам данных Web of Science и Scopus более чем для 1250 организаций в рамках централизованной (национальной) подписки. Кроме того, национальный доступ к журнальным коллекциям издательства Springer Nature, а также книгам и журнальным коллекциям издательства Elsevier был обеспечен Российским фондом фундаментальных исследований. <...>

Также на заседании Межведомственного совета были затронуты вопросы о мерах поддержки российских научных журналов,

повышения их показателей и рейтинга, механизмах повышения видимости публикаций отечественных исследователей.

Планируется, что одной из мер поддержки российских научных журналов станет организация подписки на наиболее востребованные российские научные журналы через оператора подписки – РФФИ. Также в целях достижения показателей Национального проекта “Наука” и интеграции в мировую науку (повышение видимости публикаций российских ученых и повышение показателей и рейтинга российских журналов посредством опубликования научных работ российских ученых в англоязычных версиях российских журналов в открытом формате) предлагается создание электронных архивов и внедрение модели Open Access» [99].

Последний тезис о внедрении открытого доступа следует всячески поддержать.

Глава 5.

ЕДИНОЕ РОССИЙСКОЕ ЭЛЕКТРОННОЕ ПРОСТРАНСТВО ЗНАНИЙ – ПОСТАНОВКА ЗАДАЧИ

5.1. Единое российское электронное пространство знаний в «Положении о Национальной электронной библиотеке»

Возможным ориентиром для обсуждения образа будущего научных коммуникаций в России могло бы стать упоминавшееся ранее понятие *Единого российского электронного пространства знаний* (ЕРЭПЗ). 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» [108].

Также утверждено предусмотренное законом Положение о Национальной электронной библиотеке [83], 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 функциони-

рующих 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 может охватывать все области научной коммуникации.

Однако далее в этом документе имеется специальный раздел, посвященный ЕРЭПЗ, который в силу его важности приводится 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» [83].

5.2. Некоторые замечания о понятии ЕРЭПЗ

По поводу приведенных формулировок возникает множество вопросов. Также возможна их разнообразная интерпретация. Это касается *«энциклопедической систематизации»*, *«семантической взаимосвязи документов»*, *«системы классификации, поиска и извлечения информации»*. По проблемам представления знаний, соотношения смысла и текста, классификации, создания онтологий и по смежным проблемам написаны горы литературы, высказано множество идей, причем совершенно неясно, на какие научные школы или подходы опирались авторы концепции ЕРЭПЗ.

Подробный анализ этой проблематики Национальной электронной библиотеки, очевидно, выходит за рамки задачи, поставленной в настоящей работе.

Ограничимся обсуждением одной проблемы – соотношения понятия «*пространство знаний*» с понятиями «*научная информация*», «*публикации и данные*», «*информационные ресурсы*», «*энциклопедический портал*», которые мы использовали в этой работе ранее. На наш взгляд, очевидно следующее: авторы «Положения о НЭБ» однозначно отождествляют пространство знаний с собранием библиотечных, архивных, музейных и аудиовизуальных документов. Иначе говоря, в документе доминирует «библиотечная», а точнее документалистская, парадигма мышления.

В то же время, как мы показали в первой главе, а также в третьей, на примере информационных ресурсов социогуманитарных наук, тенденции развития научных коммуникаций совершенно иные, они выходят за рамки документалистской парадигмы.

С подходом концепции ЕРЭПЗ согласиться, конечно, нельзя, хотя понять ее 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 (библиотечные, архивные и музейные) системы. Может быть, именно поэтому авторы не касаются соотношения ЕРЭПЗ и Национального энциклопедического портала, хотя их задачи и содержание явным образом пересекаются. Вообще проблема «*данных*», управление которыми становится мейнстримом научных коммуникаций, выведена за пределы рассмотрения в концепции ЕРЭПЗ.

Авторы «Положения о НЭБ» также не связывают организацию пространства знаний с проблемой выявления нового знания и оценки научных результатов, что, как показано выше, является главной задачей современной научной коммуникации.

Автор и его коллеги в опубликованной ранее статье сформулировали ряд проблем создания ЕРЭПЗ. Приведем цитату из этой работы.

«1. Авторы рассматривают не пространство знаний в целом, а его научную часть (подпространство научных знаний), тем самым выделяя те информационные объекты, которые верифицированы мировым научным сообществом, и, соответственно, отделяя их от информационных объектов, которые носят идеологический, религиозный и другой спорный с научной точки зрения характер.

2. Из наименования предмета обсуждения исключается признак “русский”, поскольку научное знание является трансграничным по своему существу и выделение русского фрагмента информационного научного пространства в контексте последующих рассуждений не имеет значения.

3. Авторы считают более адекватной и соответствующей современным представлениям формулировку *цифровое*, а не *электронное* пространство.

Таким образом, предметом рассмотрения данной статьи будет являться Единое цифровое пространство научных знаний (ЕЦПНЗ), что и вынесено в ее заголовок.

Об определении основных понятий

Говоря о создании цифрового пространства научных знаний, необходимо, прежде всего, определить каждое понятие, входящее в это словосочетание. Пространство в его математическом (формализованном) определении есть совокупность элементов с определенными 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 научные знания – совокупность практических результатов применения теоретических знаний, подтвержденных 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 ученого и заканчивая 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 “фасетов”, вообще говоря, связанных между собой, относящихся к отдельным научным направлениям; каждый фасет должен иметь фундаментальную (статическую) основу, включающую основные постулаты и базовые результаты, положенные в основу данного научного направления, и динамическую часть, включающую информацию о новейших достижениях в данной научной области» [75].

Признавая аббревиатуру ЕЦПНЗ более адекватной для именования пространства научного знания, автор тем не менее в этой книге использует аббревиатуру ЕРЭПЗ, поскольку она уже стала официальной.

Заметим также, что проблематика ЕРЭПЗ уже обсуждалась в ряде публикаций [33]. Публикации автора по данной проблеме, на которых строится дальнейшее изложение, приведены в общем списке литературы.

5.3. Общие принципы создания ЕРЭПЗ

В настоящей главе автор представляет видение основных принципов создания ЕРЭПЗ, учитывающее современное состояние и тенденции развития российской научной инфосферы, под которой понимается совокупность институций, ресурсов и сервисов, используемых в научной коммуникации, и которая была охарактеризована в предыдущих главах.

Интеграция (агрегация, конвергенция) существующих информационных ресурсов и сервисов в ЕРЭПЗ потребует мотивировать основных участников информационной деятельности на участие в соответствующих процессах. Этого можно добиться экономическими, административными и правовыми методами, однако наиболее перспективный путь – обеспечить заинтересованность научных и образовательных учреждений, отдельных ученых в коллаборации как основной методологии формирования, поддержки и развития ЕРЭПЗ. Эффективность такого подхода демонстрирует, в частности, Википедия.

Что касается соотношения ЕРЭПЗ и Национальной электронной библиотеки, то автор, конечно, сознает важность библиотечных ресурсов для формирования ЕРЭПЗ. Но библиотечные ресурсы не могут быть единственным источником этого пространства. Все

более важное место в научных коммуникациях занимают формы представления научных результатов, не прошедшие через издательские и библиотечные процессы, и этот факт должен учитываться в стратегии ЕРЭПЗ. Но то, что Национальная библиотека должна быть одним из важнейших источников или компонентов ЕРЭПЗ, не вызывает сомнений.

Целью работ по формированию ЕРЭПЗ 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.

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

1. ЕРЭПЗ должно формироваться из существующих в России ресурсов научной информации на основе конвергенции, не нарушая их функционирования.

2. ЕРЭПЗ должно обеспечивать преемственность в отношении Государственной системы научной и технической информации (ГСНТИ), функционирующей в современных организационных, правовых, экономических и технологических условиях.

3. ЕРЭПЗ должно строиться на принципах открытой науки.

4. ЕРЭПЗ должно включать централизованное ядро (Онтологию научного знания и Базу знаний) и множество распределенных ресурсов публикаций и данных.

5. База знаний должна включать контент Национального энциклопедического портала.

6. Ресурсы, используемые для формирования ядра ЕРЭПЗ, должны быть отобраны по критериям качества.

7. Онтология научного знания должна наследовать существующие информационные языки, номенклатуры и системы метаданных.

8. Информационные объекты, включаемые в Базу знаний ЕРЭПЗ, должны отвечать критериям достоверности, проверяемости, авторитетности и научности, что не исключает включения в Базу знаний положений, отражающих различные точки зрения.

9. Технологии формирования ядра ЕРЭПЗ должны обеспечивать возможность коллаборации участников.

10. ЕРЭПЗ должно использоваться для оценки результатов научной деятельности, но с учетом существенной модернизации используемых систем оценки.

5.4. Цели и задачи ЕРЭПЗ

Основные задачи ЕРЭПЗ можно сформулировать следующим образом:

- 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;
- согласованность с программами цифровизации науки и платформами цифрового развития науки;
- координация и оптимизация научно-информационной деятельности в России, в том числе 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 и стран постсоветского пространства, выявление новых и перспективных научных направлений.

Перечисленные основные задачи ЕРЭПЗ можно назвать внешними по отношению к системе научной информации.

Кроме того, при создании ЕРЭПЗ необходимо обеспечить решение задач, которые можно назвать внутренними, направленными на оптимизацию собственно научной инфосферы:

- мониторинг научно-образовательных электронных ресурсов, создаваемых в России и странах постсоветского пространства;
- экспертиза научных ресурсов и отбор наиболее качественных в качестве составных частей или источников ядра ЕРЭПЗ;
- согласованное развитие качественных и востребованных научных ресурсов;

- организация доступа к научным ресурсам, навигация и поиск в них, осуществляемые с помощью единого окна;
- создание новых информационных продуктов и сервисов, обеспечивающих предоставление актуальной, достоверной информации в современной и удобной форме;
- разработка логико-лингвистических и программных средств, обеспечивающих качественный поиск, извлечение знаний, проверку научной информации на новизну и других задач, диктуемых развитием науки и стандартами научной этики;
- согласованная программа оцифровки традиционных (библиотечных, архивных музейных) ресурсов;
- архивирование и сохранность научных ресурсов;
- обмен с международными системами представления знаний;
- стандартизация информационных процессов и продуктов;
- разработка и реализация экономической модели функционирования научно-информационных сервисов;
- совершенствование правового обеспечения сферы научной информации.

5.5. Этапы формирования ЕРЭПЗ

Современное состояние научной инфосферы диктует стратегию ее развития на основе конвергенции существующего хаотического множества информационных систем и ресурсов в упорядоченную совокупность документов и данных. Можно представить несколько этапов реализации этой стратегии. В данной книге этот процесс представлен как самостоятельная программа. Но на практике создание ЕРЭПЗ должно входить как отдельный проект в Национальный проект «Наука» и должно быть согласовано с формированием цифровых платформ науки, описанных в главе 4.

Первый этап – мониторинг инфосферы

Формирование ЕРЭПЗ включает проведение инвентаризации научно-образовательной инфосферы, организацию постоянного мониторинга и учета научно-образовательных ресурсов и сервисов, а также институций – владельцев, генераторов и операторов этих ресурсов и сервисов. Результатом первого этапа должно стать создание общегосударственного каталога научно-образовательных

электронных ресурсов. Этот каталог может быть единым, но, вероятнее, должно быть создано несколько каталогов по основным научным направлениям:

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

Прототипом одного из таких каталогов может послужить описанный выше Навигатор информационных ресурсов по общественным наукам (НИРОН) [60].

Конечно, мониторинг научных и образовательных электронных ресурсов и формирование каталогов могут осуществляться и на более детальном уровне – по отдельным научным дисциплинам, субдисциплинарным направлениям или по отдельным категориям научных ресурсов. Например, в настоящее время разрабатывается проект создания специализированного Центра лингвистических ресурсов на базе Института русского языка им. В.В. Виноградова РАН. Очевидно, что первичными задачами этого центра должны быть мониторинг лингвистических ресурсов России, их учет и каталогизация.

Возможно, будет целесообразным организовать учет и мониторинг по отдельным типам научных ресурсов или по региональным отделениям научных учреждений. Например, учет и мониторинг университетских репозиторий осуществляет НЭИКОН в рамках проекта «Национальный агрегатор открытых репозиторий» (НОРА) [70]. Самостоятельный проект по мониторингу научных ресурсов ведет Сибирское отделение РАН.

Важно, чтобы мониторинг научно-информационных продуктов происходил согласованно.

Мониторинг инфосферы и формирование каталога научных ресурсов должны производиться регулярно, например ежегодно, поскольку создание ресурсов, их исчезновение или развитие происходят весьма динамично.

Второй этап – оценка и отбор научных ресурсов

Очевидно, что ресурсов, которые их создатели относят к категории научных и / или образовательных, слишком много, их качество и надежность далеко не всегда соответствуют современным

стандартам и пользовательским требованиям. Поэтому необходим второй этап – оценка и отбор наиболее качественных научных ресурсов.

Критериями отбора научно-образовательных электронных ресурсов (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 (например, качество поиска, надежность программного обеспечения).

Важными критериями для отбора ресурсов должны стать целесообразность их включения в процессы коллаборации, интеграции и агрегации, а также их использование для формирования ядра системы НТИ, иначе называемого ядром ЕРЭПЗ – Единого российского электронного пространства знаний. Назовем отобранные ресурсы основными ресурсами ЕРЭПЗ.

В состав научных и образовательных электронных ресурсов цифровой инфраструктуры НТИ могут быть включены ресурсы различных типов, содержащие как первичную, так и обработанную научную и образовательную информацию, в том числе:

- электронные библиотеки;
- архивы периодических изданий;
- репозитории, другие коллекции препринтов и неопубликованных документов;
- оцифрованные архивные и музейные коллекции, фонды, выставки и т.п.;
- лингвистические корпуса текстов;
- электронные представления рукописей и книжных памятников;
- фактографические базы данных, реестры, регистры, каталоги, перечни;
- энциклопедии, словари, справочники, указатели;
- каталоги, библиографии, архивные описи;
- учебная и методическая литература, онлайн-курсы;
- первичные данные научных исследований (журналы наблюдений и др.);
- пространственные данные географических информационных систем;
- экспертные системы, базы знаний;

- банки изображений, видео- и аудиодокументов, фильмов;
- интерактивные продукты, 3D-модели, продукты виртуальной и дополненной реальности.

Данный список составлен на основе мониторинга информационных ресурсов академической социальной и гуманитарной науки. Но, разумеется, возможно появление и других научных информационных продуктов в электронной форме, что обусловлено динамичным развитием так называемой E-science. Мы исходим из расширительной интерпретации E-science как применения компьютерных технологий для проведения современных научных исследований, включая подготовку, постановку экспериментов, сбор данных, распространение результатов, долгосрочное хранение и обеспечение доступности всех материалов, полученных в ходе научного процесса. Они могут содержать моделирование и анализ данных, электронные / оцифрованные лабораторные записи, необработанные и сгруппированные данные, черновые версии, препринты, а также печатные и / или электронные публикации.

В настоящее время количество российских электронных ресурсов, претендующих на осуществление научной, образовательной и просветительской функций, по экспертной оценке составляет 50–100 тыс., поэтому процедура их отбора для включения в ЕРЭПЗ представляется необходимой.

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

В первую очередь в число основных ресурсов, вероятно, должны быть включены электронные ресурсы, имеющие официальный статус государственных информационных систем, такие как Национальная электронная библиотека, АИС Федерального института промышленной собственности (ФИПС) или Государственный каталог Музейного фонда.

Дискуссионным является включение в число основных российских: ресурсов на иностранных языках (например, переводных журналов); ресурсов, создаваемых совместно с иностранными участниками; русскоязычных ресурсов, владельцами которых являются нерезиденты; и других спорных категорий ресурсов. В то же время в число основных могут быть включены ресурсы, создаваемые, например, в рамках Союзного государства России и Белоруссии

или в процессе функционирования наднациональных органов Евразийского экономического союза.

В любом случае методика отбора научных ресурсов и организация их экспертизы должны стать результатом широкого обсуждения научно-информационным сообществом и быть зафиксированы в нормативных документах. Тем более это необходимо для образовательных ресурсов.

Основные ресурсы, прошедшие экспертизу и отбор, должны получить официальный статус и соответствующую поддержку, включая государственное архивирование, и, как было отмечено, стать источниками ядра ЕРЭПЗ. Вероятно, в рамках экономической модели ЕРЭПЗ основные ресурсы должны получить определенные преференции.

Третий этап – формирование ядра ЕРЭПЗ

Формирование ядра ЕРЭПЗ – третий и наиболее сложный этап. Это ядро включает две части – Онтологию научных знаний и Базу знаний.

Онтология научных знаний должна унаследовать терминологию, номенклатуру, парадигматику и концептуальные схемы информационных языков и систем метаданных всех основных ресурсов научной информации, с тем чтобы обеспечить поиск и навигацию в этих ресурсах из единого окна. Важное условие – формирование онтологии не должно приводить к переработке основных ресурсов. Поиск в основных ресурсах следует поддерживать с помощью поисковых средств, которые были заложены в этих ресурсах.

Естественно, кроме специфических видов поиска, необходим и обычный сквозной лексический поиск для тех ресурсов, которые формируют поисковый лексический индекс.

Онтология должна обеспечивать поиск и навигацию в Базе знаний ЕРЭПЗ. Под Базой знаний здесь понимается формализованное представление сведений энциклопедического и справочного характера в форме, пригодной для компьютерной обработки. База знаний ЕРЭПЗ может представлять собой совокупность энциклопедических и справочных сведений, как в виде, пригодном для компьютерной обработки, так и в человекочитаемом виде. В качестве прототипа Базы знаний ЕРЭПЗ можно указать на проект DBpedia [121].

Автору представляется очевидным, что формирование Базы знаний ЕРЭПЗ должно быть согласовано с созданием Национального энциклопедического портала. Однако эта простая мысль до сих пор не зафиксирована ни в одном из нормативных документов.

Процедуру формирования исходной версии Базы знаний можно представить следующим образом:

- формирование перечня наиболее авторитетных и популярных энциклопедий, терминологических словарей и справочников по различным наукам;

- загрузка этих энциклопедий, терминологических словарей и справочников в единую среду, их интеграция;

- выявление противоречий, ошибок, неоправданного дублирования информации;

- содержательное редактирование выявленных ошибок;

- индексирование информационных объектов Базы знаний с помощью онтологии;

- разработка технологии пополнения Базы знаний.

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

Формирование Базы знаний следует осуществлять на принципах открытости, актуальности, достоверности, авторизованности и тематического универсума НТИ.

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

В Базу знаний необходимо заложить возможность отражения различных (в том числе противоречивых) научных теорий и концепций, а также механизм предотвращения включения в Базу знаний лженаучной и паранаучной информации. Очевидно, что решающая роль в этом должна принадлежать РАН как основной экспертной организации страны. Информация, содержащая государственную или коммерческую тайну, в Базу знаний не включается.

На основе Базы знаний и онтологии могут разрабатываться и реализовываться дополнительные (монетизированные) сервисы, развитие которых будет способствовать повышению экономической эффективности системы научной информации.

Четвертый этап – выявление новых знаний

Разработка системы выявления новых знаний на основе потока научной информации, генерируемого российскими научными организациями, – четвертый этап формирования ЕРЭПЗ.

Эта система может быть основана на существующих информационно-аналитических и реферативных службах. Специалистам-аналитикам необходимо иметь в своем распоряжении сервис, позволяющий определять новизну в документе или ином научно-информационном продукте по сравнению с Базой знаний. Если такая новизна обнаружена, то 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 может опираться на алгоритмы существующих систем выявления плагиата, но использовать более высокий уровень семантических технологий.

Процедуры выявления нового знания и включения их в ядро ЕРЭПЗ должны непосредственно отражаться в системе оценки эффективности научной деятельности. Эта система постепенно заменит существующие способы оценки на основе публикационной активности, которые демонстрируют устойчивое снижение своей адекватности и полезности.

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

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

5.6. Функциональный состав участников ЕРЭПЗ

Обсуждая организационную структуру ЕРЭПЗ, мы исходим из высказанного выше тезиса, что ЕРЭПЗ должно заменить в новых современных условиях Государственную систему научно-технической информации, прежде всего в части управления созда-

нием научных и образовательных информационных ресурсов, их агрегации, интеграции и хранения.

В состав ЕРЭПЗ должны входить следующие категории участников.

Головная организация

Ее основные функции:

- организация проектирования ЕРЭПЗ;
- программно-технологическая поддержка ЕРЭПЗ;
- разработка и поддержка стандартов ЕРЭПЗ;
- организация работ по созданию Онтологии и Базы знаний (ядра ЕРЭПЗ) и их дальнейшая поддержка;
- обеспечение архивации и сохранности основных ресурсов ЕРЭПЗ.

Функции головной организации ЕРЭПЗ, очевидно, следует возложить на ту организацию, которая обладает наилучшими возможностями для координации научно-информационной деятельности в стране на основе современных информационных технологий.

Кажется предпочтительным, чтобы роль головной организации выполняла государственная научная организация. По нашему мнению, в настоящее время наибольшим потенциалом для этой роли среди организаций, подведомственных Министерству науки и высшего образования РФ, обладает Федеральный исследовательский центр «Информатика и управление». Несомненно, однако, что этот вопрос требует отдельного обсуждения.

Ведущие организации по отраслям науки

Эти учреждения должны обеспечивать мониторинг научных и образовательных электронных ресурсов, их экспертизу, оценку и отбор основных ресурсов, координацию деятельности по их созданию, организацию информационно-аналитической деятельности, а также наукометрические исследования. Эти функции целесообразно возложить на информационно-библиотечные организации по тематическому профилю.

Наиболее хорошо подготовлены для этой цели организации, подведомственные Министерству науки и высшего образования РФ, в том числе:

• ВИНИТИ РАН (совместно с БЕН РАН) – естественные науки;

• ГПНТБ России – технические науки;

• ЦНСХБ РАН – сельскохозяйственные науки;

• ИНИОН РАН – общественные и гуманитарные науки.

Для медицинских наук эти функции могут быть возложены на ЦНМБ Минздрава РФ как бесспорного лидера научно-информационной деятельности в области медицины в России.

Как было отмечено, применительно к отдельным наукам и категориям научных ресурсов эти функции могут быть возложены и на другие научные учреждения. Например, для лингвистических информационных ресурсов в таком качестве может выступить проектируемый Центр ЛИР при ИРЯ РАН, а для исторических ресурсов – Президентская библиотека им. Б.Н. Ельцина.

К выполнению этих функций разумно подключить ведущие региональные информационно-библиотечные организации. Например, это БАН, ГПНТБ СО РАН, ЦНБ УРО РАН, ЦНБ ДВО РАН, а также вновь создаваемые научно-образовательные центры.

Операторы основных научно-образовательных электронных ресурсов

В этом качестве должны выступать организации – владельцы научных ресурсов, отобранных на втором этапе, например:

• Российская государственная библиотека (РГБ);

• Центр информационных технологий и систем (ЦИТиС);

• Федеральный институт промышленной собственности (ФИПС);

• МГУ им. М.В. Ломоносова;

• Президентская библиотека им. Б.Н. Ельцина;

• НЭБ eLIBRARY;

• Национальный электронно-информационный консорциум (НЭИКОН);

• КиберЛенинка.

Очевидно, что количество научных, образовательных и информационно-библиотечных учреждений, осуществляющих создание и поддержку основных научных и образовательных электронных ресурсов, должно быть велико. Окончательный перечень операторов основных ресурсов определится в ходе проектирования

ЕРЭПЗ. Существенно, что операторами основных ресурсов могут быть организации различных организационно-правовых форм.

Службы формирования ядра ЕРЭПЗ и выявления нового знания

В этом качестве должны выступать научные учреждения, научные коллективы (общества) и отдельные специалисты, осуществляющие в распределенном и удаленном режиме анализ российской научно-технической информации с целью определения новизны и пополнения новыми данными основных научных ресурсов, Онтологии и Базы знаний ЕРЭПЗ.

5.7. Технологическая платформа ЕРЭПЗ

Технологическая платформа ЕРЭПЗ должна поддерживать коллаборации по созданию, использованию и обеспечению сохранности научных ресурсов, созданию и использованию Онтологии и Базы знаний, совместимость с ведущими международными проектами в сфере НТИ. В качестве такой платформы может быть предложен Semantic Web.

Однако технологические решения по ЕРЭПЗ должны быть согласованы с Единой цифровой платформой науки и другими общесистемными решениями по цифровизации науки, в частности АС УСНИКП и ЦПСИ. Очевидно, что эти решения должны учитывать реальные технологические возможности существующих и вновь создаваемых операторов основных научно-образовательных электронных ресурсов.

5.8. Организационные принципы функционирования ЕРЭПЗ

Управление ЕРЭПЗ должно иметь межведомственный характер. Оперативное руководство, вероятно, должно быть возложено на Министерство науки и высшего образования РФ или Министерство цифрового развития РФ. Научное и методическое руководство может быть возложено на Российскую академию наук. Возможно, при Минобрнауке или при РАН должен быть создан с этой целью Межведомственный координационный (наблюдательный) совет ЕРЭПЗ.

ЕРЭПЗ создается и функционирует в рамках программы «Цифровая экономика» либо в тесной координации с проектами, реализуемыми в рамках этой программы.

С целью формирования и развития ЕРЭПЗ может быть создан консорциум, в состав которого должны войти ведущие научно-информационные центры различных организационно-правовых форм, в том числе государственные, коммерческие и некоммерческие, взаимодействующие на принципах государственно-частного партнерства. Возглавлять, обеспечивать организационную и финансовую поддержку ее создания и функционирования должен специальный орган управления (Межведомственный координационный или наблюдательный совет).

С органами управления ЕРЭПЗ должны сотрудничать координационные и совещательные органы, действующие в настоящее время в России в сфере научно-технической информации, в том числе:

- Межведомственный совет по развитию Национальной электронной библиотеки;
- Межведомственный координационный совет по организации предоставления лицензионного доступа к информационным наукометрическим базам данных и полнотекстовым научным ресурсам;
- Научно-издательский совет Российской академии наук;
- Межведомственная рабочая группа по вопросам, связанным с созданием на общенационального научно-образовательного интерактивного энциклопедического портала.

Моделирование некоторых аспектов создания ЕРЭПЗ предусмотрено в рамках Комплексного гранта РФФИ «Исследование и разработка принципов, методов и средств конвергенции естественно-научных и социогуманитарных ресурсов как составляющих единого электронного пространства знаний». В работе по этому гранту участвуют четыре организации РАН (ФИЦ ИУ, НИИСИ, БЕН и ИНИОН), каждая из которых реализует свой частный проект. В заключительной главе описываются исследования, которые выполняются ИНИОН РАН совместно с Институтом русского языка РАН им. В.В. Виноградова в рамках гранта № 18-00-00218 «Интеграция научно-информационных ресурсов учреждений РАН (на примере языкознания) как части единого цифрового пространства РАН».

5.9. Экономическая модель ЕРЭПЗ

Финансирование ЕРЭПЗ должно осуществляться из различных источников, включая субсидии из бюджетов разных уровней, грантовую и спонсорскую поддержку, а также поступления от монетизированных сервисов. ЕРЭПЗ должно функционировать на принципах частичной окупаемости.

Поскольку важнейшим принципом функционирования ЕРЭПЗ должен быть принцип открытой науки, экономическая модель ЕРЭПЗ должна опираться на экономические принципы открытой науки, которые в настоящее время формируются мировым научным сообществом.

Следует рассмотреть возможность создания специального российского фонда поддержки ЕРЭПЗ (эндаумента).

Для основных технологических процессов создания и функционирования ЕРЭПЗ необходимо разработать экономические нормативы и показатели. При разработке экономической модели инфраструктуры НТИ следует учесть, что ее создание позволит сократить многие существующие затраты на научно-информационные продукты, системы и процессы.

5.10. Правовые основы ЕРЭПЗ

Общие принципы создания и функционирования ЕРЭПЗ должны быть закреплены отдельным федеральным законом. В нем должен быть определен статус (правовой режим) ЕРЭПЗ в целом и его составных частей.

Приоритетный принцип функционирования ЕРЭПЗ – это открытость и общедоступность научной информации. Возможные исключения могут быть сформулированы в регламенте функционирования ЕРЭПЗ.

В настоящее время в российском законодательстве практически не затронут вопрос об использовании открытых научных данных, как первичных результатов исследований, так и извлеченных из научных публикаций. Здесь также целесообразно использование международной системы лицензирования Creative Commons, но, вероятно, модифицированной с учетом специфики научных данных.

Необходимо разработать правовой режим научных и образовательных электронных ресурсов, создаваемых в коллаборации

государственных научных организаций с негосударственными и зарубежными партнерами.

На правовом уровне следует обеспечить поддержку и сохранность ядра ЕРЭПЗ и основных научных и образовательных ресурсов. Причем сохранность этих ресурсов должна быть обеспечена независимо от организационных изменений среди участников ЕРЭПЗ. Особенно это актуально в отношении ресурсов, создаваемых при поддержке грантов.

В ходе проектирования ЕРЭПЗ должны быть сформулированы необходимые изменения в информационном законодательстве, а также в части 4 Гражданского кодекса и представлены законодателям.

5.11. Международное сотрудничество при создании и развитии ЕРЭПЗ

В процессе создания и дальнейшего становления ЕРЭПЗ чрезвычайно важно анализировать опыт развития систем научной информации стран, лидирующих в науке и образовании, и внедрять положительные результаты этого опыта. Необходимо учитывать, что международное сотрудничество в области научной информации достаточно чувствительно к общему состоянию международных отношений, а совместные инициативы отдельных научных организаций, групп и объединений научного сообщества в данной сфере являются неотъемлемой частью научной дипломатии. Кроме того, отчасти продолжает функционировать созданный еще в советское время задел в кооперации с центрами научно-технической информации ряда стран, прежде всего государств, входивших в состав СЭВ. Эти функции осуществляет Международный центр научно-технической информации.

Нет никаких сомнений, что современная российская система научной информации должна обеспечивать максимально полный и быстрый доступ к зарубежным источникам и массивам научно-технической информации.

Очень важным представляется взаимодействие ЕРЭПЗ с ведущими мировыми проектами по представлению знаний. Разумеется, такое взаимодействие следует осуществлять на равноправной (недискриминационной) основе. Ведущим участникам ЕРЭПЗ необходимо своевременно реагировать на значимые инициативы зарубежных партнеров в сфере научной информации и своевременно

доводить свои предложения по тематике этих инициатив до структур государственного управления, а также инициировать их экспертное обсуждение. К такого рода важным инициативам относится, например, «План S», разработанный в 2018 г. ассоциацией Science Europe – объединением ведущих исследовательских организаций и финансирующих агентств из 12 стран Европейского союза (см. главу 1).

В перспективе ЕРЭПЗ должно интегрировать всю научную информацию на русском языке, включая зарубежные русскоязычные ресурсы. Мы полагаем, что вполне оправданно стремление к включению в это информационное пространство научной информации на языках ряда других государств постсоветского пространства, включая – несмотря на удручающее состояние современных российско-украинских отношений – украинский язык. Соответственно, приоритетными для международного сотрудничества в сфере НТИ должны стать страны ЕАЭС, СНГ, ШОС и БРИКС.

Основными координаторами международного сотрудничества в сфере НТИ могут быть Международный центр научно-технической информации (МЦНТИ) и Евразийский информационно-аналитический консорциум.

5.12. ГСНТИ и ЕРЭПЗ

Таким образом, можно сформулировать один из центральных вопросов книги: означает ли формирование ЕРЭПЗ воссоздание в новых условиях и на другой технологической базе Государственной системы научно-технической информации (ГСНТИ)?

Мы исходим из того, что структурные составляющие ГСНТИ, которые продолжают функционировать, производят пользующиеся спросом научного сообщества информационные продукты и услуги, а также проводят оригинальные исследования научной инфосферы, нуждаются в поддержке и развитии, что позволит им успешно интегрироваться в новую цифровую реальность. В то же время очевидно, что далеко не весь прежний функционал ГСНТИ востребован российским научным сообществом и не все задачи, поставленные перед ГСНТИ в советскую эпоху, сохраняют свою значимость для государственных институций современной России. Поэтому следует иметь в виду не сохранение элементов ГСНТИ ради самого их сохранения, а формирование эффективно действующей национальной системы научной инфор-

мации, отвечающей актуальным политико-экономическим задачам российского государства. Как нам представляется, ЕРЭПЗ может выполнять эту функцию.

Сегодня решающую роль государства в научной инфосфере нельзя сводить к простому статистическому доминированию информационных организаций с государственной формой собственности. Создание ЕРЭПЗ, по нашему убеждению, предполагает организационную деятельность государства, направленную на повышение эффективности научно-информационного обеспечения исследований и разработок, включая минимизацию инфраструктурных издержек. В конечном счете речь идет об обеспечении значимого вклада инфраструктуры научно-технической информации в решение политико-экономических задач, связанных с достижением целей национального развития России в условиях ожесточенной глобальной конкуренции и перехода к цифровому обществу.

Глава 6.

ИНТЕГРАЦИЯ ЛИНГВИСТИЧЕСКИХ ИНФОРМАЦИОННЫХ РЕСУРСОВ¹

6.1. Введение

Исследование проблем интеграции научно-информационных ресурсов учреждений РАН по гуманитарным наукам (на примере языкознания) как части единого цифрового информационного пространства РАН является, с одной стороны, самостоятельным проектом, а с другой – частью комплексного проекта РФФИ КОМФИ «Разработка фундаментальных основ, методов и средств обеспечения конвергенции естественно-научных и социогуманитарных ресурсов как составляющих единого цифрового пространства знаний».

Проект по языкознанию включает как общие для комплексного проекта исследования, например разработку стратегии и методики оценки научных информационных ресурсов, так и специальные исследования социогуманитарного информационного пространства знаний, включая каталогизацию лингвистических информационных ресурсов РАН. Важной частью проекта является интеграция сведений по языкознанию, отраженных в информационно-поисковых языках лингвистических информационных ресурсов, в единую онтологию научного знания с целью их конвергенции с естественно-научными знаниями.

Конкретные задачи проекта:

- разработка стратегии для действий создателей информационных ресурсов и руководящих органов науки по интеграции информационных ресурсов в единое цифровое информационное пространство знаний;

¹ Данная глава включает результаты исследований по гранту РФФИ 18-00-00298 «Интеграция научно-информационных ресурсов учреждений РАН по гуманитарным наукам (на примере языкознания) как части единого цифрового информационного пространства РАН» (рук. А.Б. Антопольский).

- разработка методики оценки гуманитарных научных ресурсов РАН на примере ресурсов по языкознанию;
- проведение инвентаризации, классификации и каталогизации информационных ресурсов РАН по социальным и гуманитарным наукам;
- участие в комплексном проекте по разработке онтологии научного знания на основе имеющихся в ИНИОН РАН информационных ресурсов по языкознанию, тезаурусов, рубрикаторов и систем метаданных и включение этих ресурсов в единое цифровое информационное пространство РАН.

6.2. Анализ зарубежного опыта

Общая характеристика

Цифровые (электронные) лингвистические информационные ресурсы (далее – ЛИР) стали в последние десятилетия важнейшим инструментом как научной деятельности в области теоретического и прикладного языкознания, так и практики применения компьютерных языковых технологий в индустрии обработки текста и устной речи (далее – NLP¹).

Наличие, доступность и качество ЛИР в значительной степени определяют эффективность коммуникации ученых и практиков, а также коллабораций в трудоемких процессах извлечения и представления лингвистической информации. С другой стороны, создание ЛИР стало одним из существенных результатов научной деятельности в лингвистике. ЛИР также стали важным инструментом образовательной деятельности филолога [28].

Эти процессы характерны как для России, так и для всего развитого мира. Проведенная недавно инвентаризация информационных ресурсов РАН (см. главу 3) показала, что только в академических учреждениях России создано свыше 1 тыс. цифровых ЛИР. Эти ресурсы представлены в Навигаторе информационных ресурсов по языкознанию [62].

Имеется еще ряд каталогов отечественных ЛИР, перечисленных в приложении 16, а также в работе [106].

Несмотря на значительный объем и большое научное и практическое значение, деятельность по созданию, поддержке и использованию ЛИР в России практически не координируется. Не реали-

¹ Natural Language Processing (NLP).

зуется функция РАН по научно-методическому руководству созданием этой важной части научной продукции. ЛИР создаются в значительной степени хаотически, дублируют друг друга. Отсутствуют учет, мониторинг, экспертиза этой деятельности. Недостаточно развита коллаборация при создании ЛИР.

Многие ЛИР, особенно созданные по грантам, после завершения проекта лишаются финансовой поддержки и исчезают или теряют легитимность.

Особо нужно отметить, что создание ЛИР плохо учитывается при оценке эффективности научных учреждений.

Для решения задач повышения эффективности по созданию и использованию ЛИР автор предлагает создать специализированный Центр ЛИР при Институте русского языка РАН, являющемся бесспорным лидером отечественной деятельности в этой сфере. Концепция создания Центра ЛИР излагается в следующем разделе.

Переход к практическому созданию такого центра требует осмысления мирового опыта в этой сфере, чему и посвящен настоящий раздел.

Создание, поддержка, распространение, обмен ЛИР стали в последние годы мейнстримом международной коллаборации в компьютерной лингвистике и индустрии NLP. Далее мы приведем описание нескольких наиболее известных международных структур, основная цель которых – организация совместной деятельности по созданию и использованию ЛИР в науке и в индустрии NLP.

ELRA – Европейская ассоциация языковых ресурсов

Основанная в 1995 г., ELRA [124] является некоммерческой организацией, основная миссия которой заключается в том, чтобы сделать общедоступными лингвистические информационные ресурсы (ЛИР), применяемые в технологиях обработки естественных языков. Для достижения этой цели ELRA проводит такие действия, связанные с ЛИР, как идентификация ЛИР, распространение, производство, проверка, оценка технологий, информирование о ЛИР. При ELRA создано Агентство по оценке и распространению ЛИР (Evaluations & Language resources Distribution Agency – ELDA), которое является оперативным органом ассоциации и реализует выполнение задач ELRA, определенных Советом ассоциации, в том числе коммерческих проектов.

ELRA и ELDA реализуют следующие сервисы на основе ЛИР:

- идентификация и каталогизация ЛИР;
- распространение ЛИР;
- производство ЛИР;
- проверка и оценка ЛИР;
- поддержка правовых вопросов, связанных с ЛИР.

Идентификация. За последние несколько лет усилия ELRA по идентификации ЛИР значительно возросли. Это напрямую связано с акцентом деятельности ELRA на выявление уже существующих ресурсов, благодаря чему достигается оптимизация усилий по развитию уже доступных ресурсов.

Для задач идентификации разработан Международный стандартный номер языковых ресурсов (ISLRN) – уникальная и универсальная модель идентификации для ЛИР, использующая стандартизированные метаданные.

Каталогизация. В каталоге ELRA ЛИР распределены по четырём категориям: «Речь и смежные ресурсы», «Письменные ресурсы», «Терминологические ресурсы» и «Мультимодальные / мультимедийные ресурсы». Приведем некоторые примеры.

1. Устные ЛИР.

– Телефонные записи. Базы данных, каталогизированные в этом разделе, были созданы с записями говорящих, сделанными по телефонной (фиксированной или мобильной) сети или через микрофон, например базы данных, созданные в рамках проекта SpeechDat.

– Записи с рабочего стола / микрофона. Базы данных, каталогизированные в этом разделе, были созданы с использованием записей динамиков, сделанных через микрофон, например базы данных, созданные в рамках баз данных проекта BABEL.

– Ресурсы вещания. Базы данных, каталогизированные в этом разделе, были созданы с записями выступающих по радио, телевидению или через Интернет.

– Ресурсы, связанные с речью. Базы данных, в которых представлены варианты произношения и фонетические варианты лексики, например базы данных BDLEX, PHONOLEX и MHATLEX.

2. Письменные ЛИР.

– Корпуса. В этом разделе содержатся одноязычные и многоязычные корпуса, например корпуса, разработанные в рамках проекта MULTEXT, MLCC, Французские научные корпуса, газетные корпуса на арабском языке и т.д.

– Одноязычные лексиконы. Раздел, посвященный одноязычным лексиконам, содержит различные типы словарей, например

словарь французских глаголов, словарь японских слов, некоторые словари PAROLE на многих языках и т.д.

– Многоязычные лексиконы. Здесь вы можете найти как двуязычные, так и многоязычные словари и лексиконы, такие как базы данных EuroWordNet.

3. *Терминологические ЛИР*. Доступны одноязычные, двуязычные и многоязычные терминологические базы данных.

4. *Мультимодальные / мультимедийные ЛИР*. Ресурсы, представленные в этом разделе, были созданы с использованием различных модальностей, включая речь. Примером таких ресурсов является база данных, созданная в рамках проекта M2 VTS.

META-SHARE. Под эгидой ELRA разработано META-SHARE, средство обмена и распространения открытых ЛИР, которое направлено на расширение доступа к таким ресурсам в глобальном масштабе. META-SHARE – это открытое, интегрированное, безопасное и совместимое средство совместного использования и обмена для ЛИР (наборов данных и инструментов) для автоматической обработки естественного языка (NLP). META-SHARE спроектировано как сеть распределенных репозиторий ЛИР, включая языковые данные и базовые инструменты NLP.

Карта LRE. Для каталогизации ЛИР разработана модель их описания (карта LRE). Схема этой карты приведена в приложении 11. В соответствии с этой схемой описано свыше 6 тыс. ЛИР.

Конференция по языковым ресурсам и их оценке (LREC). Инициатива LREC «Поделитесь своими ЛИР», инициированная в 2014 г. в Рейкьявике и продолженная в 2016 г. в Портороже, имела большой успех. Наборы общих ЛИР были проверены вручную, и теперь доступны очищенные версии списков ЛИР, которые включают доступные ЛИР – наборы данных.

Специальная группа по интересам: «Языки с ограниченными ресурсами (SIGUL)». Созданная в 2017 г., SIGUL является совместной группой по интересам Европейской ассоциации языковых ресурсов (ELRA) и Международной ассоциации речевого общения (ISCA). Эта группа намерена сосредоточиться на конкретных потребностях и требованиях для языков с меньшими ресурсами.

OLAC – Сообщество открытых лингвистических архивов

OLAC, сообщество открытых языковых архивов [133], является международным партнерством учреждений и частных лиц,

которые создают всемирную виртуальную библиотеку языковых ресурсов путем:

1) выработки консенсуса в отношении наилучшей современной практики цифрового архивирования языковых ресурсов;

2) создания сети взаимодействующих хранилищ и служб для обеспечения доступа к таким ресурсам.

OLAC объединяет 60 лингвистических архивов из всех регионов мира, в которых содержатся до 300 тыс. описаний ЛИР, включающих данные на свыше чем 8 тыс. языков.

OLAC поддерживает фундаментальный каталог ЛИР самых разных типов. Имеется сводный анализ метаданных для ЛИР, включая инструменты поддержки и использования ЛИР. Также имеется развитая поисковая машина для каталога, позволяющая получать различную статистику.

OLAC уделяет большое внимание технологиям работы с ЛИР, протоколам и процедурам обмена ЛИР, а также разработке и поддержке системы метаданных. Основу метаданных составляет известный международный стандарт Dublin Core. OLAC поддерживает модель лингвистических связанных открытых данных с использованием технологий Семантического веба.

Приведем перечень основных документов OLAC.

Стандарты

Принципы и процессы OLAC-2006-04-05. Этот документ обобщает руководящие идеи OLAC (т.е. цель, видение и основные ценности), а затем описывает, как OLAC организован и как он работает.

Метаданные OLAC-2008-05-31. Этот документ определяет формат, используемый сообществом открытых языковых архивов OLAC для обмена метаданными в рамках инициативы открытых архивов OAI. Набор метаданных основан на полном наборе Дублинских основных терминов метаданных DCMT, но формат позволяет использовать расширения для выражения специфичных для сообщества квалификаторов.

Репозитории OLAC – 2008-07-28. Этот документ определяет стандарты, которым должны следовать архивы OLAC при реализации хранилища метаданных.

Рекомендации

Рекомендации по лучшей практике описания языковых ресурсов – 2008-07-11.

Словарь «Тип дискурса» – 2012-02-04.

Словарь «Лингвистические типы данных» – 2006-04-06.

Лингвистический предметный словарь – 2006-04-06.

Словарь ролей (Contributor в формате Dublin Core) – 2006-04-06.

Рекомендуемые расширения метаданных – 2008-02-22.

Методики и проекты

Запросы для выборочного сбора метаданных OLAC – 2003-07-29.

Создание записей OLAC и репозитория в Toolbox – 2006-09-28.

Метрики метаданных OLAC – 2009-06-29.

Руководство по использованию метаданных OLAC – 2008-07-11.

Спецификации формата отображения метаданных OLAC и перехода OLAC в OAI_DC – 2009-07-23.

Сторонние расширения метаданных – 2002-12-04.

Viser – средство отображения метаданных OLAC – 2003-07-29.

TELRI – Трансъевропейская инфраструктура лингвистических ресурсов

TELRI II – это общеевропейский альянс, состоящий из 28 основных национальных языковых (технологических) учреждений с акцентом на страны Центральной и Восточной Европы и СНГ [145]. От России в состав TELRI II входит Институт русского языка РАН.

Основные цели TELRI II:

- укрепить общеевропейскую инфраструктуру для сообщества многоязычных исследователей и разработчиков;

- собирать, продвигать и предоставлять одноязычные и многоязычные ЛИР, включая инструменты для NLP;

- предложить комплексный сервис для академических и промышленных пользователей;

- организовывать НИР и ОКР в области NLP;

- обеспечить форум, где эксперты из академических кругов и индустрии делятся и оценивают инструменты, ресурсы и новые

тенденции, исследуют альтернативные подходы и участвуют в совместной деятельности;

- предоставлять опыт своих партнеров исследовательскому сообществу, общественности и языковой индустрии.

Задачи и функции TELRI могут быть описаны через функции рабочих органов этой организации.

РГ 1: Координация. Финансовое, административное и договорное управление. Научный менеджмент.

РГ 2: TELRI Сеть. Подготовка, продвижение и организация семинаров TELRI. Информационный бюллетень, веб-страница. Сетевая организация TELRI. Рекламная деятельность. Пленарные заседания TELRI. Сотрудничество с партнерами из стран, не являющихся членами TELRI.

РГ 3: ТРАКТОР Сервис. Продвижение TRACTOR (TELRI Archive of Computational Tools and Resources) для академических и промышленных языковых ресурсов / сообщества языковых технологий. Реализация сети ТРАКТОР с использованием сети ELAN. Администрирование сети. Сервисный телефон доверия. Сервисный справочник. Информация о ресурсах и программном обеспечении, проекты, опыт и участие членов TELRI II. Каталог метаданных существующих ресурсов, предлагаемых партнерами проекта.

РГ 4: TRACTOR Инструменты и ресурсы. Координация приобретения инструментов и ресурсов для ТРАКТОРа. В частности, установление стандартов, которым должны соответствовать инструменты и их документация. Приобретение программных средств. Приобретение ресурсов.

РГ 5: Организация совместных исследований. Целями РГ 5 являются разработка, структурирование и подготовка многосторонних исследовательских и опытно-конструкторских проектов, отвечающих требованиям европейского многоязычного сообщества NLP. Вопросы для обсуждения: извлечение лингвистических знаний из текстовых и / или лексических ресурсов для задач в области многоязычной лексикографии, средств перевода, идентификации терминологии и поиска информации.

CLARIN – Общие языковые ресурсы и технологическая инфраструктура

CLARIN – исследовательская инфраструктура [119], которая была основана на представлении о том, что все цифровые языко-

вые ресурсы и инструменты со всей Европы и за ее пределами должны быть доступны через единый онлайн-вход для поддержки исследователей в области гуманитарных и социальных наук. CLARIN обеспечивает легкий и устойчивый доступ к ЛИР (в письменной, устной или мультимодальной форме).

CLARIN также предлагает расширенные инструменты для поиска, исследования, использования, аннотирования, анализа или объединения ЛИР. Эти функции реализуются через сетевую федерацию центров: хранилищ языковых данных, сервисных центров и центров знаний с единым входом для всех членов академического сообщества во всех участвующих странах. Инструменты и данные из разных центров являются функционально совместимыми, поэтому процессы сбора данных можно объединять, а инструменты из разных источников можно интегрировать для выполнения сложных операций. Инфраструктура CLARIN полностью функционирует во многих странах и продолжает создаваться в странах, которые присоединились совсем недавно. ЛИР и сервисы CLARIN постоянно обновляются и совершенствуются.

Консорциумы-участники. Большая часть операций, услуг и центров инфраструктуры CLARIN обеспечивается и финансируется членами CLARIN. Членами и наблюдателями могут быть страны или межправительственные организации. Они создают национальные консорциумы, обычно включающие университеты, исследовательские институты, библиотеки и публичные архивы, из которых по крайней мере одно учреждение имеет статус центра CLARIN (см. приложение 12). Члены и наблюдатели должны обеспечить доступ к создаваемым ими ЛИР, а также к инструментам и экспертным знаниям для работы с ними.

Все консорциумы придерживаются одинаковых критериев совместимости данных и услуг, условий доступа, качества данных и услуг. Функциональная совместимость обеспечивается с помощью стандартов, принятых в рамках CLARIN. Доступ к ЛИР соответствует принципам открытой науки. Члены и наблюдатели свободны решать, что именно они вносят в инфраструктуру CLARIN.

Депозитные услуги. Одним из основных сервисов инфраструктуры CLARIN является обеспечение архивирования ЛИР и предоставление их сообществу надежным способом. Чтобы помочь исследователям устойчиво хранить свои ЛИР (например, корпуса, лексиконы, аудио- и видеозаписи, грамматики и т.д.), многие центры CLARIN предлагают услугу депозита. Это имеет ряд преимуществ:

– долгосрочное архивирование: гарантия хранения может быть предоставлена на длительный период (в некоторых случаях до 50 лет);

– ЛИР удобно цитировать через идентификатор;

– ЛИР и их метаданные будут интегрированы в инфраструктуру, что позволит эффективно осуществлять их поиск;

– защищенные ЛИР будут доступны для легальных пользователей.

– если ЛИР интегрированы в инфраструктуру CLARIN, их можно легче обрабатывать с помощью различных лингвистических инструментов.

Виртуальная языковая обсерватория. Цель – предоставить простой в использовании интерфейс, обеспечивающий единый процесс поиска большого количества ЛИР. Развитый синтаксис запроса позволяет выполнять более целенаправленный поиск. Виртуальная языковая обсерватория позволяет создавать виртуальные коллекции на основе результатов поиска.

Легкий доступ к защищенным ресурсам. Благодаря федеративному входу в систему защищенные приложения и наборы данных доступны всем участникам из многих европейских стран, у которых есть учетная запись компьютера. Однако если нужно получить доступ к этим услугам из другой страны или из учреждения, которое не участвует в этих федерациях идентификации, можно запросить учетную запись CLARIN.

Коммутатор языковых ресурсов. Инструмент, который поможет найти соответствующий язык обработки веб-приложения для ваших данных. После загрузки файла или ввода URL вы можете выбрать, какую задачу выполнять. Затем коммутатор предоставит вам список доступных инструментов CLARIN для анализа ввода. Есть и другие способы вызова коммутатора. Коммутатор – не единственный способ использовать инструменты CLARIN. Примеры конкретных веб-сервисов приведены в приложении 13.

Виртуальные коллекции. Представляют собой последовательные наборы ссылок на цифровые объекты (например, размеченный текст, видео). Ссылки могут происходить из разных архивов, отсюда и термин «виртуальный». Виртуальная коллекция удобна для ручного доступа (с помощью веб-браузера), а также для автоматической обработки (например, веб-сервисом).

CLARIN предоставляет реестр, где ученые могут создавать и публиковать свои виртуальные коллекции. Он тесно интегрирован с инфраструктурой и обеспечивает постоянные идентификаторы и

федеративный вход в систему. Метаданные коллекции доступны через Виртуальную языковую обсерваторию.

Инвентаризация ЛИР. Предоставляется инструмент (LINDAT / CLARIN), удобный для инвентаризации (каталогизации) ЛИР. Данный каталог отличается от других депозитных услуг тем, что нет необходимости загружать данные или метаданные (достаточно ссылки на веб-сайт и описание) и его можно использовать немедленно, без предварительного обращения в хостинг-центр.

Каталог полностью интегрирован в инфраструктуру CLARIN. Его метаданные (собранные с 2008 г.) тщательно курируются и общедоступны. Данные могут быть представлены в стандартном формате CMDI. Долгосрочное архивирование гарантировано.

Федеративный поиск. Чтобы обеспечить исследователям возможность расширенного поиска с использованием конкретных моделей коллекции данных, CLARIN предлагает поисковую систему (пока прототип) данных, которые доступны в центрах хранения. Сами данные остаются у владельца, поэтому поиск называется федеративным. Поисковая система суммирует и отображает то, что доступно. Чтобы выполнить более сложный запрос, нужно перейти к специализированному интерфейсу поиска в центре – владельце ЛИР. Технологией этого федеративного поиска контента является SRU / CQL, и CLARIN – специфичное расширение этого протокола. Федеративный поиск отличается от поиска метаданных, например, в Виртуальной обсерватории, где все метаданные сначала собираются (копируются на один сервер), а затем индексируются централизованно. Федеративный поиск применяется по нескольким причинам:

- правовые проблемы делают невозможным копирование некоторых ЛИР;

- размер многих ЛИР делает децентрализованную индексацию наиболее приемлемым вариантом;

- большинство ЛИР индексируются в зависимости от коллекции, что затрудняет использование или разработку единой поисковой системы.

Хотя федеративный поиск контента является масштабируемым, он обходится дешевле, чем локальный поиск, и некоторые функции отсутствуют, например ранжирование. Поэтому федеративный поиск будет особенно полезен в качестве первого шага для определения того, где размещены интересные ЛИР и в каких центрах можно использовать более специализированный поиск.

CLARIN для исследователей. CLARIN для исследователей – это онлайн-коллекция учебных материалов, тематических исследований и контактов с экспертами из всей сети CLARIN, которые предназначены для исследователей и студентов всех ступеней, которые работают в области цифровой гуманитаристики.

Ресурсные семьи. Целью данной инициативы CLARIN является предоставление обзоров доступных ЛИР для исследователей в области цифровой гуманитаристики и NLP. Обзоры организованы в соответствии с типами ЛИР и языком. Помимо ЛИР, имеющих в инфраструктуре CLARIN, обзор включает сведения о других ценных ЛИР, которые еще не были интегрированы в CLARIN. В настоящее время предлагаются обзоры следующих девяти семейств корпусов и пяти семейств лексических ресурсов.

Корпуса:

- исторические корпуса;
- учебные корпуса;
- литературные корпуса;
- газетные корпуса;
- параллельные корпуса;
- вручную размеченные корпуса;
- парламентские корпуса;
- речевые корпуса;
- корпуса блогов.

Лексические ресурсы:

- лексиконы;
- словари;
- концептуальные ресурсы;
- глоссарии;
- списки слов.

Реестр курсов по цифровой гуманитаристике. Содержит выбор курсов, предлагаемых европейскими академическими организациями. Платформа для обмена информацией о курсах совместно разработана CLARIN ERIC и DARIAH-EU. Студенты, преподаватели и исследователи могут осуществлять поиск в базе данных на основе дисциплин, местоположения, кредитов ECTS или приружаемых академических степеней.

Обмен знаниями. В основе инфраструктуры обмена знаниями лежит техническая инфраструктура, включающая как технические средства, которые предоставляют пользователям доступ к данным и инструментам, так и людей, которые управляют этими средствами. Имеется набор общих согласованных организационных пра-

вил, мер и соглашений, которые должны обеспечивать бесперебойное взаимодействие между пользователями инфраструктуры, операторами и компонентами, включая, например, стандарты, условия доступа, лицензии, обеспечение качества и т.д.

Мероприятия. CLARIN ежегодно проводит 12 конференций, семинаров и др.

LDC – Консорциум лингвистических данных

Миссия. Консорциум лингвистических данных (Linguistic Data Consortium – LDC) [130] реализует принцип, согласно которому широкий доступ к данным стимулирует инновации. LDC поддерживает образование, исследования и развитие технологий путем создания и обмена ЛИР, такими как данные, инструменты и стандарты.

Языковые ресурсы (ЛИР) – это коллективные материалы, используемые теми, кто занимается языковым образованием, исследованиями и развитием технологий. Можно выделить следующие основные типы ЛИР.

- *Данные.* Предоставление ЛИР этого типа сообществу – основа деятельности LDC. Ежегодно адаптируются и рассылаются подписчикам и членам LDC 30–36 ЛИР.

- *Инструменты* содержат программное обеспечение, разработанное и распространяемое LDC. Все инструменты доступны бесплатно по лицензиям, предполагающим открытый исходный код. Условия лицензий могут отличаться.

- *Документы* содержат публикации (с 1998 г.) сотрудников LDC об их работе, а также доклады, презентации, статьи и книги.

Каталог LDC содержит описания около 1000 ЛИР. Возможен поиск в каталоге по названию корпуса, номеру в каталоге, языку и ограничения поиска по году, по проектам, по членству в LDC, по типу ЛИР на основе типологии Дублинского ядра метаданных.

Получение данных. Членство в LDC обеспечивает организациям права, привилегии и скидки на доступ. Для организаций, не являющихся членами, большинство ЛИР, распространяемых LDC, доступны на основании лицензий «только на исследования» и облагаются лицензионным сбором. ЛИР «только для членов» доступны для текущих участников, которые могут запрашивать данные с указанной сниженной лицензионной оплатой.

Стипендии данных. Программа стипендий для данных предоставляет студентам, имеющим на это право, бесплатный доступ к ЛИР LDC. Эта программа субсидируется LDC, чтобы помочь талантливым людям выполнить задачи, связанные с исследованием языка.

Управление данными. Эффективность NLP в немалой степени зависит от доступности ЛИР, удобства использования и возможности архивирования. LDC предоставляет информацию о «распаковке» ЛИР, извлечении файлов и других способах использования данных. Предлагается простая, проверенная временем модель лицензии LDC. В сервисе управления данными объясняется, как LDC помогает исследователям в разработке и реализации планов управления данными для конкретных проектов. Наконец, есть инструкции для цитирования корпусов LDC в библиографии.

Сотрудничество. Приводится перечень текущих и законченных проектов в области применения ЛИР, организованных LDC. Особое место занимает направление – оценка технологий, реализуемых с помощью ЛИР, предоставленных LDC.

ISCA – Международная ассоциация речевого общения

Целью ассоциации ISCA [129] является содействие в международном мировом контексте деятельности и обмену во всех областях, связанных с наукой и технологиями речевого общения. Ассоциация предназначена для всех лиц и учреждений, заинтересованных в фундаментальных исследованиях и технологическом развитии, целью которых являются описание, объяснение и воспроизведение различных аспектов человеческого общения с помощью речи. Прежде всего, это касается фонетики, лингвистики, компьютерной речи, распознавания и синтеза, компрессии речи, распознавания говорящего, средств медицинской диагностики патологий голоса.

Задачи ISCA:

- стимулировать научные исследования и образование;
- организовывать конференции, курсы и семинары;
- публиковать и содействовать публикации научных работ;
- содействовать обмену научными взглядами в области речевого общения;
- поощрять изучение разных языков;
- сотрудничать со всеми связанными ассоциациями;

– исследовать промышленное применение результатов исследований;

– в более общем плане – содействовать развитию отношений между государственным и частным секторами, а также между наукой и техникой.

SIL – Летний институт лингвистики

Миссия SIL [143] – служение языковым сообществам в качестве международного защитника. SIL обслуживает языковые сообщества по всему миру, наращивая их потенциал для устойчивого развития языка посредством исследований, перевода, обучения и разработки материалов. SIL участвует в 1660 языковых проектах в 162 странах. Обладает значительными языковыми ресурсами. ЛИП SIL включают коллекции данных, публикуемых SIL различными способами на региональном уровне. В каталог SIL также включены некоторые сайты с широким охватом языковых семей, не входящие в SIL. Раздел «Лингвистическое программное обеспечение» посвящен, прежде всего, программному обеспечению, связанному с развитием языка, например используемому для лингвистического анализа и организации данных.

Национальные центры ЛИП

Национальные центры или консорциумы ЛИП созданы практически во всех европейских странах, а также во многих странах за пределами Европы; их обзор занял бы слишком большой объем. Перечень национальных центров ЛИП, входящих в сеть CLARIN, представлен в приложении 12. В настоящем обзоре мы ограничимся описанием деятельности одной, наиболее развитой национальной системы ЛИП, а именно французской системы ORTOLANG, входящей в программу TGIR Huma-Num¹.

¹ Название программы расширяется как «Очень большая научная инфраструктура для цифровой гуманитаристики».

ORTOLANG – Открытые ресурсы и инструменты для обработки языка

ORTOLANG [137] объединяет 220 университетов, научных и промышленных организаций, работающих в сфере NLP. ORTOLANG является французским координатором европейской сети CLARIN.

ORTOLANG осуществляет следующие функции.

Идентификация / подготовка данных:

– каталогизация ЛИР с помощью набора стандартных метаданных;

– мониторинг и оценка ЛИР;

– поддержка текущих международных стандартов и рекомендаций;

– поддержка развития ЛИР.

Архивное хранение ЛИР:

– хранение, обслуживание и поддержка ресурсов и инструментов;

– постоянное архивирование на основе решений TGIR Huma-Num.

Распространение ЛИР:

– поддержка пользователей;

– внедрение процедур, позволяющих пользователям использовать многопользовательские ресурсы и инструменты.

Модель ORTOLANG использует базовые объекты модели Open Archival Information System (OAIS).

ORTOLANG интегрирует достижения науки о языке, базы данных и доступ к ЛИР через два ресурсных центра – CNRTL и SLDR. Цель ORTOLANG также заключается в объединении ЛИР партнеров, реализующих разные подходы к изучению языка.

Ресурсы ORTOLANG:

□ корпуса – 221 ресурс;

○ письменные – 33;

○ устные – 162;

○ мультимодальные – 26;

□ лексика – 16 ресурсов;

□ терминология – 23 ресурса;

□ инструменты – 34 ресурса;

□ интеграционные проекты – 4 ресурса;

□ количество участников – 220 учреждений.

Все ресурсы, депонированные на ORTOLANG, собираются через виртуальную языковую обсерваторию CLARIN и доступны с помощью фасетного поиска.

В рамках ORTOLANG реализуются четыре интеграционных проекта;

- CNRTL (Centre National de Ressources Textuelles et Lexicales) – национальный центр текстовых и лексических ресурсов¹;

- SLDR (Speech & Language Data Repository) – репозиторий речевых и языковых данных²;

- портал для поиска в CNRTL и SLDR³;

- тематический центр устных и мультимодальных корпусов Нантьер-Орлеан⁴.

*Ресурсы SLDR / ORTOLANG*⁵

Количественные данные:

- количество ЛИР: 341;

- документы: 790856;

- скачиваний: 47013;

- участники: 767 из 61 страны;

- языки общения: 175.

Типы ресурсов. ЛИР сгруппированы в четыре основных типа элементов:

- первичные данные;

- вторичные данные;

- инструменты;

- коллекции.

Созданные описания ЛИР соответствуют основным стандартам метаданных: они могут собираться международными каталогами и поисковыми системами. ЛИР назначаются постоянные идентификаторы.

Корпуса. Модель описания:

- имя (ссылка на детальное описание);

- координатор, лаборатория, контакт;

- тематика;

- назначение;

- объем данных;

¹ <https://www.cnrtl.fr>

² <http://www.sldr.org>

³ <https://www.ortolang.fr/market/applications/portail>

⁴ <http://modyco.inist.fr/>

⁵ <http://www.sldr.org/>

- языки;
- дата записи;
- количество дикторов;
- метаданные;
- долгосрочное хранение.

Вторичные данные. Пример описания:

miscellaneous – prax000929 => wiki Corpus St-Maur (documents) (Praxiling – UMR 5267 (Montpellier FR)) Documents et travaux portant sur le corpus “Corpus St-Maur“ (applied_linguistics, anthropological_linguistics) French.

Инструменты. Пример описания:

software – sldr000841 => wiki (Misc publications) MarsaTag (Stéphane Rauzy, CNRS – LPL) Laboratoire parole et langage – UMR 7309 (LPL, Aix-en-Provence FR) Tools for textual data enrichment (written text and speech transcription): tokenizer, morphosyntactic & POS tagger,.. (general_linguistics, computational_linguistics) French.

Коллекции – наборы первичных данных, ресурсов, инструментов и рекурсивных коллекций. Пример описания:

sldr000804 => wiki Travaux d'étudiants Master LEX. Laboratoire parole et langage – UMR 7309 (LPL, Aix-en-Provence FR). Département de sciences du langage, Université d'Aix-Marseille (Aix-en-Provence FR) Travaux des étudiants Master LEX à l'Université d'Aix-Marseille (computational_linguistics).

Публикации. Приводятся библиографическое описание, реферат и ссылка на ЛИР.

Терминологические веб-сайты и банки данных

Среди ЛИР и, соответственно, центров поддержки ЛИР следует выделить терминологические ЛИР (веб-сайты, службы, банки данных), которые по количеству существенно превосходят другие типы ЛИР. Это объясняется тем, что терминологические ЛИР во множестве создаются не лингвистическими, а отраслевыми структурами как на национальном, так и на международном уровне. Терминологические ЛИР обычно не ориентированы на NLP, хотя и могут использоваться в этих технологиях. В приложении 14 приводятся адреса и аннотации 20 наиболее известных терминологических служб и ресурсов. Нужно отметить, что в их число включен и один российский терминологический банк (РОСТЕРМ).

Каталоги ЛИР

Наиболее полные сведения о количестве и номенклатуре ЛИР можно получить из каталогов ЛИР. Таких каталогов в Интернете представлено достаточно много. Эти каталоги существенно различаются по объему (от нескольких десятков до нескольких тысяч цифровых объектов), по глубине и детальности описания, по возможностям поиска. Но главное – в исследованных каталогах существенно различается понятие ЛИР. Например, включаются ли в понятие ЛИР сайты учреждений и отдельных лиц, инструментальные средства NLP, документные массивы, блоги, списки персон-лингвистов и другие типы ЛИР.

Нужно отметить, что международные и национальные каталоги ЛИР существенно различаются и практически не пересекаются. В настоящем обзоре приводится перечень из 15 международных (приложение 15) и 10 отечественных (приложение 16) каталогов ЛИР.

Из представленного обзора с очевидностью следует, что деятельность не только по созданию ЛИР, но и по обеспечению их поддержки, открытости, доступности и совместимости, стала важным направлением индустрии NLP. К сожалению, Россия недостаточно представлена на мировом рынке услуг в области ЛИР. Количество и качество отечественных ЛИР позволяют претендовать на гораздо более достойное место.

6.3. Концепция создания Центра лингвистических ресурсов РАН

Постановка задачи

Быстрое развитие информационных технологий в лингвистике как науке, а также необходимость решения различных прикладных задач, использующих методы и средства компьютерной лингвистики (таких как обработка текста, анализ и синтез речи, автоматический перевод, обучение языкам и др.), привели к созданию большого числа электронных лингвистических информационных ресурсов (ЛИР) различного назначения. Деятельность по их формированию, обеспечению доступа, сохранности, открытости, возможности повторного использования требует оптимизации, координации и системного подхода, т.е., обобщенно говоря, – управления деятельностью в сфере ЛИР.

Управление деятельностью в сфере ЛИР должно стать частью Цифровой системы управления сервисами научной инфраструктуры коллективного пользования (АС УСНИКП) [54], создание которой предусмотрено в Национальном проекте «Наука».

Создание системы управления ЛИР обеспечит значительный экономический эффект.

Предлагается решать эти задачи на основе создания Центра лингвистических ресурсов РАН при Институте русского языка РАН.

Определения

Лингвистические информационные ресурсы (ЛИР) – организованные массивы речевых и языковых данных, находящихся на машинных носителях, предназначенных для использования в науке и различных сферах практической деятельности, а также массивы документов, создаваемых как результат научных исследований в сфере языкознания.

Интеграция информационных ресурсов – создание централизованных ресурсов, полностью включающих данные и функции интегрируемых ресурсов.

Агрегация – создание распределенных ресурсов, в которых централизованы метаданные и некоторые функции, как правило, поиск и навигация.

Состояние создания ЛИР в РАН

Зарубежный опыт по созданию центров ЛИР, достаточно обширный, описан в предыдущем разделе.

Состояние ЛИР в России может быть представлено на основании мониторинга академических информационных ресурсов. В результате мониторинга был создан Навигатор информационных ресурсов по языкознанию (НИРЯЗ) [62]. Количество электронных ЛИР по каждому типу, отраженных в НИРЯЗ по состоянию на декабрь 2019 г., приведено в табл. 8.

Следует указать, что эти данные показывают количество не только собственно ресурсов по языкознанию, но также универсальных по тематике ресурсов, включающих лингвистические ресурсы. Особенно это касается документных ресурсов (библиографии, каталоги, электронные библиотеки, периодика).

**Функциональная типология и статистика
лингвистических информационных ресурсов в РАН**

№	Типы лингвистических ресурсов	Кол-во
1.	Вторичные документы (библиографии, библиотечные каталоги, описи, каталоги ссылок)	646
2.	Электронные коллекции и библиотеки полных текстов (книги, диссертации, отчеты, труды конференций и др.)	383
3.	Периодические, продолжающиеся издания и архивы периодики	77
4.	Корпуса	15
5.	Лексикографические ИР	83
6.	Этнолингвистические и социоллингвистические БД	10
7.	Лингвистические географические системы, атласы	5
8.	Электронное представление памятников письменности	5
9.	Активные ресурсы (алгоритмы, процессоры, программы)	28
10.	Граматики	11
11.	Описания языков и комплексные лингвистические сайты	14
12.	Информационные языки	10
13.	Энциклопедии, справочники, реестры языков	75
14.	Сведения об отдельных персонах (сайты и страницы ученых, личные фонды, биобиблиографии)	473
15.	Списки, перечни, указатели персон	31
16.	Медиаресурсы	8
17.	Сайты учреждений – владельцев ЛИР	116
18.	Сайты-спутники и ресурсы во внешних АИС	46
	Всего	2036

Приведенные данные свидетельствуют о том, что в академических учреждениях создается значительное число лингвистических информационных ресурсов, поэтому координация деятельности по их созданию даст значительный экономический эффект.

Конечно, нужно иметь в виду, что множество документных ресурсов по языкознанию имеются в общенациональных информационных системах, таких как Национальная электронная библиотека, Электронная библиотека диссертаций, КиберЛенинка, Научная электронная библиотека и др. Поэтому в рамках взаимодействия с этими организациями следует выработать решения, минимизирующие дублирование при обработке соответствующих документов.

Значительный объем ресурсов по языкознанию практически всех категорий представлен в информационных системах вузов, а также в различных общественных и коммерческих проектах, которые на данном этапе не рассматриваются.

Типология лингвистических ресурсов

Функциональная типология электронных лингвистических ресурсов, разработанная на основе мониторинга российских академических ресурсов, представлена в табл. 8. В данном варианте она несколько упрощена по сравнению с двухуровневой типологией, представленной в Навигаторе и описанной выше.

С точки зрения интероперабельности и открытости данных можно разделить ресурсы на категории, например на основе пяти-звездочной модели Тима Бернерса Ли:

- уровень «1 звезда»: представление данных в любом формате с открытой лицензией (например, скан документа);
- уровень «2 звезды»: представление данных в структурированном виде (например, таблица Excel);
- уровень «3 звезды»: представление данных в неproprietary формате (например, формат CSV вместо Excel);
- уровень «4 звезды»: Данные представлены в открытых стандартах консорциума W3C (RDF и SPARQL), предназначенных для идентификации данных;
- уровень «5 звезд»: размещение связанных данных, когда любые данные могут быть связаны между собой.

Также выделяют активные и пассивные лингвистические ресурсы. К пассивным формам относят словари, письменные текстовые массивы (корпуса текстов), фонетические ресурсы, электронные библиотеки и т.д.,

Активные формы – алгоритмы, модели, программы, базы знаний

В ходе мониторинга использовалась также тематическая классификация информационных ресурсов на основе Государственного рубрикатора научно-технической информации (ГРНТИ). Подробнее вопросы применения ГРНТИ для навигации по ЛИР рассмотрены в разделе 6.4.

Задачи Центра лингвистических ресурсов

1. Мониторинг, учет, каталогизация российских ЛИР.

2. Экспертиза ЛИР, в том числе обеспечение экспертной функции РАН в отношении создания ЛИР как результатов научной деятельности.

3. Координация деятельности по созданию ЛИР, стратегическое планирование деятельности по созданию ЛИР.

4. Интеграция и агрегация ЛИР (по отдельным типам ЛИР).

5. Организация открытого доступа к ЛИР и его поддержка.

6. Архивирование и обеспечение сохранности ЛИР.

7. Разработка стандартов, типовых программных средств, метаданных и методик по созданию ЛИР.

8. Создание и поддержка централизованных сервисов для создателей и пользователей ЛИР, в том числе монетизированных.

9. Организация и выпуск журнала по проблемам прикладной и компьютерной лингвистики, а также ЛИР.

10. Организация площадки для обмена опытом в области создания и использования ЛИР.

11. Создание онтологии научного знания по лингвистике, в том числе в рамках создания Единого российского электронного пространства знаний.

12. Взаимодействие с универсальными российскими системами научной информации.

13. Взаимодействие с международными и зарубежными национальными центрами лингвистических ресурсов.

Этапность создания и развития Центра лингвистических ресурсов

• 1 этап – решение перечисленных задач относительно ЛИР, создаваемых в учреждениях РАН.

• 2 этап – решение перечисленных задач относительно ЛИР, создаваемых в государственных научных и образовательных учреждениях.

• 3 этап – создание национального Центра лингвистических ресурсов, решение перечисленных задач относительно всех российских ЛИР.

Целесообразность интеграции и агрегации ЛИР по отдельным типам

Возможность и целесообразность интеграции или агрегации ЛИР с целью создания централизованных многофункциональных ресурсов или сервисов существенно различаются для каждого типа ресурсов, с учетом различия языков, структуры и назначения ресурсов. Кратко рассмотрим эту возможность для каждого типа ЛИР.

Библиографические ресурсы учреждений РАН, включая тематические и предметные библиографии, перечни трудов сотрудников и учреждений, включая неопубликованные документы, а также описи архивных фондов, бесспорно, нуждаются в интеграции. Отчасти эта задача решается в универсальных библиографических системах-агрегаторах (ЛИБНЕТ, ЭКБСОН, АРБИКОН), но представляются очевидными преимущества создания централизованной библиографической БД для учреждений и сотрудников РАН, работающих в сфере языкознания. Эта библиографическая БД должна выполнять функции корпоративной каталогизации в учреждениях РАН, учета научных трудов и публикаций учреждений и сотрудников РАН, обеспечивать наукометрические исследования, обеспечивать подготовку отчетности для вышестоящих организаций. Прототипом такой БД может служить АИС «ИСТИНА», подобные прототипы имеются также в ряде научных учреждений РАН. В создании библиографической БД по языкознанию мог бы принять активное участие ИНИОН РАН.

Библиографическая БД должна также служить инструментом координации программ оцифровки научных документов в библиотеках и архивах РАН.

Полнотекстовые документные ресурсы, включающие как опубликованные, так и неопубликованные документы, также могут быть собраны в единую отраслевую электронную библиотеку, которая может выполнять и функцию тематического репозитория. На первом этапе в эту библиотеку могут быть включены документы, находящиеся в открытом доступе, а далее в нее поэтапно могут включаться документы, авторы и правообладатели которых осознают преимущества открытого доступа. Предлагаемая электронная библиотека по языкознанию могла бы развиваться во взаимодействии с Национальной электронной библиотекой, КиберЛенинкой, Электронной библиотекой диссертаций и другими общероссийскими проектами научной информации.

Предлагаемая электронная библиотека могла бы также служить базой для выявления плагиата в трудах по языкознанию. Настройка алгоритмов антиплагиата на специфику лингвистических исследований позволит существенно улучшить оценку научных трудов в этой области.

В любом случае количество коллекций научных документов в учреждениях РАН свидетельствует о необходимости координации и оптимизации деятельности по их созданию, предоставлению в доступ, исключению дублирования и другим задачам. Важной функцией Электронной библиотеки по языкознанию могло бы стать архивирование и обеспечение долговременного хранения научных документов *born digital*.

Периодические издания по языкознанию вряд ли нуждаются в создании отдельного централизованного хранилища, поскольку эти функции уже выполняются Научной электронной библиотекой, а также КиберЛенинкой. Однако включение поискового аппарата этих проектов в общую навигационную систему Центра лингвистических ресурсов существенно повысит качество поиска научной литературы.

Корпуса

Для лингвистических корпусов в Центре ЛИР должна быть обеспечена единая система навигации, основанная на единой системе метаданных (тегов) для всех видов корпусов и для различных языков.

Лексикографические ресурсы представляют собой один из важнейших видов ЛИР. С одной стороны, их в учреждениях РАН достаточно много, с другой – создание и актуализация лексикографических ресурсов весьма трудоемки. На первом этапе представляется возможным выстроить качественную навигацию этих ресурсов на основе поиска по метаданным имеющихся ресурсов и сквозного лексического поиска. В перспективе, вероятно, имеет смысл рассмотреть создание централизованных лексикографических БД по отдельным языкам, а также по межъязыковым словарям. Отдельной задачей является передача лексикографических ресурсов в открытый доступ в соответствии со стандартами открытых данных.

Этнолингвистические и социоллингвистические БД, очевидно, вряд ли нуждаются в интеграции в единый ресурс, однако для них возможно построить удобный общий интерфейс.

Лингвистические географические системы, атласы, электронные представления памятников письменности, активные ЛИР (алгоритмы, процессоры, программы) не требуют интеграции. Для этих категорий ресурсов достаточно создать полный и актуальный каталог, возможно, включающий некоторые инфометрические или другие оценочные параметры, которые позволят организовать экспертизу новых разработок этих же классов. Также могут быть предложены инструменты, обеспечивающие открытость этих данных, а также организовано их архивирование.

Граматики, описания языков и комплексные лингвистические сайты должны соответствовать некоторым стандартам представления данных. Эти стандарты должны быть разработаны или адаптированы Центром лингвистических ресурсов. Аналогично другим типам ЛИР могут быть предложены инструменты, обеспечивающие открытость этих данных, а также обеспечено их архивирование

Информационные языки, которые будут собраны в Центре лингвистических ресурсов, в том числе тезаурусы, рубрикаторы, классификаторы, системы и словари метаданных, используемые в различных ЛИР, могут быть в перспективе интегрированы в онтологию научного знания по лингвистике. Эта онтология сможет служить для навигации (метапоиска) во всех ЛИР, доступ к которым будет обеспечен через Центр лингвистических ресурсов. Указанная онтология может служить также частью общей онтологии научного знания, которая должна быть создана в рамках Единого российского электронного пространства знаний. В перспективе онтология научного знания по лингвистике может служить инструментом для выявления нового знания в лингвистических трудах.

Лингвистические энциклопедии и справочники, включая реестры языков, желательно интегрировать в общую энциклопедическую базу данных, создание которой, во-первых, повысит качество энциклопедических сведений за счет выявления разночтений и ошибок; во-вторых, предоставит дополнительные возможности для справочно-информационного обслуживания пользователей; в-третьих послужит важным источником данных для упомянутой онтологии. Как и для других типов ЛИР, могут быть предложены инструменты, обеспечивающие открытость энциклопедических данных, а также обеспечено их архивирование в электронном виде.

Персональные (просопографические) данные об ученых-лингвистах и других заметных личностях должны обрабатываться в соответствии с теми же правилами, что и другие энциклопедические данные. По существу, эти данные должны формировать отдельную (персональную) подсистему энциклопедической БД и – в перспективе – онтологии научного знания по языкознанию. Существенным отличием этих данных является то, что они должны отвечать требованиям российского законодательства о персональных данных. Выполнять эти требования проще в централизованной системе, поэтому Центр лингвистических ресурсов может предложить соответствующий сервис учреждениям РАН.

Медиаресурсы, включая банки изображений, аудио- и видеoinформацию, а также кинофильмы, на данном этапе достаточно учитывать и каталогизировать. Возможно, следует организовать сквозной поиск по библиографическим описаниям медиadoкументов. Следует учитывать, что в Институте русской литературы (Пушкинский дом) РАН создан и функционирует Национальный электронный звуковой депозитарий, в котором хранится большинство диалектологических и этнографических аудиодокументов. Естественно, дублировать эту деятельность нецелесообразно.

Прочие интернет-ресурсы должны быть проанализированы дополнительно, и решение о включении этих ресурсов в тот или иной режим интеграции и агрегации должно приниматься индивидуально. Однако в каталог ЛИР эти ресурсы должны быть включены.

Организационные аспекты создания Центра лингвистических ресурсов

Обязательным условием создания центра в любой организационной форме должна быть поддержка РАН и Министерства науки и высшего образования РФ. При получении предварительного одобрения должны быть подготовлены проекты устава центра и технико-экономическое обоснование функционирования центра.

Важным условием практического функционирования центра является его включение в состав научной инфраструктуры в соответствии с Национальным проектом «Наука», а именно – Федеральным проектом «Развитие передовой инфраструктуры для проведения исследований и разработок в Российской Федерации» [80].

Центр лингвистических ресурсов должен входить в состав Цифровой системы управления сервисами научной инфраструкту-

ры коллективного пользования, создание которой предусмотрено данным федеральным проектом, п. 1.4.

Центр лингвистических ресурсов должен также входить в состав современной цифровой инфраструктуры для хранения и анализа научно-технической информации, а также для обмена такой информацией, создание которой предусмотрено Поручением Президента России от 29.12.2018 г. [81].

Могут быть предложены две организационные формы для создания центра:

1) обособленное (хозрасчетное) подразделение Института русского языка РАН;

2) самостоятельная некоммерческая организация в форме автономной некоммерческой организации (АНО) или фонда, учредителями которой выступят Институт русского языка и / или ФГБУ РАН. Возможно также привлечение в качестве учредителей других институтов РАН лингвистического профиля.

Окончательный выбор организационно-правовой формы должен быть сделан после консультаций в вышестоящих организациях.

Технологические аспекты функционирования Центра лингвистических ресурсов

Технологические аспекты функционирования центра должны быть проработаны в соответствии с общими требованиями к Цифровой системе управления сервисами научной инфраструктуры коллективного пользования.

Конечно, целесообразно при выборе технологической платформы учесть зарубежный опыт. Так, например, Сообщество лингвистических архивов OLAC (см. раздел 6.2.3) в качестве технологической платформы использует основные стандарты Семантического веба: стандарт метаданных Dublin Core, протокол OAI-PMH, модель данных LOD. Такую платформу автор считает предпочтительной.

Экономические аспекты создания и функционирования Центра лингвистических ресурсов

Основой функционирования центра должно быть бюджетное финансирование, обеспечивающее существенную долю (вероятно, не менее 50%) расходов центра.

Предпочтительной является форма субсидий, предполагающих выполнение государственного задания по формированию, поддержке и предоставлению в открытый доступ лингвистических информационных ресурсов, определенных в государственном задании.

Остальные расходы должны быть компенсированы из других источников, а именно:

- программы Президиума РАН;
- участие в других программах, предполагающих использование лингвистических ресурсов, особенно в программах искусственного интеллекта;
- гранты фондов, поддерживающих научные исследования;
- хозяйственные договора, предполагающие создание и развитие дополнительных монетизированных сервисов.

Правовые аспекты создания и функционирования центра

Основой создания и функционирования должны быть:

- учредительные документы по созданию центра;
- устав центра.

Взаимодействие центра с создателями ЛИР в рамках РАН и Минобрнауки осуществляется на основе нормативно-методических документов Минобрнауки и РАН, а также договоров о совместной деятельности между центром и партнерскими организациями.

Взаимодействие центра с общегосударственными проектами в области научной информации определяется правительственными документами, а также договорами о совместной деятельности между центром и партнерскими организациями.

Взаимодействие центра с международными и зарубежными проектами по созданию и использованию ЛИР определяется международными соглашениями, а также договорами между центром и соответствующими организациями.

Экспертиза ЛИР и проектов по их созданию и использованию осуществляется на основе нормативно-методических документов Минобрнауки и РАН.

6.4. Навигатор информационных ресурсов по языкознанию (НИРЯЗ)

Структура и функциональность НИРЯЗ

НИРЯЗ разработан¹ в рамках проведения исследований по гранту РФФИ № 18-00-002\18 «Интеграция научно-информационных ресурсов учреждений РАН (на примере языкознания) как части единого цифрового пространства РАН». НИРЯЗ является модификацией Навигатора информационных ресурсов по общественным наукам (НИРОН), описанного выше, в главе 3.

НИРЯЗ предназначен для сбора и организации сведений о лингвистических информационных ресурсах РАН. Эти сведения должны быть использованы для разработки концепции и проектирования Центра лингвистических ресурсов РАН как составной части Единого российского электронного пространства знаний. Сведения, содержащиеся в НИРЯЗ, могут быть использованы для навигации в информационном пространстве РАН, а также для наукометрических исследований.

Содержание информации, включенной в НИРЯЗ

НИРЯЗ содержит перечень ресурсов по языкознанию, а также специфических лингвистических ресурсов, создаваемых учреждениями РАН и республиканских академий наук. Особенностью НИРЯЗ является то, что он включает сведения как о традиционных ресурсах (библиотечных, архивных, музейных), так и об электронных (электронные библиотеки, базы данных, информационные системы, сайты, социальные сети и др.). В НИРЯЗ включены сведения о коллекциях и массивах документов и данных, и лишь в некоторых случаях – сведения об отдельных документах. Сведения о конференциях и других научных мероприятиях, диссертациях, архивных фондах и описях ограничиваются ссылками на страницы, где эти ресурсы представлены. Сведения о персональных страницах ученых-лингвистов представлены выборочно и могут

¹ Разработка программного обеспечения НИРЯЗ осуществлена А.М. Антоновым (ИНИОН РАН).

быть удалены или дополнены по желанию субъектов информации – ученых, работающих в учреждениях РАН.

На данном этапе НИРЯЗ содержит сведения о приблизительно 1,2 тыс. информационных ресурсов.

Навигация и поиск

В НИРЯЗ предусмотрены следующие виды навигации и поиска ресурсов.

1. Поиск по владельцам информационных ресурсов осуществляется в разделе *Владельцы ИР*. Предъявляется алфавитный список кратких наименований учреждений. При клике на название учреждения приводится список ресурсов, которыми оно владеет, а также другие данные об учреждении.

2. Навигация по типам ресурсов осуществляется в разделе *Типы ИР*. Предъявляется список из 15 типов ресурсов, часть которых разделены на виды. Типология ресурсов унаследована от исходной версии навигатора НИРОН (см. главу 3). При выборе типа или вида ресурсов предъявляется список наименований ресурсов, относящихся к этому типу и виду.

3. Навигация по тематике осуществляется в разделе *Тематика ЛИР*. Пользователю предъявляется тематический рубрикатор, разработанный на основе раздела 16 «Языкознание» ГРНТИ, но значительно модернизированный. Один ресурс может иметь от одной до трех рубрик.

4. Языки и группы языков. В разделе *Языки и группы языков* пользователю предъявляются алфавитные перечни конкретных языков, а также генеалогических, ареальных и других групп языков, которым посвящены ресурсы, представленные в НИРЯЗ.

5. Лексический поиск по наименованию ресурсов и владельцев. В поисковом поле нужно указать слово или фрагмент слова. В результате поиска пользователю предъявляется список ресурсов, а также владельцев, наименования которых содержат это слово или фрагмент слова.

Во всех случаях найденные наименования ресурсов содержат активные ссылки, кликнув на которые, можно получить сведения об этих ресурсах или перейти на веб-сайт, где содержится ресурс или сведения о нем.

Рубрикатор НИРЯЗ

Настоящий рубрикатор¹ разработан для Навигатора информационных ресурсов по языкознанию (НИРЯЗ). По мнению разработчиков, он отражает современную тематическую структуру научных исследований по языкознанию. Рубрикатор НИРЯЗ приведен в приложении 17.

Рубрикатор разработан на основе раздела 16 «Языкознание» Государственного рубрикатора научно-технической информации (ГРНТИ) и сохраняет его структуру, методические принципы, расположение и правила кодирования рубрик. Для совпадающих рубрик сохранены коды ГРНТИ.

К рубрикации прилагается таблица перекодировки, включающая сведения об изменениях кодировки отдельных рубрик (приложение 18). Предполагается, что вновь введенные рубрики с новыми кодами будут использоваться в НИРЯЗ. Они могут быть также использованы при применении ГРНТИ для индексирования информационных материалов, начиная с 2020 г., в других информационных организациях.

Главные изменения Рубрикатора НИРЯЗ по сравнению с ГРНТИ заключаются в следующем.

На второй уровень вынесены тематические классы «Методы лингвистических исследований» и «Уровни языка», что позволило систематизировать понятия, именующие различные направления и объекты лингвистических исследований.

Также существенно пополнена рубрика «Прикладное языкознание», в которую включены основные современные направления компьютерной лингвистики. Предыдущая версия данного раздела ГРНТИ, разработанная еще в 1970-е годы, безусловно, устарела.

При разработке рубрикатора учтены замечания, полученные от специалистов ИНИОН РАН и ВИНТИ РАН.

Данная версия рубрикатора направлена в Методический совет по классификационным системам при ВИНТИ РАН как предложения по модернизации ГРНТИ. Получен положительный ответ, подтверждающий возможность и целесообразность модернизации данного раздела ГРНТИ.

¹ Разработчики рубрикатора: Антопольский Александр Борисович, канд. филол. наук, д-р техн. наук, ИНИОН РАН; Маркарова Тамара Сергеевна, канд. филол. наук, фрилансер; Плунгян Владимир Александрович, акад. РАН, д-р филол. наук, ИРЯ РАН; Савчук Светлана Олеговна, канд. филол. наук, ИРЯ РАН.

Языковой фасет НИРЯЗ

В данной версии рубрикатора НИРЯЗ отсутствует рубрика ГРНТИ 16.41 «Языки мира». Наименования языков, а также групп языков в рубрикаторе НИРЯЗ выделены в отдельный фасет. Принцип фасетизации языков применяется также в рубрикаторе ИНИОН и в ведущих библиотечно-библиографических классификациях (УДК, ББК), где языки выделены в специальные определители.

Преимущества фасетного принципа очевидны: тематическая принадлежность ресурса никак не связана с именованием языков и языковых групп, которым посвящены ресурсы. Поэтому при описании ресурсов тематика и языки должны использоваться независимо.

Кроме того, применение фасетного принципа значительно упрощает классификацию и делает ее практическое применение более однозначным.

Однако прямое применение этого принципа в ГРНТИ является невозможным, поскольку в ГРНТИ фасеты отсутствуют. Поэтому индексировать ресурсы (так же как и документы) в отношении языков по ГРНТИ неудобно.

Существует и еще одна методическая причина выделения языков в отдельный фасет. При навигации должен быть простой способ поиска ресурсов, относящихся к конкретному языку, например, русскому. Использование ГРНТИ не дает такой возможности, потому что в рубрике 16.41 «Языки мира» на третьем уровне не находится лишь рубрика «Индоевропейские языки» и русский язык окажется на 5–6 уровне. Это, конечно, тоже неудобно.

Языковой фасет НИРЯЗ включает три алфавитных указателя:

- указатель отдельных языков;
- указатель генеалогических языковых групп;
- указатель ареальных и других языковых групп.

К «другим» группам отнесены такие не строго определенные или устаревшие понятия, как древние, мертвые, восточные языки, которые тем не менее соответствуют содержанию реальных информационных ресурсов.

Предполагается, что при описании ресурса, а также при поиске можно будет использовать как отдельные языки, так и наименования групп, имея в виду, что каждый ресурс может иметь до трех рубрик по фасету «Язык».

Во всех случаях в указатели включены только те языки и языковые группы, которым посвящены ресурсы, представленные в

НИРЯЗ по состоянию на август 2019 г. Вводить в фасет полные списки языков и языковых групп невозможно, это сделает фасет невероятно громоздким. Поэтому языковой фасет предлагается сделать динамичным и пополнять по мере появления в НИРЯЗ описаний ресурсов, посвященных новым языкам и языковым группам.

Содержание языкового фасета приведено в приложении 19.

6.5. Методика оценки качества лингвистических электронных ресурсов

Методы оценки ресурсов

Поскольку речь идет об электронных сетевых ресурсах общего доступа, то задача сводится к рассмотрению сетевых информационных систем, аккумулирующих эти ресурсы. Действительно, оценка собственно ресурсов, т.е. контента, вне представляющей его сетевой информационной системы (сайта, портала) в настоящее время вряд ли является конструктивной. Мы уже имеем не просто текст, как, например, в книге, а некоторую мультимедийную среду, которой можно управлять. Способы отображения, функциональные возможности и другие особенности современных систем позволяют говорить о формировании принципиально нового типа взаимодействия человека и информации.

Существующие методы оценки качества информационных ресурсов (сайтов) можно разделить на две категории: универсальные и экспертные, ориентированные на область знания. Универсальные методы, в том числе популярные методы библиометрии, вебометрики, альтметрики и веб-аналитики, подробно описаны во многих работах, в том числе в цитируемой выше монографии «Инфосфера общественных наук России» или в работе С.В. Бредихина и А.Ю. Кузнецова [32].

Совокупность методов и технологий измерения свойств сайта в целях его продвижения в выдачах поисковых систем называют обычно *SEO (Search Engine Optimization)*. Известны также термины: *SEO-оптимизация, поисковое продвижение, поисковая оптимизация, раскрутка сайта, продвижение сайта*.

Большинство специалистов признают, что необходимо сочетание универсальных и экспертных методов оценки. В настоящей

статье мы рассмотрим вариант экспертной оценки, ориентированной на пользователей в области языкознания.

Оценка качества сайта

Под оценкой качества сайта понимается, с одной стороны, интегральная характеристика сайта, позволяющая сравнивать его с другими, ранжировать сайты, с другой – процесс получения этой характеристики. Этот термин будет употребляться в обоих смыслах без уточнения, если выбор ясен из контекста. Оценка качества сайта как интегральная характеристика должна позволять конечному пользователю осуществлять выбор при обращении к интернет-ресурсам. Эта интегральная характеристика строится на основе оценки свойств (параметров) сайта. Во множестве имманентных свойств объекта, качество которого предполагается оценивать, в нашем случае – сайта, выделяется подмножество тех свойств, которые при потреблении (эксплуатации, применении, использовании – здесь рассматриваемых как синонимы термина «потребление») – вообще при контакте с человеком на любом этапе жизненного цикла объекта оказывают положительное или отрицательное влияние на человека (или на важные в некотором смысле для человека другие объекты).

Параметры оценки сайтов

Строго определенного набора параметров сайта, которые рассматриваются при поисковой оптимизации, не существует. Разные организации и отдельные специалисты, выполняющие такие работы, пользуются наборами, отличающимися друг от друга как по составу, так и по названиям параметров. Тем не менее между ними есть много общего, что позволяет сформировать перечень групп параметров, в той или иной степени использующихся при SEO:

- дизайн;
- контент;
- функциональность;
- навигация;
- поиск;
- экспорт информации;
- интерактивность;

- юзабилити (удобство использования);
- технические аспекты.

Однако эти параметры, в той или иной степени влияя на качество сайта, непосредственно в оценку качества не входят.

Однако многие параметры, используемые при SEO, несомненно, должны учитываться при оценке качества сайта. Действительно, с одной стороны, специалист-филолог, заказывая такой сайт или самостоятельно взявшись за его создание, должен ясно понимать, в какую коммуникативную среду он выходит, и учитывать особенности и требования этой среды. С другой – реакция поисковых систем, как уже отмечалось выше, весьма существенна при оценке качества сайта.

Следует обратить внимание на следующий важный момент. Для оценки качества сайта используются имманентно присущие ему свойства (характеристики), а не характеристики результата использования (потребления) сайта, не характеристики взаимодействия человека с сайтом. Поэтому посещаемость сайта, его позиция в поисковой выдаче не всегда прямо зависят от свойств сайта. (Аналогия: благодаря рекламе некоторое лекарство пользуется большим спросом, в то время как оно не является высококачественным.) Ситуация усугубляется тем, что количество потенциальных пользователей данного сайта в силу его специфичности может быть в принципе невелико и ожидать высокой посещаемости нельзя. При формировании оценки качества мы не должны забывать, что целями являются сравнение сайтов, их ранжирование, выработка рекомендаций для пользователей по выбору сайта. В связи с этим, помимо экспертных оценок собственно характеристик сайта, будет представлять интерес мнение людей, реально использующих сайт, а не случайно зашедших на него. По аналогии с книгой: не всякий, кто держал / просматривал / купил некоторую книгу, ее прочитал, но эти люди создают число обращений к книге. А вот мнение тех, кто внимательно прочитал книгу, может иметь для нас, для оценки книги намного большее значение. Применительно к сайту косвенным представлением такого мнения являются следующие характеристики:

- глубина просмотра;
- время, проведенное пользователем на сайте;
- показатель отказов.

Хотя эти характеристики отражают не собственные свойства сайта, их целесообразно также учитывать при оценке качества.

Характеристики оценки качества сайта

В данном разделе представлен список характеристик, которые целесообразно использовать при оценке качества сайта. Всего в список включено 47 характеристик сайта, распределенных по семи группам. За основу списка взяты характеристики, которые наиболее часто используются в SEO при работе с информационными сайтами и сайтами СМИ. Представленный список является предварительным и требует анализа на полноту свойств сайта, влияющих на его качество. Некоторые характеристики являются взаимосвязанными или функционально зависимыми. Например, в группе «Контент» есть характеристика «частота и регулярность обновления», которая взаимосвязана с характеристиками «наличие подписки на новости» и «наличие сведений о новых поступлениях». А параметр «время реакции системы на действия пользователя», в частности скорость загрузки вызванной страницы, взаимосвязан с параметром «показатель отказов»: пользователь уходит с сайта не потому, что там нет нужной информации, а потому, что не хочет тратить свое время. Требуется также проверка списка характеристик на непротиворечивость по совокупности их конкретных значений.

Дизайн:

- целостность стилевого оформления;
- читабельность основного материала;
- сбалансированность верстки страниц; не перегружены ли страницы информацией;
- сбалансированность цветов на страницах;
- наличие и уместность рекламы, не мешает ли она работе.

Контент:

- релевантность заявленной тематике;
- актуальность;
- уникальность;
- представительность (объем представленной информации);
- полнота (в какой степени покрывается заявленная тематика);
- засоренность (оценка доли ненужной информации);
- частота и регулярность обновления;
- наличие и корректность метаинформации, в частности библиографических описаний к представляемым материалам;
- логичность, полнота и удобство структуры всего контента, ее соразмерность объему контента.

Функциональность

Навигация:

- наличие и удобство использования меню;
- наличие и удобство указателей;
- наличие карты сайта;
- наличие гипертекстовых ссылок, корректность их работы;
- возможность доступа к любым представленным материалам за минимальное число кликов;
- доступ к необходимым навигационным средствам с любой страницы сайта;
- в многостраничных материалах (статьях, книгах и т.п.) простой переход на определенную страницу, в начало / конец материала, к структуре / содержанию.

Поиск:

- наличие атрибутного поиска;
- наличие полнотекстового (лексического) поиска во всем контенте, в заданном разделе;
- наличие комбинированного поиска (атрибутного и полнотекстового одновременно);
- репрезентативность списка результатов поиска и возможность их сортировки.

Экспорт информации:

- возможности экспорта и импорта информации, допустимые форматы.

Интерактивность:

- наличие обратной связи;
- наличие ссылок на социальные сети;
- наличие подписки на новости;
- возможность размещения кнопки на другом сайте;
- возможность ставить закладки, оставлять комментарии, выделять текст и т.п.

Юзабилити:

- работоспособность сайта на разных платформах;
- доступность сайта через различные браузеры;
- адаптивность сайта;
- разнообразие форм / форматов представления материала на сайте;
- время реакции системы на действия пользователя;
- наличие настроек (шрифты, масштаб, повороты и т.п.);
- ограничение доступа (регистрация, плата и т.п.);
- наличие сведений о новых поступлениях;

- наличие сопроводительной информации (цели / задачи проекта, контакты, подсказки и т.п.).

Технические аспекты:

- надежность функционирования сайта;
- безопасность обмена данными между сайтом и пользовательским устройством;
- защита данных от несанкционированного доступа.

Поведение пользователей на сайте:

- глубина просмотра;
- время, проведенное на сайте;
- показатель отказов;
- количество повторных обращений.

Сделаем некоторые пояснения для ряда характеристик.

Уникальность – доля заимствованного контента. Количество материалов, представленных на сайте, взятых из других источников.

Под *атрибутным поиском* понимается поиск информационных материалов, представленных на сайте, по их формальным признакам (метаданным). При этом предполагается, что соответствующие метаданные нормируются и существует удобный способ просмотра и выбора для поиска их значений. Невыполнение этих условий существенно снижает удобство работы пользователя и, следовательно, оценку качества.

С точки зрения оценки качества полнотекстовый поиск нельзя оценивать бинарным признаком «есть – нет», так как существует много вариантов его реализации, например:

- с помощью внутренней или внешней поисковой машины;
- поиск с учетом / без учета грамматики;
- поиск с учетом / без учета диакритики;
- поиск во всем информационном фонде или только в выбранном фрагменте.

Примерно то же относится к *комбинированному* поиску, представляющему собой совместное использование атрибутного и полнотекстового поисков, то, что часто называется расширенным поиском.

Учет всех аспектов поиска информации на сайте – специальная задача, требующая отдельной проработки. При этом эти параметры распространяются только на текстовые ресурсы, в то время как значительную и все возрастающую роль играют структурированные данные – лексикографические, корпуса, ГИСы и др.

Доступность сайта через различные браузеры: возможность обращения пользователя к сайту с помощью различных

браузеров, очевидно, является важной характеристикой для оценки качества сайта. Проверку доступности сайта через все существующие браузеры вряд ли можно назвать разумной задачей – количество их исчисляется сотнями, – однако свыше 90% обращений к сетевым системам (визитов) приходится на 10 наиболее популярных (см. табл. 9). Видимо, на доступность через эти 10, а может быть, только через первые пять браузеров следует обращать внимание при оценке качества сайта.

Таблица 9

Популярность топ-10 браузеров

№	Браузер	Доля визитов, %
1.	Google Chrome	40,17
2.	Яндекс.Браузер	20,61
3.	Safari	10,32
4.	Opera	5,39
5.	Firefox	4,46
6.	Android Browser	2,91
7.	Samsung Internet	2,61
8.	Internet Explorer	1,98
9.	Edge	1,57
10.	MIUI browser	1,36

Адаптивность сайта – способность сайта корректно отображаться на экранах всех размеров и на всех видах устройств. При этом сайт подстраивается под определенный экран в режиме реального времени, т.е. динамически. Основными плюсами адаптивного сайта являются его универсальность и простота использования, а также отсутствие надобности в отдельной версии сайта. Сайт сам будет принимать приемлемый вид, независимо от того, какое устройство его просматривает – ноутбук, планшет, смартфон или телевизор, – все будет отображаться корректно и удобно для пользователя.

Существуют два основных варианта адаптивности:

- мобильная версия сайта – это дублирующая версия основного сайта, для которой применяется специальная верстка;
- адаптивный сайт – это не отдельная версия ресурса, это и есть основной сайт, и он автоматически подстраивается под разрешение устройства, с которого происходит вход.

Время реакции системы на действия пользователя – это характеристика показывает время ожидания пользователем результа-

тов выполненных операций, прежде всего, открытия новой страницы сайта, получения результатов поиска, переходов в системе навигации. Люди совершенно не готовы ждать лишние пару секунд, не зная точно, что это того стоит. Легче закрыть вкладку с медленно загружающимся сайтом и перейти на другой ресурс. Согласно исследованиям Microsoft Bing, задержка всего лишь в несколько секунд приводит к ухудшению впечатления пользователя на 3,8%, снижению прибыли на 4,3% и количества кликов на 4,4%. Учитывать в количественном выражении реакцию на многие операции и затем строить комплексную оценку формально было бы более правильно, чем получать качественную оценку эксперта. Однако сложность и трудоемкость количественного подхода, а также незначительное повышение точности делают целесообразным выбор мнения эксперта.

Группа *Поведение пользователей на сайте* является важным показателем, который может давать представление о привлекательности сайта. Эта группа включает в себя три характеристики.

Показатель отказов (BR – bounce rate) – средний процент людей, которые перешли на сайт, но сразу (например, в Яндексe – в течение 15 сек.) покинули его. Google считает отказом любое посещение, при котором была просмотрена только одна страница сайта. Даже если пользователь читал текст несколько минут или даже часов, а потом покинул сайт, сеанс будет считаться нулевым. В нашем случае целесообразно принять подход Яндекса.

Глубина просмотра (PPV – pages per visit) – количество страниц сайта, которые посетил пользователь за один сеанс.

Время пребывания на сайте (время на сайте / длительность визита; TSS – time spent on site) – среднее время, которое посетители проводят на сайте. Учитывается время с начала визита до последнего активного действия.

Безусловно, характеристики этой группы не очень точные. Так, глубина просмотров зависит от способа представления информации на сайте: один и тот же документ можно представить одной веб-страницей, а можно – многими. Длительность визита не в полной мере характеризует время работы с данным сайтом: например, пользователь не покинул сайт, а ушел на несколько часов от компьютера, не закрыв сессию. Однако заменить эти важные характеристики нечем, но в среднем и совместно они должны давать приемлемый результат.

Методика оценки качества сайта

Сайт описывается набором характеристик. Каждая характеристика сайта может быть количественной или качественной.

Для результатов, допускающих количественное измерение, используют количественные показатели. Значения таких показателей выражаются в некоторой метрической шкале в виде некоторого действительного числа, имеющего определенный физический или экономический смысл.

Большинство характеристик и результатов строгому количественному измерению не поддаются. Для их оценивания используют качественные показатели. Качественные показатели измеряют с помощью экспертных оценок, т.е. субъективно, путем наблюдения за процессом и результатами. Значения их выражаются значениями на номинальной или порядковой шкале. Более того, даже в тех случаях, когда характеристика по сути своей является количественной, ее можно оценивать как качественную и получать ее экспертную оценку на порядковой шкале. Например, оценку информационного фонда можно проводить в соответствии с порядковой шкалой: 2 – малый, 3 – средний, 4 – большой, 5 – очень большой, – получая тем самым субъективное мнение эксперта. Представление этой характеристики как количественной приведет к очень большим сложностям. На какой единице измерения остановиться: байт, файл, книга, произведение, страница?.. Можно ли ее применить ко всем интересующим сайтам?

Поскольку подавляющее число характеристик сайта являются качественными и для их получения требуются специалисты-эксперты, то основным становится метод экспертных оценок, т.е. метод организации работы со специалистами-экспертами и обработки их мнений. Такие методы к настоящему времени достаточно хорошо изучены, отработаны и даже стандартизованы [36].

Метод экспертных оценок – это метод организации работы со специалистами-экспертами и обработки мнений экспертов. Сущность метода экспертных оценок заключается в том, что в основу принятого решения закладывается мнение специалиста или коллектива специалистов, основанное на их знаниях и практическом профессиональном опыте. Экспертной может называться только та оценка, которая основана на правилах объективности и честности.

Предполагаются следующие этапы работы по экспертному оцениванию.

- Подготовительный этап, включающий:

- назначение модератора экспертизы;
- уточнение целей и задач экспертизы;
- отбор сайтов, подлежащих оценке;
- подготовка анкет для экспертизы каждого сайта.
 - Формирование группы экспертов.
 - Проведение самой экспертизы.
 - Проведение полного анализа результатов – обработка экспертных оценок.
 - Подготовка и составление отчета с результатами экспертного оценивания.

Применительно к нашей задаче экспертизу целесообразно проводить в форме индивидуальных оценок, т.е. использовать мнения независимых друг от друга отдельных экспертов. Количество независимых экспертов – 5–10 человек. Экспертиза должна проводиться в течение определенного времени одновременно всеми экспертами.

Эксперты должны быть профессиональными квалифицированными филологами, имеющими достаточный опыт работы с компьютерами и сетевыми ресурсами. Эксперт не должен быть заинтересован в каком-то определенном или конкретном результате оценки каждого сайта.

Параметры, по которым оценивается сайт, не являются неизменными. Они зависят от:

- изменения содержания сайта;
- изменения объема информационного фонда;
- реорганизации структуры информационного фонда и / или сайта;
- проведения кампаний по продвижению сайта;
- сезонных изменений трафика;
- привязки к циклам образования;
- привязки к периодам отдыха (каникулы, праздники, отпуска и т.п.).

Поэтому необходимо получать среднюю, устойчивую оценку за год.

Поскольку оценка сайтов необходима для их сравнения, то аналогичные регулярные процедуры следует производить и для других сайтов. Однако ситуация в этом случае несколько сложнее. Действительно, в результате сравнения в определенный момент сайт-1 оказался лучше сайта-2; через год ситуация может радикально измениться – например, если сайт-1 не изменялся, а сайт-2 развивался.

Учитывая «сезонность» в работе сайтов, т.е. колебания показателей, по которым оценивается сайт, в течение года, а также изменения, происходящие на самом сайте, представляется целесообразным производить проверки регулярно – 3–4 раза в течение года.

Экспертиза проводится методом анкетирования. Анкета составляется для каждого оцениваемого сайта и включает в себя:

- сведения об эксперте и времени проведения экспертизы;
- сведения об оцениваемом сайте (название, электронный адрес);
- перечень вопросов, на которые должен ответить эксперт, – перечень характеристик сайта, по которым он должен дать оценку;
- дополнительные сведения о сайте, которые эксперт считает необходимым сообщить;
- общий вывод о качестве сайта.

Перечень характеристик сайта, по которым эксперт будет давать оценку, должен содержать для каждой характеристики ее наименование и возможные варианты оценок. Оценка заключается в выборе одного из вариантов. Исключение составляют количественные характеристики, определяемые по данным систем Google- / Яндекс-аналитика. Эти оценки получаются и проставляются в анкетах на этапе их подготовки.

Таким образом, экспертиза проводится по методу непосредственной оценки. В этом случае диапазон изменения характеристик объекта разбивается на отдельные интервалы, каждому из которых приписывается определенная оценка (балл), например, от 0 до 10. Именно поэтому метод непосредственной оценки иногда именуют также балльным методом.

Для анализа результатов применяются различные методы математической статистики. Причем они могут комбинироваться и варьироваться в зависимости от типа задачи и необходимого результата.

В упрощенном варианте сайт оценивается по формуле:

$$K = \alpha_d D + \alpha_c C + \alpha_f F + \alpha_u U + \alpha_n N + \alpha_a A,$$

где K – оценка качества сайта,

D, C, F, U, N, A – частные оценки сайта по соответствующей группе характеристик,

$\alpha_d, \alpha_c, \alpha_f, \alpha_u, \alpha_n, \alpha_a$ – весовые коэффициенты соответствующих групп.

Частная оценка качества сайта – оценка по определенной группе характеристик. Она определяется по формуле:

$$X = \sum_1^m \beta_i x_i ,$$

где $X = D, C, F, U, N, A$

m – количество характеристик в группе,

x_i – оценка i -ой характеристики в группе,

β_i – весовой коэффициент i -ой характеристики в группе.

Примеры оценки сайтов

Для понимания правильности восприятия задачи экспертами и представления о получаемых результатах трем специалистам-филологам (д-р филол. наук и два канд. филол. наук, МГУ), работающим на стыке языкознания и литературоведения, было предложено оценить несколько хорошо знакомых им сайтов. Выбор сайтов оставался на их усмотрение. Эти специалисты в интересах своей профессиональной деятельности активно используют средства вычислительной техники и постоянно обращаются к сетевым информационным ресурсам. Оценка должна была производиться в удобной им форме по рубрикам параметров оценки качества, рассмотренным выше. При этом им было разрешено достаточно произвольно использовать эти рубрики, так как важно было выявить их предпочтения при выборе параметров оценки. В результате ими был подготовлен формуляр, по которому и производилась оценка качества.

Параметры оценки:

- дизайн (design);
- контент (content);
- функциональность:
- навигация (navigation);
- поиск (search);
- юзабилити (usability):
- настольный компьютер (desktop version);
- мобильные устройства (responsive version);
- безопасность (https).

Надо отметить, что эти эксперты не стали использовать целый ряд характеристик сайта, которые достаточно формальны и не требуют специальной оценки филологов. Так, ими полностью не использовались рубрики «Интерактивность» и «Поведение пользователей на сайте», а также такие параметры, как доступность сайта через различные браузеры, наличие карты сайта и др.

Кроме того, они рекомендовали получать оценку каждого сайта не от одного, а от двух экспертов: филолога и IT-специалиста, например специалиста по SEO. При этом опросники (перечни оцениваемых характеристик сайта) должны пересекаться, в частности по дизайну. В нашем случае, учитывая целесообразность применения метода экспертных оценок, речь должна идти о двух соответствующих группах экспертов. Причем эти группы должны быть независимы. Результаты в виде анкет сайтов представлены в подробном отчете по разработке данной методики [153].

Еще одной важной работой, проделанной теми же специалистами-филологами, был ответ на вопрос о том, чем они руководствуются при выборе того или иного сайта. Сводный ответ представлен ниже.

Наличие и корректность метайнформации – очень важны, поскольку в противном случае крайне затруднена перепроверка. Без корректной метайнформации любой текст – это просто цитата «неизвестно откуда», чтобы ни было написано рядом с текстом.

Формат материалов – важен в случае копирования всего файла (предпочтительно pdf и / или djvu, сохраняющие исходные печатные страницы), не важен – в случае поиска информации (при условии, что можно скопировать найденное или выделенное), в этом случае главное – простота копирования найденного. При работе со смартфона – очень важна респонсивность / адаптивность; она хуже достигается при pdf-представлении, чем при html-представлении; иногда удобны другие image-форматы (jpg, png), встроены в html-страницу.

Возможность копирования – очень важна. В противном случае крайне замедляется работа с информацией, а при работе со смартфоном становится невозможной (скопировать нельзя, а переписать некуда).

Частота и регулярность обновления – зависит от типа информации.

Наличие обратной связи – важно в случае, если нужно связаться с теми, кто ведет сайт. Обычно это не нужно, но когда не-

обходимость возникает, а на сайте информации нет, – это вызывает раздражение.

Время реакции системы – если речь идет о работе с корпусом или библиотекой, когда делаются подряд десятки запросов, медленное действие, конечно, очень мешает.

Наличие настроек (шрифты, масштаб, повороты и т.п.) – на ноутбуке или ПК очень помогает, иначе приходится пользоваться возможностями браузера (увеличение и пр.), а это часто неудобно и приводит к непредсказуемым последствиям (например, увеличение шрифта может привести к потере навигационной панели). На смартфоне, наоборот, удобнее адаптивность: специально настраивать что-то на экране телефона трудно; но должна быть возможность дополнительного увеличения с помощью стандартных средств (сейчас это простая растяжка изображения пальцами).

Наличие и навязчивость рекламы – совершенно не беспокоит. Понятно, что людям нужно на что-то содержать ресурс. Кроме того, можно установить блокировщик рекламы.

Ограничение доступа (регистрация, плата и т.п.) – если ресурс нужен постоянно, на нем регистрируются; если нет, ищут информацию в другом месте; платными ресурсами обычно не пользуются, но если очень нужно, платят.

Наличие сопроводительной информации (цели / задачи проекта, его спонсоры, исполнители, контактная информация и т.п.) – очень полезно при первом знакомстве с ресурсом. Иногда для того, чтобы решить, остаться здесь или поискать в другом месте, сопроводительная информация принципиально важна (в случае если нужна профессиональная, проверенная информация, должно быть понятно, кто ее в принципе может подготовить, а кто нет).

Целесообразность учета. Ответ на вопрос о целесообразности учета при оценке сайта по тематическим признакам сводился примерно к следующему. Во-первых, для этих целей достаточно каталогов, а во-вторых – общая оценка по этим признакам неизбежно будет достаточно груба, а точная – субъективна.

Проведение описанного эксперимента позволяет сформулировать методику проведения экспертизы ресурсов по языкознанию для отбора ресурсов для включения в информационную систему Центра лингвистических ресурсов. Однако окончательную форму данная методика приобретет после появления утвержденного документа (например, технического задания или технического проекта), определяющего функциональность и этапы создания Центра лингвистических ресурсов.

6.6. Онтология поисковых терминов по лингвистике

Участие в создании онтологии по комплексному гранту

В рамках упомянутого комплексного гранта РФФИ «Разработка фундаментальных основ, методов и средств обеспечения конвергенции естественно-научных и социогуманитарных ресурсов как составляющих Единого цифрового пространства знаний» одной из центральных задач является разработка понятийного аппарата конвергенции, который, очевидно, должен иметь форму некоторой онтологии.

О том, что онтология научного знания должна быть элементом ядра ЕРЭПЗ, автор неоднократно писал, в том числе и в главе 5 данной книги. Однако структуру и механизмы формирования подобной онтологии еще предстоит определить.

В настоящей книге автор не предполагал делать подробный анализ понятия онтологии и модификации этого понятия в российской и мировой науке. Укажем, что некоторый обзор трактовок этого понятия и соответствующей литературы сделан в монографии [22].

Там же представлено описание результатов исследований по гранту РФФИ «Интерактивная система создания и поддержки онтологии научного знания на базе динамического комплекса терминологических словарей», в котором принимал участие автор. Результаты исследований, выполненных в рамках этого гранта и не вошедших в цитированную выше монографию, представлены в публикациях [76; 93].

Эти результаты, в том числе созданная в результате исследований по гранту информационная система «Терминобаза» [104], стали одним из источников исследований в рамках комплексного проекта по конвергенции.

Участие в комплексном проекте по конвергенции автора, как и ИНИОН РАН, кроме собственных исследований, определялось следующей формулировкой в заявке: «Участие в комплексном проекте по разработке онтологии научного знания на основе имеющихся в ИНИОН РАН информационных ресурсов по языкознанию, тезаурусов, рубрикаторов и систем метаданных и включение этих ресурсов в единое цифровое информационное пространство РАН».

В рамках выполнения этого пункта заявки соисполнителям по гранту в ФИЦ «Информатика и управление» были переданы следующие материалы:

- тезаурус по языкознанию ИНИОН РАН;
- словарь ключевых слов базы данных по языкознанию ИНИОН РАН;
- фрагмент базы данных по языкознанию ИНИОН РАН.

Эти данные были загружены в платформу ЛибМета, которая играет роль метапоисковой машины. ЛибМета позволяет модифицировать запросы к различным базам данных, в том числе расширять запрос при помощи тезаурусов, используя ассоциативные, иерархические и синонимические связи, а также учитывать морфологию терминов запроса. Технология ЛибМета и ее применение для поиска в междисциплинарном информационном пространстве описаны в ряде публикаций [26; 115; 116].

Онтология поисковых терминов по лингвистике (ОПТЕЛ)

Кроме описанного выше решения об участии в создании общей онтологии по комплексному гранту, было принято решение о разработке лингвистического ресурса, специально предназначенного для метапоиска в рамках концепции Центра лингвистических ресурсов.

Этот ресурс получил название *Онтология поисковых терминов по лингвистике (ОПТЕЛ)*.

ОПТЕЛ предполагается формировать как сводную базу данных терминов информационно-поисковых языков разных типов, используемых в разнородных лингвистических информационных ресурсах. Среди них классификационные языки (ГРНТИ, УДК, ББК, Рубрикатор ИНИОН), дескрипторные (тезаурус ИНИОН, словари ключевых слов), теги для разметки корпусов, входящих в состав Национального корпуса русского языка, и других корпусов, лексика признаков и значений признаков лексикографических БД, реестров языков и, возможно, иных ЛИР.

ОПТЕЛ на первом этапе может рассматриваться как лексический указатель к разнородным ЛИР, отраженным в Навигаторе НИРЯЗ.

В перспективе, учитывая богатство представленной лексики, а также семантическую информацию, заложенную в информаци-

онно-поисковых языках, ОПТЕЛ может стать одним из источников полноценной онтологии по лингвистике.

В настоящее время работа по созданию ОПТЕЛ только начата, в частности, подготовлено Техническое задание на ОПТЕЛ. Оно представлено в приложении 20.

Перспективы развития ОПТЕЛ, как и других частных разработок по формированию онтологии научного знания, будут определяться как общим ходом работ по проектированию ЕРЭПЗ, так и его компонентов, в том числе Центра ЛИР.

Заключение

Мы, конечно, не осветили все важные тенденции и факторы, влияющие на развитие научных коммуникаций и определяющие образ будущего. За пределами нашего рассмотрения остались такие важные вопросы, как:

- динамика цифровизации научных коммуникаций;
- дезагрегация и персонализация коммуникативных функций в единой цифровой среде;
- экономическая модель коммуникаций в условиях открытого доступа;
- развитие правового статуса информационных ресурсов и системы копирайта в цифровой среде;
- развитие семантических технологий и искусственного интеллекта.

Но даже на основе рассмотренных вопросов можно утверждать, что анализ существующего положения дел в области научных коммуникаций и тенденций их развития не дает достаточно возможностей, чтобы с надежностью предсказать будущее научных коммуникаций и научно-информационных продуктов и систем. Слишком многое зависит от политики, проводимой руководством российской науки. Это касается и движения в сторону открытой науки, и концепций пространства знаний, экономических и правовых проблем коммуникации, а также организационной модели системы научной информации в России.

Роль нашего сообщества должна заключаться в выработке образа будущего, приемлемого для различных участников научной коммуникации. Как показывает опыт борьбы вокруг открытого доступа или проблем копирайта, поиск баланса интересов представляет собой совсем не простую задачу

В качестве заключения хотел бы поддержать проф. В.М. Московкина [59]: переход российских научных коммуникаций к от-

крытой науке и их оптимизация требуют осмысленной, системной и целенаправленной политики в этой области российских государственных органов и фондов, прежде всего Министерства науки и высшего образования, РАН, РФФИ и РНФ.

Эта политика должна учитывать происходящие изменения научного информационного пространства, намечать контуры системы научной информации в цифровой среде и служить ориентиром для создателей информационных ресурсов и сервисов.

К сожалению, отсутствие такой политики означает, что многие финансовые ресурсы, которые сейчас направляются на информационное обеспечение науки, будут затрачены неэффективно.

Для исправления этого положения требуется активизация научного и информационного сообщества для формирования адекватного образа будущего научных коммуникаций, который может быть предложен правительству.

Очевидно, что для автора, который работает в ИНИОН РАН, важнейшее значение имеют перспективы этого учреждения, которое долгое время было одним из лидеров отечественной научной информации, да и сейчас обладает огромным потенциалом в научно-информационной деятельности.

Какую же роль в новой модели организации научной информации может играть ИНИОН?

Во-первых, роль координатора формирования Единого электронного пространства знаний в области общественных и гуманитарных наук. Для этого необходимы проведение мониторинга научной инфосферы, оценка и отбор наиболее качественных российских информационных ресурсов, определение рациональных способов их интеграции в Единое электронное пространство знаний. Ключевое направление всей этой работы – создание онтологии научного знания и базы знаний, отражающих актуальную, достоверную и оригинальную информацию, вероятно, в форме, близкой к энциклопедической.

Хотя для этой работы имеются серьезные заделы (в частности, опыт создания тезаурусов по социально-гуманитарным дисциплинам), в целом для ИНИОН это станет существенно новым направлением. Не следует бояться того, что, взявшись за решение этой задачи, ИНИОН непременно пожертвует основными направлениями своей текущей деятельности. На деле, для организации этой работы нужно сформировать внутри института одно сильное подразделение с ограниченной штатной численностью (до 20 человек), которое станет ядром, вокруг которого сформируется

внешняя сеть экспертов по профильным социально-гуманитарным дисциплинам.

Во-вторых, созданная в ИНИОН база данных по общественным наукам (АИСОН) может стать основой для формирования сводной библиографической базы данных по гуманитарным и общественным наукам научных учреждений РАН или Минобрнауки в целом. Такая база сможет освободить ученых и преподавателей от трудоемкой и неэффективной деятельности по учету публикаций и других результатов научной деятельности и отражения их в отчетности, а также позволит решить многие инфометрические задачи.

В-третьих, важно сохранить существующие реферативные службы института, но при этом актуализировать их задачи, сделав упор на аналитику и извлечение нового знания из потока научной информации.

Понятно, что эти задачи требуют, прежде всего, привлечения квалифицированных специалистов по всем областям социального и гуманитарного знания. По сути, это то, чем ИНИОН уже обладает. Мнение о том, что обновление научно-информационной миссии института потребует замены работающих в ИНИОН специалистов на информационных технологов, является глубоко ошибочным. Напротив, опыт цифровой гуманитаристики показывает, что основная роль при создании информационных ресурсов и систем должна принадлежать именно отраслевым специалистам, а функция IT-специалистов – адаптация или разработка программно-технических средств, обеспечивающих выполнение поставленных задач.

Весьма перспективное направление – использование возможностей искусственного интеллекта и больших данных для социально-гуманитарных исследований. Но делать это в рамках одной научной организации едва ли разумно. Тут нужна коллаборация с участием IT-специалистов, информационщиков, математиков, специалистов по общественным и гуманитарным наукам. Представители ИНИОН, несомненно, могли бы играть видную роль в такой коллаборации. Фактически уже сейчас мы начинаем осваивать такую модель сотрудничества, разрабатывая совместно с учеными Института русского языка им. В.В. Виноградова РАН концепцию создания общероссийского Центра лингвистических ресурсов.

Предлагаемая читателям книга поднимает актуальные вопросы организации научно-информационной деятельности в России. Автор надеется, что эта книга поможет в осознании этих проблем как научному сообществу, так и руководителям российской науки.

Литература и электронные ресурсы

1. Акт передачи произведений в общественное достояние. – URL: <http://opencitations.wordpress.com/2013/01/03/open-letter-to-publishers>
2. Амстердамский призыв для действий по открытой науке. – Режим доступа: <https://www.government.nl/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science>
3. Антопольский А., Каленов Н., Серебряков В. Единая система информационного обслуживания науки ФАНО (Проект концепции) // Информационные ресурсы России. – 2016. – № 5. – С. 2–5.
4. Антопольский А.Б. О разработке Единого российского электронного пространства знаний // Научное издание международного уровня – 2019: стратегия и тактика управления и развития: материалы VIII Международной научно-практической конференции / отв. ред.: О.В. Кириллова. – 2019. – С. 17–29.
5. Антопольский А.Б. Инвентаризация информационных ресурсов академического сектора общественных наук // Историческая информатика. – 2017. – № 3. – С. 20–42. – DOI: 10.7256/2585–7797.2017.3.24014. – Режим доступа: https://nbpublish.com/library_read_article.php?id=24014
6. Антопольский А.Б. Единое российское электронное пространство знаний как ключевая задача информатики в обозримом будущем // Информационное общество: образование, наука, культура и технологии будущего: 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. – Выпуск 3. – С. 9–23. – Режим доступа: <https://doi.org/10.17586/2587-8557-2019-3-9-23>
7. Антопольский А.Б. О возможности и целесообразности воссоздания национальной системы научно-технической информации в России // Информация и инновации: Оценки, тенденции, перспективы. – 2017. – № 1/2. – С. 15–21.
8. Антопольский А.Б. О создании центра лингвистических ресурсов РАН // Известия Российской академии наук. Серия литературы и языка. – 2019. – Т. 78, № 4. – С. 5–12. – DOI: 10.31857/S241377150006107-. – Режим доступа: <http://ras.jes.su/sliya/s241377150006107-0-1>

9. Антопольский А.Б. Определение границ при проведении мониторинга информационных ресурсов социально-гуманитарных наук // Информационные ресурсы России. – 2017. – № 3 – С. 6–10.
10. Антопольский А.Б. Перспективы развития научных коммуникаций // Культура: теория и практика: Электронный научный журнал. – 2019. – Выпуск 2 (30), март-апрель. – Режим доступа: <http://theoryofculture.ru/issues/104/>
11. Антопольский А.Б. Алгоритмы и процедуры определения уровня развития интеллектуальных ресурсов и научной и образовательной организации // Научно-техническая информация. Сер. 1: Организация и методика информационной работы. – 2015. – № 1. – С. 14–24.
12. Антопольский А.Б. Будущее научных коммуникаций и научной информации // Информация и инновации. – 2019. – Т. 14, № 1 – С. 7–17. – DOI: 10.31432/1994-2443-2019-14-1-7-17
13. Антопольский А.Б. Информационные ресурсы академических организаций социогуманитарного профиля: опыт мониторинга и классификации // Международная конференция, посвященная 65-летию ВИНИТИ РАН, «Информация в современном мире», Москва, 25–26 окт., 2017: Материалы конференции. – 2017. – С. 7–12. – русский.
14. Антопольский А.Б. Информационные ресурсы общественных наук. Опыт организации мониторинга // Библиосфера. – 2017. – № 3. – С. 78–84.
15. Антопольский А.Б. О представлении информационно-компьютерных наук в различных классификационных системах: Текст доклада на 28 заседании совместного семинара ИПИ РАН и ИНИОН РАН «Методологические проблемы наук об информации» (26 мая 2016 г., г. Москва, ИНИОН РАН). – Режим доступа: http://inion.ru/index.php?page_id=605
16. Антопольский А.Б. О путях оптимизации академических социогуманитарных информационных ресурсов // Научно-техническая информация. Сер. 1: Организация и методика информационной работы. – 2018. – № 8. – С. 21–27.
17. Антопольский А.Б., Антонов А.М. О навигаторе информационных ресурсов по общественным наукам // Информация и инновации. – 2018. – Т. 13, № 2. – С. 18–24. – DOI:10.31432/1994-2443-2018-13-2-18-24
18. Антопольский А.Б., Белоозеров В.Н., Маркарова Т.С. О разработке онтологии на основе классификаций и терминологических словарей // Информационные ресурсы России. – 2017. – № 5. – С. 2–7.
19. Антопольский А.Б., Вигурский К.В.. Экспертная оценка качества сетевых ресурсов по языкознанию: Препринт. – Режим доступа: http://inion.ru/site/assets/files/1206/ekspertiza_saitov_preprint.pdf
20. Антопольский А.Б., Ефременко Д.В. К вопросу о едином электронном пространстве знаний // Вестник Российской академии наук. – 2018. – Т. 88, № 2. – С. 163–170. – DOI: 10.7868/S086958731802007 X. – Режим доступа: <https://doi.org/10.1134/S1019331618010070>
21. Антопольский А.Б., Ефременко Д.В. О создании современной цифровой инфраструктуры для хранения и анализа научно-технической информации // Научно-техническая информация. Сер. 1: Организация и методика информационной работы. – 2019. – № 6. – С. 8–17.

22. Антопольский, А.Б., Ефременко Д.В. Инфосфера общественных наук России: монография / под ред. В.А. Цветковой. – М.; Берлин: Директ-Медиа, 2017. – 676 с. – DOI: 10.23681/468227
23. Архивы РАН. – Режим доступа: <http://www.aran.ru/?q=ru/welcome>
24. Ассоциация научных редакторов и издателей (АНРИ). – Режим доступа: <http://rasep.ru/>
25. Ассоциированные региональные библиотечные консорциумы (АРБИКОН). – Режим доступа: <https://arbicon.ru>
26. Атаева О.М., Серебряков В.А., Тучкова Н.П. Расширение предметной области информационного запроса на основе онтологии знаний цифровой библиотеки LibMeta // Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23–28 сентября 2019 г., г. Новороссийск). – М.: ИПМ им. М.В. Келдыша, 2019. – С. 63–75.
27. Базы данных STN International. – Режим доступа: http://www.nioch.nsc.ru/sibstn/onlin_db.htm
28. Беляева Л.Н. Сетевые лингвистические ресурсы в подготовке филолога // Вестник ПНИПУ. Проблемы языкознания и педагогики. – 2015. – № 4 (14). – Режим доступа: <https://cyberleninka.ru/article/n/setevye-lingvisticheskie-resursy-v-podgotovke-filologa>
29. Берлинская декларация об открытом доступе к научным и гуманитарным знаниям. – Режим доступа: <https://openaccess.mpg.de/67987/BerlinDeclarati>
30. Богатов В.В., Сыроежкина Д.С. Коллаборации научных организаций как элемент инфраструктуры науки // Управление наукой и наукометрия. – 2016. – № 4. – Режим доступа: <https://cyberleninka.ru/article/n/kollaboratsii-nauchnyh-organizatsiy-kak-element-infrastruktury-nauki>
31. Бородин Л.И., Гарскова И.М. Историческая информатика: перезагрузка? // Вестник Пермского университета. Серия «История». – 2011. – Вып. 2 (16). – С. 5–11.
32. Бредихин С.В., Кузнецов А.Ю. Методы библиометрии и рынок электронной научной периодики. – Новосибирск: ИВМиМГ СО РАН: НЭИКО, 2012. – 256 с.
33. Вершинин А.П. Единое российское электронное пространство знаний: вопросы права // Университетская книга. – 2016. – Декабрь. – Режим доступа: <http://www.unkniga.ru/biblioteki/bibdelo/6630-edinoe-rossiyskoe-elektronnoe-prostranstvo-znaniy-voprosy-prava.html>
34. Встреча с Юрием Осиповым и Сергеем Кравцом // Официальный сайт Президента России. – 2017. – 20 декабря. – Режим доступа: <http://kremlin.ru/events/president/news/56450>
35. Гагская декларация по открытию знаний в цифровую эпоху / Европейская ассоциация научных библиотек. – Режим доступа: <https://www.rd-alliance.org/hague-declaration-knowledge-discovery-digital-age.html>
36. ГОСТ Р 56713–2015 (ISO/IEC 25010:2011) Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модели качества систем и программного обеспечения. – Режим доступа: <http://docs.cntd.ru/document/1200127275>
37. Государственный каталог музейного фонда РФ. – Режим доступа: <https://goskatalog.ru/portal>

38. Декларация Будапештской конференции по открытому доступу. – Режим доступа: <http://legacy.earlham.edu/~peters/fos/boaifaqr.htm>
39. Диссертационные советы – Режим доступа: https://vak.minobrnauki.gov.ru/dc#tab=_tab:dc~
40. Долгин А.Б. Манифест новой экономики. Вторая невидимая рука рынка. – М.: АСТ, 2010. – 244 с.
41. Долгин А.Б. Экономика символического обмена. – М.: Прагматика культуры, 2007. – 640 с.
42. Егоров С.В., Захарова С.А. Краудсорсинг в науке // Социологический альманах. – 2015. – № 6. – Режим доступа: <https://cyberleninka.ru/article/n/kraudsorsing-v-nauke>
43. Единая государственная информационная система учета результатов научно-исследовательских, опытно-конструкторских и технологических работ гражданского назначения (ЕГИСУ НИОКТР). – Режим доступа: <https://rosrid.ru>
44. Единая цифровая платформа научного и научно-технического взаимодействия, организации и проведения совместных исследований в удаленном доступе, в том числе с участием зарубежных ученых (ЦПСИ). – Режим доступа: https://minobrnauki.gov.ru/common/upload/library/2019/07/20190329_Sovet_20190329_TSPSI.pdf
45. Информационно-библиотечный совет РАН. – Режим доступа: https://www.benran.ru/IBS_RAN/doc.html
46. Информационно-поисковая система «Русская словесность». – URL: <http://www.pushkinskiydom.ru/Default.aspx?tabid=10526>
47. Информационные языки в XXI веке: Библиографический указатель / сост. А.Б. Антопольский, Т.С. Маркарова. – Режим доступа: <http://inion.ru/about/news/informatsionnye-iazyki-v-21-veke-bibliograficheskii-ukazatel/>
48. Информация о конференциях, симпозиумах, международных мероприятиях, проведенных Институтом российской истории РАН. – Режим доступа: <http://iiran.ru/?q=conf>
49. Инфосфера ОПК: теория, история, перспективы: Коллективная монография / Антопольский А.Б. и др. – М.: МАКС Пресс, 2016, – 328 с.
50. Калёнов Н.Е. Задачи и функции библиотек РАН в современных условиях // Информатика и ее применения. – 2012. – Т. 6, № 2. – С. 51–58.
51. Каталог «Наука в Рунете». – Режим доступа: <https://elementy.ru/catalog>
52. Книжный рынок России: состояние, тенденции и перспективы развития: Отраслевой доклад ФАПМК. – 2018. – Режим доступа: <http://www.fapmc.ru/rospechat/activities/reports/2018/pechat2>
53. Концепция создания Единой цифровой платформы науки и высшего образования Минобрнауки России. – Режим доступа: https://minobrnauki.gov.ru/common/upload/library/2019/07/20190705_Kontseptsiya_ETSP_1.4.9.pdf
54. Концепция цифровой автоматизированной системы предоставления сервисов научной инфраструктуры коллективного пользования (АС УСНИКП). – Режим доступа: https://minobrnauki.gov.ru/common/upload/library/2019/07/Kontseptsiya_AS_USNIKP_11_03_2019_Odobren.pdf
55. Маркарова Т.С. Модель развития отраслевой академической библиотеки в современной информационной среде // Проблемы современного образования. –

2011. – № 2. – С. 123–129. – Режим доступа: http://www.pmedu.ru/res/2011_2_10.pdf
56. Московкин В., Шерстюкова Е. Будущее научных публикаций – за открытым доступом? Сделать научное знание доступным всем // Троицкий вариант. Бытие науки. – 2019. – 17.01 – Режим доступа: <https://trv-science.ru/2019/01/17/budushhee-za-otkrytym-dostupom>
57. Московкин В. Десять принципов Плана S // Троицкий вариант. Бытие науки. – 2018. – 10.11. – № 267. – С. 5. – Режим доступа: <https://trv-science.ru/2018/11/20/10-principov-plana-s-eu/>
58. Московкин В. Революционный шаг Евросоюза: размещение в открытом доступе всех результатов научных исследований // Троицкий вариант. Бытие науки. – 2017. – 18.07, № 233. – С. 4. – Режим доступа: <https://trv-science.ru/2017/07/18/revolyucionnyj-shag-eu>
59. Московкин В. Руководство по выполнению Плана S Евросоюза. Вызов для России // Троицкий вариант. Бытие науки. – 2019. – 26.02, № 273. – С. 14. – Режим доступа: <https://trv-science.ru/2019/02/26/rukovodstvo-po-vypolneniyu-plana-s-eu>
60. Навигатор информационных ресурсов по общественным наукам. – Режим доступа: <http://http://niron.inion.ru>
61. Навигатор информационных ресурсов по управлению наукой. – Режим доступа: <http://naukoved.alexo.beget.tech>
62. Навигатор информационных ресурсов по языкознанию. – Режим доступа: <http://niryaz2.alexo.beget.tech>
63. Насколько дорога память: ученые обсудили современное положение академического архивного дела // Научная Россия. – 2017. – 17 декабря. – Режим доступа: <https://scientificrussia.ru/articles/arhiv-ran-kruglyj-stol-05-12-2017>
64. Научная электронная библиотека eLIBRARY.ru. – Режим доступа: <https://elibrary.ru>
65. Научная электронная библиотека «КиберЛенинка». – Режим доступа: <https://cyberleninka.ru/>
66. Научное наследие России. – Режим доступа: <http://e-heritage.ru/about.html>
67. Научный совет по музеям Сибирского отделения РАН. – Режим доступа: http://www.history.nsc.ru/structure/council/council_museum.htm
68. Национальная электронная библиотека. – Режим доступа: <https://нэб.рф>
69. Национальный портал открытого образования. – Режим доступа: <http://npood.ru/about>
70. Национальный агрегатор открытых репозиторий (НОРА). – Режим доступа: <https://openrepository.ru>
71. Национальный информационно-библиотечный центр ЛИБНЕТ. – Режим доступа: <http://www.nilc.ru>
72. Национальный корпус русского языка. – Режим доступа: <http://www.ruscorpora.ru/>
73. Национальный электронный звуковой депозитарий. – Режим доступа: <http://nezd.ru/en/catalog/>
74. Номенклатура ВАК. – Режим доступа: <http://docs.cntd.ru/document/542610966>
75. О едином цифровом пространстве научных знаний / А.Б. Антопольский, Н.Е. Калёнов, В.А. Серебряков, А.Н. Сотников // Вестник Российской ака-

- демии наук. – 2019. – Т. 89, № 7. – С. 728–735. – DOI: 10.31857/S0869-5873897728-735
76. О развитии терминологической базы данных в виде комплекса отраслевых информационно-поисковых тезаурусов / А.Б. Антопольский, В.Н. Белоозеров, Н.Е. Калёнов, Т.С. Маркарова // Информационные ресурсы России. – 2018. – № 5. – С. 15–24.
 77. Общественные консультации по законопроекту о научной и научно-технической деятельности. – Режим доступа: <https://www.preobra.ru/nauka>
 78. Организации и персонал, выполняющие научные исследования и разработки // ФГБНУ НИИ РИНКЦЭ. – 2018. – Режим доступа: http://csrs.ru/archive/stat_2018_institutions/institutions_2018.pdf
 79. Открытый каталог научных конференций, выставок и семинаров. – Режим доступа: <http://konferencii.ru>
 80. Паспорт Национального проекта «Наука». – Режим доступа: <http://static.government.ru/media/files/vCAoi8zEXRVsuy2Yk7D8hvQbpbUSwO8y.pdf>
 81. Перечень поручений по итогам заседания Совета при Президенте по науке и образованию, состоявшегося 27.11.2018 г. Утвержден 29 декабря 2018 года. Пр-2558, п. 3 а) 2 «О создании современной цифровой инфраструктуры для хранения и анализа научно-технической информации, а также для обмена такой информацией». – Режим доступа: <http://kremlin.ru/acts/assignments/orders/59632>
 82. План S и рекомендации по его внедрению. – Режим доступа: <https://www.coalition-s.org>
 83. Положение о федеральной государственной информационной системе «Национальная электронная библиотека». Утверждено Постановлением Правительства РФ от 20 февраля 2019 г. № 169. – Режим доступа: <http://government.ru/docs/35825>
 84. Портал Музеев СО РАН. – Режим доступа: <http://sbras.mmc.nsu.ru/?int=VIEW&class=ROOT&templ=VIEW&target=XML>
 85. Постановление Правительства РФ от 20.02.2019 г. № 169 «Об утверждении Положения о федеральной государственной информационной системе “Национальная электронная библиотека” и методики отбора объектов Национальной электронной библиотеки». – Режим доступа: <https://www.garant.ru/products/ipo/prime/doc/72084144/>
 86. Постановление Правительства РФ № 1109 от 29 августа 2019 г. «Об утверждении Правил предоставления субсидии из федерального бюджета открытому акционерному обществу “Научное издательство ‘Большая Российская энциклопедия’ ” на создание и функционирование общенационального интерактивного энциклопедического портала». – Режим доступа: <http://government.ru/docs/all/123635/>
 87. Постановление Правительства РФ от 24 июля 1997 г. N 950 «Об утверждении Положения о государственной системе научно-технической информации (с изменениями и дополнениями)». – Режим доступа: <http://base.garant.ru/11901351/#ixzz67ym1Z596>
 88. Призыв Жюрье к открытой науке и библиоразнообразию. – Режим доступа: <https://jussieucall.org/jussieu-call/>

89. Приказ Министерства образования и науки РФ от 22 октября 2015 г. N 1201 «О межведомственном координационном совете по организации предоставления лицензионного доступа к информационным наукометрическим базам данных и полнотекстовым научным ресурсам». – Режим доступа: <https://base.garant.ru/71779726>
90. Программа «Цифровая экономика Российской Федерации». Утверждена распоряжением Правительства РФ от 28 июля 2017 г. № 1632-р. Режим доступа: <http://static.government.ru/media/files/9gFM4FHj4PsB79I5v7yLVuPgu4bvR7M0.pdf>
91. Проект обновленного Положения о Национальной электронной библиотеке. – Режим доступа: <http://www.unkniga.ru/images/docs/2017/polozhene-neb-2-variant.pdf>
92. Разработка онтологии информационного пространства знаний на основе дефинитивных связей / Антопольский А.Б. (ИНИОН РАН, Москва, Россия), Белоозеров В.Н. (ВИНИТИ РАН, Москва, Россия), Калёнов Н.Е. (БЕН РАН, Москва, Россия), Маркарова Т.С. (Москва, Россия), Шабурова Н.Н. (ИФП СО РАН, Новосибирск, Россия), Якшин М.М. (БЕН РАН, Москва, Россия) // Перспективные направления научных исследований и критические технологии в классификационных системах: Научно-практическая конференция: Сборник докладов / ВИНТИ РАН. – Москва, 2017. – 25–27 октября. – С. 16–20.
93. Разработка семантической сети ключевых слов на основе дефинитивных связей / А.Б. Антопольский, В.Н. Белоозеров, Н.Е. Калёнов, Н.Н. Шабурова, М.М. Якшин // Научно-техническая информация. Сер. 1: Организация и методика информационной работы. – 2017. – № 11. – С. 19–23.
94. Разработка стратегии проекта «общественное достояние» и оценка экономического эффекта ее реализации: отчет ЦЭМИ РАН / рук. А.Н. Козырев. – 2016. – Режим доступа: <https://roem.ru/wp-content/uploads/2016/03/report.pdf>
95. Распоряжение Правительства РФ от 25.08.2016 № 1791-р «Об образовании межведомственной рабочей группы по вопросам, связанным с созданием на основе электронной версии Большой российской энциклопедии и других российских научных энциклопедий Общенационального научно-образовательного интерактивного энциклопедического портала». – Режим доступа: <https://rulaws.ru/goverment/Rasporyazhenie-Pravitelstva-RF-ot-25.08.2016-N-1791-g/>
96. Распоряжение Правительства РФ № 2804-р от 26 ноября 2019 г. «О создании Общенационального интерактивного энциклопедического портала». – Режим доступа: <http://publication.pravo.gov.ru/Document/View/0001201911290032?>
97. Распоряжение Правительства РФ от 26 ноября 2015 г. N 2421-р «Об утверждении состава Межведомственного совета по развитию Национальной электронной библиотеки». – Режим доступа: <https://base.garant.ru/71261096/>
98. Распоряжение Правительства РФ от 28 августа 2019 г. № 1904-р «Об утверждении плана развития Национальной электронной библиотеки». – Режим доступа: <http://static.government.ru/media/files/m2vIkqFcLuTerGCxh H94FPICAaJkAC1o.pdf>

99. Российские ученые должны быть обеспечены всеми необходимыми информационными ресурсами. – Режим доступа: https://minobrnauki.gov.ru/ru/press-center/card/?id_4=2241
100. Система Музеев РАН. – Режим доступа: <http://www.artan.ru/?q=ru/sysmus>
101. Совершенствование системы научных публикаций: рекомендации круглого стола при Комитете Государственной Думы по науке и образованию. – Режим доступа: <https://raser.ru/dokumenty/vneshnie-dokumenty>
102. СОЦИОНЕТ. Научное информационное пространство. – Режим доступа: <https://socionet.ru>
103. Список журналов Scopus. – Режим доступа: <http://ores.su/ru/journals/scopus>
104. Терминологическая система «Терминобаза». – Режим доступа: http://class.labs.benran.ru/users/sign_in
105. Указ Президента РФ № 808 от 24.12.2014 г. «Об утверждении Основ государственной культурной политики». – Режим доступа: <https://base.garant.ru/70828330/>
106. Усталов Д.А. Каталоги лингвистических ресурсов: состояние и перспективы // Молодой ученый. – 2012. – № 12. – С. 148–152. – Режим доступа: <https://moluch.ru/archive/47/5955/>
107. Федеральная резервная система банков знаний. – Режим доступа: <https://noosphere.ru>
108. Федеральный закон «О библиотечном деле» от 29.12.1994 N 78-ФЗ ст. 18.1 (ред. от 03.07.2016). – Режим доступа: <http://fzrf.su/zakon/o-bibliotechnom-dele-78-fz/st-18.1.php>
109. Фундаментальная электронная библиотека «Русская литература и фольклор». – Режим доступа: <http://www.feb-web.ru>
110. Центральная научная медицинская библиотека Первого МГМУ им. И.М. Сеченова. – Режим доступа: <http://www.scsml.rssi.ru>
111. Центральный фондовый каталог (ЦФК) Архивного фонда РФ. – Режим доступа: <http://cfc.rusarchives.ru/CFC-search>
112. ЭКБСОН. Информационная система доступа к электронным каталогам библиотек сферы образования и науки. – Режим доступа: <http://www.vlibrary.ru>
113. Экономика Рунета. Экосистема цифровой экономики России. Российская ассоциация электронных коммуникаций: Ежегодный доклад 2017 г. – Режим доступа: <https://raec.ru/live/raec-news/10096/>
114. Электронная библиотека диссертаций. – Режим доступа: <http://diss.rsl.ru/>
115. Ataeva O.M., Serebryakov V.A., Tuchkova N.P. Mathematical Physics Branches: Identifying Mixed Type Equations // Lobachevskii Journal of Mathematics. – 2019. – Vol. 40, N 7. – P. 876–886. – DOI: 10.1134/S1995080219070047
116. Ataeva O.M., Serebryakov V.A. Information Model of LibMeta Digital Library // Lobachevskii Journal of Mathematics. – 2019. – Vol. 40, № 7. – P. 861–875. – DOI:10.1134/S1995080219070035
117. Bar-Ilan J. Scientometrics. – 2008. – Bibl. p. 74, 257–271.
118. Bibliometrics // The Leiden Manifesto for research metrics [Электронный ресурс] / Diana Hicks, Paul Wouters, Ludo Waltman, Sarah de Rijcke, Ismael Rafols Nature. – 2015. – 23 April. – P/ 520, 429–431. – DOI: 10.1038/520429 a. – URL: <https://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351>

119. CLARIN – European Research Infrastructure for Language Resources and Technology. – URL: <https://www.clarin.eu/>
120. Confederation of Open Access Repositories (COAR.). – URL: <https://www.coar-repositories.org/?s=Plan+S>
121. DBpedia. – URL: <https://ru.wikipedia.org/wiki/DBpedia>
122. EBSCO Information Services. – URL: <https://www.ebsco.com/>
123. Establishing correspondences between Web of Science subject areas and UDC / V.N. Beloozerov, M.R. Biktimirov, A.B. Antopolskij, O.A. Antoshkova, T.S. Astaxova, O.V. Smirnova // International UDC seminar 2017 «Faceted Classification Today» 14–15 September, 2017, London UK. – URL: <http://seminar.udcc.org/2017/abstracts.php#beloozero>
124. European Language Resources Association. – URL: <http://www.elra.info/en/>
125. Fair Open Access Alliance. – URL: <https://www.fairopenaccess.org/>
126. Fields of Science. – URL: <http://www.oecd.org/science/inno/38235147.pdf>
127. Future of Scholarly Publishing and Scholarly Communication: Report of the Expert Group to the European Commission. – URL: <https://op.europa.eu/en/publication-detail/-/publication/464477b3-2559-11e9-8d04-01aa75ed71a1>
128. Guidance on the Implementation of Plan S. – URL: <https://www.coalition-s.org/implementation>
129. International Speech Communication Association. – URL: <https://www.isca-speech.org/iscaweb/index.php/about-isca>
130. Linguistic Data Consortium (LDC). – URL: <https://www ldc.upenn.edu/>
131. Мнемосине. Территория исторической памяти. – URL: <http://isaran.ru/?q=welcome>
132. OECD Principles and Guidelines for Access to Research Data from Public Funding / Open Knowledge Foundation. – URL: <http://okfn.org>
133. OLAC, the Open Language Archives Community. – URL: <http://www.language-archives.org/>
134. Open Access 2020 Initiative. – URL: <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access>
135. Open Knowledge Foundation – URL: <http://okfn.org/>
136. Open Library. – URL: <http://openlibrary.org/>
137. Outils et Ressources pour un Traitement Optimisé de la LANGue. – URL: <https://www.ortolang.fr/>
138. Principles for Open Data in Science. – URL: <https://pantonprinciples.org/>
139. Public Discussion on Publishing Reform. – URL: <https://publishing-reform.gitlab.io/>
140. Research Excellence Framework. – URL: <http://www.ref.ac.uk/>
141. San Francisco Declaration on Research Assessment. – URL: <https://sfidora.org> – Пепевод: В.И. Карнышева. Источник: http://piohelp.ru/myfiles/dora/DORA_rus.html
142. Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level. 29–30 January 2004. Final Communiqué. – URL: https://www.oecd.org/sti/sciencetechnologyandinnovationforthe21stcenturymeetingoftheoecdcommitteefor_scientificand-technologicalpolicyatministeriallevel29-30january2004-finalcommunique.htm
143. SIL International. – URL: <https://www.sil.org/>
144. STN International. – URL: http://www.stn-international.de/stn_home.html

145. TELRI – Trans-European Language Resources. – URL: <http://telri.nytud.hu>
146. The Concordat on Open Research Data. – URL: <https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf>
147. The Development of a Semantic Network of Keywords Based on Definitive Relationships /A.B. Antopol'skii, V.N. Beloozerov, N.E. Kalenov, N.N. Shaburova, M.M. Yakshin // Scientific and Technical Information Processing. – 2017. – Vol. 44, N 4. – P. 261–265. – DOI: 10.3103/S0147688217040062
148. Ulrich's Periodicals Directory. – URL: <http://ulrichsweb.serialssolutions.com/login>
149. Web of Science. – URL: <https://clarivate.ru/products/web-of-science/>

ПРИЛОЖЕНИЯ

Приложение 1. Общественные науки в ГРНТИ

- 00 Общественные науки в целом
- 02 Философия
- 04 Социология
- 06 Экономика и экономические науки
- 10 Государство и право. Юридические науки
- 11 Политика и политические науки
- 12 Науковедение
- 13 Культура. Культурология
- 14 Народное образование. Педагогика
- 15 Психология
- 16 Языкознание
- 17 Литература. Литературоведение. Устное народное творчество
- 18 Искусство. Искусствоведение
- 19 Массовая коммуникация. Журналистика. Средства массовой информации
- 20 Информатика
- 21 Религия. Атеизм
- 23 Комплексное изучение отдельных стран и регионов
- 26 Комплексные проблемы общественных наук

Приложение 2. Фрагмент номенклатуры ВАК по общественным наукам

ГУМАНИТАРНЫЕ НАУКИ

- 07.00.00 Исторические науки и археология
- 10.00.00 Филологические науки
- 10.01.00 Литературоведение
- 10.02.00 Языкознание
- 09.00.00 Философские науки

ИСКУССТВОВЕДЕНИЕ И КУЛЬТУРОЛОГИЯ

- 17.00.00 Искусствоведение
 - 24.00.00 Культурология
 - 05.25.00 Документальная информация
 - 05.25.02 Документалистика, документоведение, архивоведение
 - 05.25.03 Библиотечковедение, библиографоведение и книговедение
 - 05.25.05 Информационные системы и процессы
- ## СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ И ОБЩЕСТВЕННЫЕ НАУКИ
- 19.00.00 Психологические науки
 - 08.00.00 Экономические науки
 - 13.00.00 Педагогические науки
 - 22.00.00 Социологические науки
 - 12.00.00 Юридические науки
 - 23.00.00 Политология

Приложение 3. Рубрикатор РГНФ

- 01 ИСТОРИЧЕСКИЕ НАУКИ
- 02 ЭКОНОМИЧЕСКИЕ НАУКИ
- 03 ОБЩЕСТВЕННЫЕ НАУКИ
- 03–100 Философские науки
- 03–300 Социологические науки
- 03–400 Юридические науки
- 03–500 Науковедение, социальная история науки и техники
- 03–600 Социальная география
- 04 ФИЛОЛОГИЧЕСКИЕ НАУКИ; ИСКУССТВОВЕДЕНИЕ
- 04–100 Филология
- 04–110 Литературоведение
- 04–120 Языкознание
- 04–130 Текстология
- 04–140 Фольклористика
- 04–200 Искусствоведение
- 04–210 Изобразительное искусство
- 04–220 Теория и история архитектуры
- 04–230 Музыковедение
- 04–240 Музыкальная фольклористика
- 04–250 Театроведение
- 06 КОМПЛЕКСНОЕ ИЗУЧЕНИЕ ЧЕЛОВЕКА; ПСИХОЛОГИЯ;
ФУНДАМЕНТАЛЬНЫЕ ПРОБЛЕМЫ ОБРАЗОВАНИЯ
- 06–100 Методология комплексного изучения человека
- 06–200 Психология
- 06–300 Фундаментальные проблемы образования
- 07 ГЛОБАЛЬНЫЕ ПРОБЛЕМЫ И МЕЖДУНАРОДНЫЕ ОТНОШЕНИЯ

Приложение 4. Рубрикатор РФФ

08 ГУМАНИТАРНЫЕ И СОЦИАЛЬНЫЕ НАУКИ

08–100 Исторические науки
08–150 Экономические науки
08–200 Философские науки
08–250 Политология
08–300 Социология
08–350 Юридические науки
08–400 Науковедение
08–450 Филологические науки
08–500 Искусствоведение
08–550 Психология
08–600 Педагогика
08–650 Социальная и экономическая география
08–700 Культурология

Приложение 5. Рубрикатор FOS (ОЭСР) по социальным и гуманитарным наукам

5. SOCIAL SCIENCES

5.1. Psychology

- Psychology (including human – machine relations);
- Psychology, special (including therapy for learning, speech, hearing, visual and other physical and mental disabilities);

5.2. Economics and Business

- Economics, Econometrics; Industrial relations;
- Business and Management;

5.3. Educational sciences

- Education, general; including training, pedagogy, didactics;
- Education, special (to gifted persons, those with learning disabilities);

5.4. Sociology

- Sociology; Demography; Anthropology, Ethnology;
- Social topics (Womens' and gender studies; Social issues; Family studies, Social work);

5.5. Law

- Law, Criminology, Penology;

5.6. Political science

- Political science; Public administration; Organisation theory;

5.7. Social and economic geography

- Environmental sciences (social aspects); Cultural and economic geography; Urban studies (planning and development); Transport planning and social aspects of transport (Transport engineering to be 2.1);
- 5.8. Media and communications
 - Journalism; Information science (social aspects); Library science; Media and socio-cultural communication;
- 5.9. Other social sciences
 - Social sciences, interdisciplinary;
 - Other social sciences;
- 6. HUMANITIES
 - 6.1. History and Archaeology
 - History (history of science and technology to be 6.3, history of specific sciences to be under the respective headings); Archaeology;
 - 6.2. Languages and Literature
 - General language studies; Specific languages; General literature studies; Literary theory; Specific literatures; Linguistics;
 - 6.3. Philosophy, Ethics and Religion
 - Philosophy, History and philosophy of science and technology;
 - Ethics (except ethics related to specific subfields); Theology; Religious studies;
 - 6.4. Arts (arts, history of arts, performing arts, music)
 - Arts, Art history; Architectural design; Performing arts studies (Musicology, Theater science, Dramaturgy); Folklore studies;
 - Studies on Film, Radio and Television;
 - 6.5. Other humanities

Приложение 6. Рубрикатор WOS по социальным наукам

Anthropology
 Area Studies
 Business
 Business, Finance
 Cultural Studies
 Communication
 Criminology & Penology
 Demography
 Economics
 Education & Educational Research
 Education, Special
 Environmental Studies
 Ergonomics
 Ethics
 Ethnic Studies
 Family Studies

Geography
Gerontology
Health Policy & Services
History
History & Philosophy Of Science
History of Social Sciences
Hospitality, Leisure, Sport & Tourism
Industrial Relations & Labor
Information Science & Library Science
International Relations
Law
Linguistics
Management
Nursing
Planning & Development
Political Science
Psychiatry
Psychology, Applied
Psychology, Biological
Psychology, Clinical
Psychology, Developmental
Psychology, Educational
Psychology, Experimental
Psychology, Mathematical
Psychology, Multidisciplinary
Psychology, Psychoanalysis
Psychology, Social
Public Administration
Public, Environmental & Occupational Health
Rehabilitation
Social Issues
Social Sciences, Biomedical
Social Sciences, Interdisciplinary
Social Sciences, Mathematical Methods
Social Work
Sociology
Substance Abuse
Transportation
Urban Studies
Women's Studies

Приложение 7.

Рубрикатор WOS по искусству и гуманитарным наукам

Archaeology
Architecture
Art
Asian Studies
Classics
Cultural studies
Dance
Film, Radio, Television
Folklore
History
History & Philosophy Of Science
Humanities, Multidisciplinary
Language & Linguistics
Literary Reviews
Literary Theory & Criticism
Literature
Literature, African, Australian, Canadian
Literature, American
Literature, British Isles
Literature, German, Dutch, Scandinavian
Literature, Romance
Literature, Slavic
Medieval & Renaissance Studies
Music
Philosophy
Poetry
Religion
Theater

Приложение 8.

Фрагмент рубрикатора Scopus

EXPAND ARTS AND HUMANITIES
Anthropology
Archaeology
Conservation
History
History and Philosophy of Science
Literature
Museology

Music
 Philosophy
 Religious Studies
 Visual Arts
 EXPAND BUSINESS, MANAGEMENT AND ACCOUNTING
 Accounting
 Administration and Marketing
 Beauty and Fashion
 Business
 Hospitality, Sport and Tourism Management
 Management Information Systems
 Strategy and Strategic Management
 EXPAND ECONOMICS AND FINANCE
 Agricultural and Natural Resource Economics
 Business Economics
 Economic Development, Technological Change, and Growth
 Economic History
 Economic Systems
 Financial Economics
 General Economics and Teaching
 Health, Education and Welfare
 Industrial Organization
 International Economics
 Labor and Demographic Economics
 Law and Economics
 Macroeconomic and Monetary Economics
 Mathematical and Quantitative Methods
 Methodology and History of Economic Thought
 Microeconomics
 Public Economics
 Urban, Rural and Regional Economics
 EXPAND PSYCHOLOGY
 Applied Psychology
 Clinical Psychology
 Developmental and Educational Psychology
 Experimental and Cognitive Psychology
 Neuroscience and Physiological Psychology
 Organizational Behavioral Psychology
 Personality, Social and Criminal Psychology
 Psychology (General)
 EXPAND SOCIAL SCIENCES
 Childhood Studies
 Education
 Geography, Planning and Development
 Health
 Language and Linguistics
 Law

Приложение 9.

Функциональная классификация типов ресурсов по результатам мониторинга

1. ПЕРВИЧНЫЕ ТРАДИЦИОННЫЕ РЕСУРСЫ

- 1.1. Библиотечные фонды
 - 1.1.1. Коллекции по тематике, источнику
 - 1.1.2. Фонды литературы на языках народов России
 - 1.1.3. Фонды иностранной литературы
 - 1.1.4. Фонды отечественной литературы
 - 1.1.5. Фонды периодики
 - 1.1.6. Фонды редкой книги
 - 1.1.7. Коллекции рукописей
- 1.2. Печатные периодические и продолжающиеся издания
- 1.3. Архивные фонды
 - 1.3.1. Личные фонды
 - 1.3.2. Разряды (тематические коллекции)
 - 1.3.3. Фонды учреждений
- 1.4. Музейные фонды
 - 1.4.1. Музейные коллекции
 - 1.4.2. Экспозиции и выставки
- 1.5. Фонотеки
 - 1.5.1. Аудиовизуальные лингвистические ресурсы
 - 1.5.2. Национальный звуковой депозитарий
 - 1.5.3. Фольклорные аудиокolleкции
- 1.6. Коллекции изображений (иконотеки), фотографий, негативов
- 1.7. Киноматериалы
- 1.8. Микроформы

2. ПЕРВИЧНЫЕ ЭЛЕКТРОННЫЕ РЕСУРСЫ

- 2.1. Публикации учреждений – полные тексты
 - 2.1.1. Издания институтов: монографии и продолжающиеся
 - 2.1.2. Публикации сотрудников в СМИ
 - 2.1.3. Труды по подразделениям
 - 2.1.4. Труды сотрудников
 - 2.1.5. Материалы научных мероприятий
 - 2.1.6. СМИ об учреждении

- 2.2. Электронные периодические издания
 - 2.2.1. По типу
 - 2.2.1.1. Самостоятельные, имеющие e-ISSN
 - 2.2.1.2. Версии печатных изданий
 - 2.2.1.3. Выборочные публикации
 - 2.2.1.4. Переводные версии
 - 2.2.2. По форме представления
 - 2.2.2.1. Полные тексты
 - 2.2.2.2. Содержание номеров
 - 2.2.2.3. Аннотации статей
 - 2.2.2.4. Указатели выпусков
 - 2.2.3. Размещение в НЭБ, КиберЛенинке, других ЭБ
- 2.3. Электронные библиотеки
 - 2.3.1. Коллекции периодических и продолжающихся изданий
 - 2.3.2. Коллекции рекомендуемой литературы к учебным программам
 - 2.3.3. Коллекции рукописей
 - 2.3.4. Репозитории
 - 2.3.5. Эпос на DVD
 - 2.3.6. Другие и универсальные ЭБ
- 2.4. Электронные архивы (не входящие в систему архивов РАН)
 - 2.4.1. Электронные системы исторической фактографии (ИСИ СО РАН)
 - 2.4.2. Социологические материалы, биографии
 - 2.4.3. Материалы экспедиций
 - 2.4.4. Массивы фольклорных записей
 - 2.4.5. Материалы Комиссии по истории ВОВ
- 2.5. Неопубликованные документы
 - 2.5.1. Данные учета археологических памятников
 - 2.5.2. Диссертационные комплекты, включая авторефераты, отзывы и др.
 - 2.5.3. Доклады учреждений
 - 2.5.4. Конкурсы
 - 2.5.5. Новостная информация
 - 2.5.6. Планы научных исследований
 - 2.5.7. Описания проектов, работ по грантам, участия в программах
 - 2.5.8. Описания услуг
 - 2.5.9. Отчеты учреждения
 - 2.5.10. Очерки направлений деятельности и результатов
 - 2.5.11. Сведения об объектах интеллектуальной собственности
 - 2.5.12. Сведения о составе и деятельности научных институций (советов, комиссий)
 - 2.5.13. Экспедиционные отчеты
 - 2.5.14. Экспертные заключения
- 2.6. Сведения о персоналиях
 - 2.6.1. «Кто есть кто»
 - 2.6.2. БД по личным архивным фондам
 - 2.6.3. Биографии, биографические справки
 - 2.6.4. Биографические энциклопедии
 - 2.6.5. Каталоги лиц

- 2.6.6. Лауреаты
- 2.6.7. Мемориальные музеи
- 2.6.8. Мемории, «книги памяти»
- 2.6.9. Персональные выставки, коллекции, экспозиции
- 2.6.10. Персональные сайты
- 2.7. Справочные и энциклопедические ресурсы
 - 2.7.1. Генеалогические справочники
 - 2.7.2. Календарный цикл обрядов и обычаев
 - 2.7.3. Летописи
 - 2.7.4. Справочники учреждений
 - 2.7.5. Терминологические словари и справочники
 - 2.7.6. Указатели имен, фактов, событий
 - 2.7.7. Электронная библиотека знаний
 - 2.7.8. Энциклопедии онлайн или на переносимых носителях
- 2.8. Фотогалереи, фотобанки
- 2.9. Видеоматериалы

3. ВТОРИЧНЫЕ РЕСУРСЫ

- 3.1. Вторичные данные о публикациях учреждений
 - 3.1.1. Перечни библиографических описаний
 - 3.1.2. Аннотации
 - 3.1.3. Обложки
 - 3.1.4. Содержание публикаций
 - 3.1.5. Предисловия, вступительные главы и др.
- 3.2. Библиографические БД
- 3.3. Библиографические указатели и перечни
 - 3.3.1. Биобиблиографические
 - 3.3.2. Источниковые
 - 3.3.3. Проблемные
 - 3.3.4. Ретроспективные
 - 3.3.5. Тематические
 - 3.3.6. Хронологические
- 3.4. Библиотечные каталоги
 - 3.4.1. Алфавитные
 - 3.4.2. Каталоги периодики
 - 3.4.3. Каталоги фондов и коллекций
 - 3.4.4. Краеведческие
 - 3.4.5. Предметные
 - 3.4.6. Систематические
 - 3.4.7. Топографические
- 3.5. Архивный справочно-поисковый аппарат
 - 3.5.1. Каталоги
 - 3.5.2. Описи
 - 3.5.3. Путеводители
 - 3.5.4. Указатели к фондам
 - 3.5.5. Указатели фондообразователей

- 3.6. Музейные каталоги
 - 3.6.1. Виртуальные выставки
 - 3.6.2. Инвентарные книги
 - 3.6.3. Каталоги коллекций
 - 3.6.4. Научные каталоги
 - 3.6.5. Онлайн экскурсии, 3D-туры
 - 3.6.6. Экспозиции
- 3.7. Сообщения о мероприятиях
 - 3.7.1. Программы
 - 3.7.2. Презентации
- 3.8. Реферативные и аналитические продукты
 - 3.8.1. Журналы
 - 3.8.2. Сборники
 - 3.8.3. Обзоры, дайджесты
- 3.9. Каталоги Интернет-ссылок

4. ЛИНГВИСТИЧЕСКИЕ РЕСУРСЫ

- 4.1. Корпуса текстов
- 4.2. Словарные БД и электронные картотеки
 - 4.2.1. БД лексической и семантической сочетаемости
 - 4.2.2. Диалектологические
 - 4.2.3. Орфографические, орфоэпические
 - 4.2.4. Рубрикаторы, классификации
 - 4.2.5. Словари корпусов
 - 4.2.6. Словари лексических функций
 - 4.2.7. Тезаурусы, семантические словари
 - 4.2.8. Терминологические
 - 4.2.9. Топонимические
 - 4.2.10. Этимологические БД
- 4.3. Лингвистические процессоры
 - 4.3.1. Многоцелевые лингвистические процессоры
 - 4.3.2. Действующая модель «Смысл – текст»
 - 4.3.3. Процессоры для машинного перевода
 - 4.3.4. Анализаторы устной речи
 - 4.3.5. Синтезаторы устной речи
 - 4.3.6. Морфологические БД, модели, анализаторы
 - 4.3.7. Орфографические, парсеры
 - 4.3.8. Синтаксические анализаторы
 - 4.3.9. Системы транскрипции
 - 4.3.10. Процессоры для представления словарей онлайн
 - 4.3.11. Специальные лингвистические процессоры
 - 4.3.11.1. Для работы с иероглификой (РАМЕЯ)
 - 4.3.11.2. Для оценки образовательных тестов
 - 4.3.11.3. Для анализа ответов обучаемого
 - 4.3.11.4. Для извлечения знаний из текста
 - 4.3.11.5. Для грамматической разметки текстов
 - 4.3.11.6. Для когнитивных экспериментов

- 4.4. Грамматические ресурсы (академическая грамматика)
- 4.5. Описания языков, реестры языков
- 4.6. Лингвистические атласы
- 4.7. Комплексные лингвистические АИС (сайты)
- 4.8. Этнолингвистические и социоллингвистические БД
- 4.9. Аудиовизуальные лингвистические ресурсы

5. СПЕЦИАЛИЗИРОВАННЫЕ РЕСУРСЫ

- 5.1. Инфометрические и вебометрические ресурсы.
- 5.2. Электронные научные издания (= базы данных)
- 5.3. Мультимедийные ресурсы
 - 5.3.1. Аудиовизуальные лингвистические ресурсы
 - 5.3.2. Комплексные АИС (Русская словесность)
 - 5.3.3. Мультимедийные издания на CD
 - 5.3.4. Фольклорные коллекции
 - 5.3.5. Экспедиционные программные комплексы
- 5.4. Археологические ресурсы
 - 5.4.1. АИС для учета археологических памятников (археограф)
 - 5.4.2. Виртуальные модели исторических объектов
 - 5.4.3. Древнерусские памятники в фотодокументах
 - 5.4.4. Сводные археологические БД
- 5.5. Комплексные АИС изображений
 - 5.5.1. БД графики
 - 5.5.2. Фотолетописи
 - 5.5.3. Фотобанки
- 5.6. Экспертные системы и базы знаний
- 5.7. Географические информационные системы, атласы
- 5.8. Информационно-аналитические системы и БД
 - 5.8.1. Банки социологических данных
 - 5.8.2. БД показателей
 - 5.8.3. Интерактивные исследовательские АИС
 - 5.8.4. Программные средства для кодикологического анализа рукописных памятников
 - 5.8.5. Системы для оценки научной эффективности
 - 5.8.6. Таблицы в разрезе основных атрибутов законов
 - 5.8.7. Хронологические БД

6. ТИПЫ РЕСУРСОВ В ИНТЕРНЕТЕ

- 6.1. Основной сайт учреждения
- 6.2. Сайты или самостоятельные страницы научных подразделений
- 6.3. Сайты-сателлиты, в том числе
 - 6.3.1. Сайты проектов учреждения
 - 6.3.2. Тематические сайты
 - 6.3.3. Сайты журналов, издаваемых учреждением
 - 6.3.4. Персональные страницы сотрудников
 - 6.3.5. Сайты аффилированных с учреждением структур

- 6.4. Сведения об учреждении или материалы учреждения на сайтах вышестоящих научных учреждений
- 6.5. Сведения об учреждении или материалы учреждения в общих информационных системах
- 6.6. Аккаунты (группы) учреждения в социальных сетях

Приложение 10.

Упрощенная функциональная классификация, используемая в Навигаторе НИРОН

- 1. Библиотеки
- 2. Архивы
- 3. Музеи
- 4. Каталоги
 - 4.1. Традиционные
 - 4.2. Электронные
 - 4.3. Смешанные
- 5. Электронные коллекции и библиотеки
 - 5.1. Публикации учреждений
 - 5.2. Электронные библиотеки
 - 5.3. Смешанные коллекции
 - 5.4. Репозитории
 - 5.5. ЭБ на переносимых носителях
- 6. Информационные системы
 - 6.1. Информационно-аналитические БД
 - 6.2. ГИС
 - 6.3. Индексы цитирования
 - 6.4. Инфометрические ресурсы
 - 6.5. Электронные представления памятников
 - 6.6. Экспертные АИС и базы знаний
 - 6.7. Мультимедийные, 3D-, VR-системы
 - 6.8. Комплексные АИС
- 7. Справочники, энциклопедии
 - 7.1. Энциклопедии
 - 7.2. Справочники
 - 7.3. Хронологии
 - 7.4. Фонды учреждений
 - 7.5. Архивные разряды
- 8. Персональные ресурсы
 - 8.1. Персональные сайты и страницы
 - 8.2. Личные фонды
 - 8.3. Сотрудники учреждений
 - 8.4. Указатели лиц
 - 8.5. Биобиблиографии

9. Лингвистические ресурсы
 - 9.1. Корпуса текстов
 - 9.2. Словарные БД и электронные картотеки
 - 9.3. Лингвистические процессоры
 - 9.4. Грамматические ресурсы
 - 9.5. Описания языков, реестры языков
 - 9.6. Лингвистические атласы
 - 9.7. Этно- и социолингвистические БД
 - 9.8. Комплексные лингвистические АИС (сайты)
 - 9.9. Информационные языки
10. Периодика
11. Библиографии
 - 11.1. Перечни трудов учреждений / сотрудников
 - 11.2. Аннотированные или реферативные указатели
 - 11.3. Предметные, тематические библиографии
 - 11.4. Библиографические БД
12. Мероприятия
13. Неопубликованные материалы
 - 13.1. Диссертации
 - 13.2. Отчеты
 - 13.3. Экспертные заключения
 - 13.4. Описания проектов, грантов, экспедиций
 - 13.5. Очерки деятельности учреждения, подразделения
14. Медиаресурсы
 - 14.1. Банки изображений и фото
 - 14.2. Аудиоресурсы
 - 14.3. Видеоресурсы
 - 14.4. Киноматериалы
 - 14.5. Смешанные ресурсы
15. Прочие интернет-ресурсы
 - 15.1. Сайты подразделений
 - 15.2. Сайты-сателлиты
 - 15.3. Аккаунты в социальных сетях
 - 15.4. Ресурсы во внешних ЭБ и АИС
 - 15.5. Каталоги ссылок

Приложение 11.

Карта LRE [электронная карта лингвистических ресурсов]

В скобках указано количество ЛИР в каталоге ELRA, в которых имеются соответствующие данные.

Оценка ЛИР

Оценочные данные (230)

Инструменты оценки (71)

- Оценочный пакет (25)
- Методология оценки / стандарты / руководящие принципы (15)
- Ресурс-данные
- Корпус (2920)
- Лексикон (666)
- Онтология (162)
- Грамматика / Языковая модель (82)
- Терминология (66)
- Банки деревьев зависимостей (42)
- Ресурс-руководство
 - Представительство-аннотации. Формализм / Руководство (62)
 - Языковые ресурсы / Технологии. Инфраструктура
 - Метаданные (10)
- Ресурс-инструмент
- Tagger / Parser (400)
- Инструмент аннотации (245)
- Корпус-инструмент (83)
- Распознаватель именованных объектов (60)
- Инструмент машинного перевода (51)
- Программный инструментарий (41)
- Токенайзер (35)
- Инструмент машинного обучения (32)
- Инструмент моделирования языков (29)
- Разногласие по смыслу слова (17)
- Распознаватель речи / Транскриптор (14)
- Обработка сигналов / Извлечение функций (14)
- Веб-сервис (9)
- Синтезатор преобразования текста в речь (9)
- Идентификатор языка (6)
- Распознаватель динамиков (4)
- Инструмент для анализа настроений (4)
- Просодический анализатор (3)
- Анализатор изображений (3)
- Инструмент разговорного общения (1)
- Состояние производства
- Существующие и используемые (2587)
- Недавно созданные (в процессе) (1408)
- Недавно созданные (законченные) (1290)
- Существующие обновленные (487)
- Другое (354)
- Не применимо (17)
- Доступность
- Свободно доступные (2772)
- Другое (1232)
- От владельца (1229)
- Из дата-центра (580)
- Нет в наличии (267)

Не применимо (63)
Модальность (семиотический тип)
Письменный (4355)
 Речь (430)
 Мультимодальные / мультимедиа (286)
 Не применимо (261)
Речь / Письменность (186)
Язык жестов (64)
Независимо от модальности и другое (561)
Использование ресурсов
Извлечение информации, поиск информации (608)
Машинный перевод, перевод речи в речь (532)
Разбор и тегирование (289)
Языковое моделирование (282)
Классификация документов, текстовая категоризация (201)
Распознавание / понимание речи (185)
Приобретение (181)
Дискурс (178)
Открытие / Представление знаний (172)
Смысл словосочетания (160)
Распознавание эмоций / генерация (154)
Оценка / Валидация (152)
Создание корпуса / Аннотирование (148)
Текстовое копирование (147)
Признание именованного субъекта (144)
Диалог (122)
Подведение итогов (96)
Ответы на вопросы (83)
Морфологический анализ (80)
Семантическая паутина (68)
Веб-сервисы (67)
Создание / аннотация аннотации (63)
Синтез речи (55)
Поколение естественных языков (54)
Машинное обучение (54)
Текстовое вложение и перефразирование (53)
Анализ мнений / Анализ настроений (40)
Распознавание / Генерация жестового языка (35)
Идентификация языка (35)
Идентификация личности (30)
Анафора, Coreference (30)
Семантическая ролевая маркировка (29)
Обнаружение и отслеживание тем (25)
Обработка мультимедийных документов (22)
Голосовое управление (3)
Другое (1566)
Тип языка

Одноязычный (2507)
 Независимый от языка (2206)
 Многоязычный (961)
 Двухязычный (380)
 Трёхязычный (84)
 Не применимо (5)
 Язык (по списку)

Приложение 12. Участники CLARIN

№	Страна	Национальный консорциум	Ведущий партнер CLARIN
	Австрия	Цифровые гуманитарные науки	ARCHE (Ресурсный центр для гуманитарных наук)
	Болгария	CLaDA-BG	
	Хорватия	HR-Кларин	Университет Загреба
	Кипр	Кларин-CY	Лаборатория исследований цифрового наследия (Технологический университет Кипра)
	Чехия	LINDAT / Кларин	Карлов университет в Праге
	Дания	Кларин-DK	Университет Копенгагена
	Эстония	CLARIN Эстония	Центр языковых ресурсов Эстонии
	Финляндия	FIN-Кларин	Университет Хельсинки
	Германия	CLARIN-Д	Тюбингенский университет
	Италия	Кларин-IT	Институт компьютерной лингвистики им. А. Дзамполли, Национальный исследовательский совет
	Греция	Кларин: Эль	Исследовательский центр ИЛСП-АФИНА
	Латвия	Кларин-LV	Институт математики и информатики Латвийского университета
	Венгрия	HunCLARIN	Научно-исследовательский институт лингвистики Венгерской академии наук
	Литва	Кларин-LT	Университет Витаутаса Магнуса
	Нидерланды	CLARIAH-NL	Утрехтский университет
	Норвегия	Кларино	Университет Бергена
	Польша	CLARIN PL	Вроцлавский технологический университет
	Португалия	ПОРТУГАНСКИЙ КЛАРИН	Университет Лиссабона
	Словения	CLARIN.SI	Институт Йозефа Стефана
	Швеция	SWE-Кларин	Språkbanken
	Франция	Numa-Num	Национальный центр научных исследований (CNRS)
	Исландия	КЛАРИН Исландия	Институт исландских исследований имени Арни Магнуссона
	Южная Африка	SADiLaR	Северо-Западный Университет
	Соединенное Королевство	Кларин-UK	Оксфордский университет

Приложение 13.

Индивидуальные веб-сервисы членов CLARIN

1. Weblicht: среда выполнения для автоматического аннотирования текстовых корпусов, созданная CLARIN-D. Лингвистические инструменты, такие как токенизаторы, часть речевых тегов и анализаторы, инкапсулированы в виде веб-сервисов, которые пользователь может объединить в пользовательские цепочки обработки.
2. Услуги NLP для польского языка от Вроцлавского технологического университета. Веб-сервис TaKIPi-WS предоставляет функциональные возможности для маркировки текста, лематизации, сегментации, морфологического анализа и токенизации. Запросы обрабатываются асинхронно. Кроме того, веб-сервис подключения к польскому WordNet.
3. Услуги NLP для португальского языка от группы NLX, Университет Лиссабона. Фрагментатор, токенизатор, разметчик (WSDL).
4. Услуги NLP для румынского и других языков от RACAI. Обработка текста, языковая идентификация и факторный перевод.
5. Веб-сервисы AVAtch. Веб-сервисы автоматической сегментации звука и распознавания видео, могут использоваться в ELAN.
6. Услуги NLP для нескольких языков от Лейпцигского университета. Широкий спектр услуг.
7. Услуги MAUC. Автоматическое фонетическое выравнивание для разных языков.
8. CLARIN-ES-лаборатории. Испанская виртуальная лаборатория по технологии CLARIN.

Приложение 14.

Топ-20 терминологических служб и банков данных

1. Центр по координации терминологии Евросоюза (TermCoord): <https://termcoord.eu/terminology-websites/>. Основная цель TermCoord заключается в оказании помощи переводчикам в выполнении их повседневных задач и содействии исследованию терминологии и управлению терминологией в подразделениях перевода, а также в увеличении вклада в терминологическую базу данных ЕС IATE. Поддерживаются каталоги терминологических ресурсов и организаций.
2. Интерактивная терминология для Европы (IATE): <https://iate.europa.eu/about>. Межинституциональная терминологическая база данных Европейского союза. ИАТЕ объединила все существующие терминологические базы данных переводческих служб ЕС в одну межведомственную базу данных, содержащую примерно 1,4 млн многоязычных записей и 9 млн терминов на 24 официальных языках ЕС. Текущая новая версия IATE была выпущена 7 ноября 2018 г.
3. Евротермбанк: <https://eurotermbank.com/>. Является крупнейшим централизованным терминологическим онлайн-банком для языков ЕС и исландского, взаимосвязан с другими терминологическими банками и ресурсами.

- EuroTermBank позволяет обмениваться данными с существующими национальными и международными базами данных, устанавливая отношения сотрудничества, согласовывая методологии и стандарты, разрабатывая и внедряя механизмы и процедуры обмена данными. Объем – 15 млн терминов.
4. WordNet: <https://wordnet.princeton.edu/>. Большая лексическая база данных английского языка. Существительные, глаголы, прилагательные и наречия сгруппированы в наборы когнитивных синонимов (synsets), каждый из которых выражает отдельную концепцию. Синсеты взаимосвязаны посредством концептуально-семантических и лексических отношений.
 5. TermNet: <https://www.termnet.org/>. Международная сеть терминологии является международным форумом сотрудничества для компаний, университетов, учреждений и ассоциаций, которые занимаются дальнейшим развитием глобального рынка терминологии.
 6. TermWiki: http://ru.termwiki.com/TermWiki_Home. Онлайн-портал терминологии, который позволяет пользователям искать, загружать, переводить и делиться терминами и определениями с другими пользователями (более 100 языков).
 7. TermScience: <http://www.termsciences.fr/termsciences/?lang=en>. Термины науки: это портал терминологии, разработанный INIST совместно с LORIA и ATILF. Его цель – продвигать, объединять и делиться терминологическими ресурсами (словарями специалистов, тезаурусами) государственного сектора и образовательных учреждений для создания общего терминологического справочного ресурса.
 8. EuroVoc: http://publications.europa.eu:80/eurovoc/index_en.htm. Многоязычный междисциплинарный тезаурус, охватывающий информацию ЕС, в частности Европейского парламента. Содержит термины на 22 языках.
 9. UNTERM: <https://unterm.un.org/UNTERM/portal/welcome>. Терминологическая база данных ООН. Содержит техническую и специализированную терминологию на каждом из шести официальных языков ООН (английский, французский, испанский, русский, мандаринский и арабский).
 10. Веб-платформа языковой индустрии (LIND-Web): https://ec.europa.eu/info/departments/translation/language-industry-platform-lind_e. Содержит факты и цифры по языковой индустрии ЕС. Они предоставляются языковыми специалистами, заинтересованными сторонами отрасли и институтами ЕС.
 11. Терминальный портал ФАО: <http://www.fao.org/faoterm/en/>. Продовольственная и сельскохозяйственная организация ООН. Он содержит более 80 тыс. записей на шести официальных языках Организации Объединенных Наций (английском, французском, испанском, русском, арабском и мандаринском).
 12. BabelNet: <https://babelnet.org>. Многоязычный энциклопедический словарь с лексикографическим и энциклопедическим охватом терминов на 50 языках и онтологией, которая объединяет понятия и именованные объекты в очень большую сеть семантических отношений, состоящую из более чем 9 млн записей.
 13. UNESCOTERM: <http://www.unesco.org/unescdi/index.php/eng/thesaurus/tous.118>. База данных терминологии ЮНЕСКО на английском, арабском, испанском, китайском, русском и французском языках.

14. Банк данных РОСТЕРМ: <http://www.gostinfo.ru/catalog/terminlist>. Единственный российский терминологический банк, представленный в международных каталогах. Содержит свыше 140 тыс. терминологических статей из ГОСТ, ГОСТ Р, стандартов ИСО и МЭК.
15. Терминология Тильды: <https://term.tilde.com/>. Услуги: извлечение терминов и поиск в облаке – около 5 млн стандартизированных и надежных терминов; облачные средства для управления терминологией и совместного использования; встроенное распознавание и поиск терминологии в инструментах перевода.
16. ВТОТЕРМ: <https://wto.sdlproducts.com/multiterm>. Терминологическая база данных, созданная Всемирной торговой организацией (ВТО). Поиск терминов и эквивалентов может быть выполнен на французском, английском и испанском языках.
17. Терминология Международной организации труда (МОТ): <https://www.ilo.org/inform/online-information-resources/databases/terminology/lang--en/index.htm>. Включает терминологическую базу (ILOterm), предметную классификацию и два тезауруса.
18. Портал лексикографических ресурсов НАТО: <https://www.lexicool.com/online-dictionary.asp?FSP=C04&FKW=nato>. Содержит 36 ЛИР.
19. TERMIUM Plus: <https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=eng>. Один из крупнейших в мире банков терминологических и лингвистических данных, предоставляет доступ к миллиону терминов на английском, французском, испанском и португальском языках.
20. Терминология ОЭСР: <https://www.oecd-ilibrary.org/glossaries>. Содержит 15 ЛИР.

Приложение 15.

Мировые и зарубежные каталоги ЛИР

1. Каталог лингвистических и терминологических ресурсов TERMSCIENCES: <http://www.termosciences.fr/termsciences/spip.php?rubrique23&lang=en>
2. Терминологические органы, ассоциации и сети: <https://termcoord.eu/terminology-bodies-networks/>
3. Глоссарии ЕС: https://termcoord.eu/discover__trashed/glossaries-by-eu-institutions-and-bodies/. Коллекция из 223 многоязычных глоссариев с нескольких веб-сайтов европейских учреждений и агентств.
4. Каталог ELRA: <http://catalogue.elra.info/en-us/>. Каталог Европейской ассоциации лингвистических ресурсов
5. Glossary Links: <https://termcoord.eu/glossarylinks/>. Инструмент поиска глоссариев с базой данных более 8000 глоссариев и постоянно растущий. Все ссылки в поиске глоссария регулярно проверяются и обновляются TermCoord с точки зрения релевантности и надежности.
6. Каталог OLAC: <http://search.language-archives.org/index.html>. Каталог, разработанный Сообществом открытых языковых архивов (OLAC), предоставляет доступ к обширной информации о тысячах языков, включая сведения о тек-

стовых коллекциях, аудиозаписи, словари и программное обеспечение, полученные из десятков цифровых и традиционных архивов.

7. Метаиндекс лингвистики, естественного языка и компьютерной лингвистики Стэнфордского университета: <https://nlp.stanford.edu/links/linguistics.html>
8. Каталог депозитариев лингвистической информации: <https://www.re3data.org/search?query=&subjects%5B%5D=104%20Linguistics>
9. Цифровая историческая лингвистика: <https://zenodo.org/communities/dighl/?page=1&size=20>. Всего 124 ресурса.
10. Ресурсы по лингвистике: <https://guides.lib.umich.edu/c.php?g=282869&p=1884907>. Электронные ресурсы для поддержки преподавания и исследований в области лингвистики. Включены указатели на статьи в научных журналах, полнотекстовые онлайн-источники, энциклопедии, словари и др.
11. Ресурсы по лингвистике. Каталог Летнего института лингвистики: <https://www.sil.org/linguistics/topical-links>
12. Проект исчезающих языков (ELP): <http://www.endangeredlanguages.com/#/5/11.942/59.788/0/100000/0/low/mid/high/dormant/awakening/unknown>. Каталог ресурсов по исчезающим языкам. В настоящее время свыше 7 тыс. ссылок на ресурсы о 3,4 тыс. языков.
13. Справочник по языковым веб-сайтам (бывший сайт «Человеческие языки»): <http://ilovelanguages.com/>. 2400 ссылок на лучшие языковые ресурсы Интернета.
14. Проекты и исследовательские сайты по лингвистике: <https://linguistlist.org/sp/GetWRListings.cfm?WRABrev=Projects>. Международное лингвистическое сообщество онлайн. Лингвисты, словари, шрифты, изучение языков, языковые ссылки, курсы лингвистики и учебные программы, программное обеспечение, тексты, корпорации.
15. Каталог сайтов и веб-ресурсов ЮНЕСКО по языковому разнообразию и многоязычию в Интернете и языкам в опасности: <http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/linguistic-diversity-and-multilingualism-on-internet/atlas-of-languages-in-danger/websites-and-online-resources/>

Приложение 16.

Российские каталоги ЛИР

1. Навигатор информационных ресурсов по языкознанию: <http://nirya2.alexo.beget.tech>
2. Каталог ресурсов для обработки естественного языка: <https://nlpub.mipt.ru/>
3. Портал знаний по компьютерной лингвистике: <https://uniserv.iis.nsk.su/cl/>
4. Каталог лингвистических программ и ресурсов в Сети: <https://rvb.ru/soft/catalogue/catalogue.html>
5. Информационный портал «Речевые технологии» (распознавание речи, синтез речи): <http://speech-soft.ru/>
6. Компьютерная лингвистика. Научно-образовательный портал «Лингвистика в России: ресурсы для исследователей»: http://uisrussia.msu.ru/linguist/_B_comput_ling.jsp
7. Каталог лингвистических программ и ресурсов в Сети / Linguistics Software: <http://rykov-ling.narod.ru/resurs.html>

8. Лингвистика. Каталог научных сайтов: <https://elementy.ru/catalog/t123/Lingvistika>
9. Поиск по лингвистическим сайтам (источники): <http://ruthenia.ru/tiutcheviana/search/sites/linglist.html>
10. Электронные ресурсы по лингвистике: <http://www.ling-expert.ru/links.html>

Приложение 17.

Рубрикатор по языкознанию НИРЯЗ

16. ЯЗЫКОЗНАНИЕ

- 16.01 Общие вопросы языкознания
 - 16.01.07 Философия языка. Философия и методология языкознания
 - 16.01.09 История языкознания. Персоналия
 - 16.01.13 Научные общества, съезды, конгрессы, конференции, симпозиумы по языкознанию
 - 16.01.17 Международное сотрудничество в области языкознания
 - 16.01.21 Организация научно-исследовательской работы в языкознании
 - 16.01.29 Информационная деятельность в области языкознания
 - 16.01.33 Терминология языкознания
 - 16.01.45 Преподавание языкознания
 - 16.01.51 Языковая политика. Государственные языки. Законы о языке
- 16.07 Методы лингвистических исследований
 - 16.07.01 Диахронические исследования. Образование и развитие языков
 - 16.07.03 Сравнительно-историческое языкознание
 - 16.07.05 Структурная лингвистика
 - 16.07.07 Математическая лингвистика. Формализованные языки. Генеративная лингвистика
 - 16.07.09 Полевая лингвистика
 - 16.07.11 Дескриптивная лингвистика
 - 16.07.13 Лингвистическая типология. Классификация языков
 - 16.07.15 Контрастивная лингвистика
 - 16.07.17 Когнитивная лингвистика
- 16.21 Общее языкознание
 - 16.21.07 Общетеоретические проблемы языкознания. Семиотика. Язык и речь. Невербальная коммуникация
 - 16.21.21 Прагматика. Теория речевых актов
 - 16.21.27 Социолингвистика
 - 16.21.29 Многоязычие и билингвизм
 - 16.21.31 Интерлингвистика
 - 16.21.33 Лингвистика текста. Анализ дискурса
 - 16.21.36 Этнолингвистика
 - 16.21.38 Психолингвистика
 - 16.21.42 Нейролингвистика

16.21.51 Семантика
 16.21.55 Стилистика. Лингвистическая поэтика. Риторика. Функциональные стили
 16.21.61 Культура речи и языковая норма
 16.21.63 Диалектология и лингвистическая география
 16.21.65 Лексикография. Словари

16.25 Уровни языка

16.25.11 Язык и письмо. Графика. Орфография. Пунктуация
 16.25.13 Фонология. Фонетика. Акцентология
 16.25.15 Грамматика
 16.25.19 Морфонология. Словообразование. Морфология.
 16.25.21 Синтаксис
 16.25.23 Лексикология. Терминоведение
 16.25.25 Ономастика. Топонимика
 16.25.27 Этимология
 16.25.29 Фразеология

16.31 Прикладное языкознание

16.31.21 Морфологический, синтаксический автоматический анализ текста
 16.31.25 Лингвостатистика. Статистическая обработка текста
 16.31.27 Автоматический (машинный) перевод
 16.31.29 Анализ и синтез устной речи
 16.31.31 Информационные языки. Лингвистические вопросы информатики
 16.31.33 Лингвистические вопросы искусственного интеллекта
 16.31.35 Компьютерная лингвистика
 16.31.37 Корпусная лингвистика
 16.31.41 Лингвистические вопросы перевода
 16.31.51 Лингвистические вопросы преподавания языков. Лингводидактика
 16.31.53 Лингвистическая экспертиза. Определение авторства текста
 16.31.61 Применение лингвистики в других отраслях

Приложение 18.

Таблица перекодировки кодов ГРНТИ в Рубрикатор НИР

Код действующего ГРНТИ	Наименование рубрики ГРНТИ	Тип изменения
1	2	3
16	Языкознание	Без изменения
16.01	Общие вопросы языкознания	Без изменения
16.01.07	Философские вопросы, методология и методы исследования в языкознании	Без изменения
16.01.09	История языкознания. Персоналия	Без изменения
16.01.11	Современное состояние и перспективы развития языкознания	Исключить коды
16.01.13	Научные общества, съезды, конгрессы, конференции, симпозиумы по языкознанию	Без изменения
16.01.17	Международное сотрудничество в области языкознания	Без изменения

Продолжение таблицы

1	2	3
16.01.21	Организация научно-исследовательской работы в языкознании	Без изменения
16.01.29	Информационная деятельность в области языкознания	Без изменения
16.01.33	Терминология языкознания	Без изменения
16.01.45	Преподавание языкознания	Без изменения
16.21	Общее языкознание	Без изменения
16.21.07	Общетеоретические проблемы языкознания	Без изменения
16.21.21	Методы лингвистических исследований	16.07
16.21.23	Классификация языков	16.07.13
16.21.25	Образование и развитие языков	16.07.01
16.21.27	Социолингвистика	Без изменения
16.21.29	Психолингвистика	16.21.38
16.21.31	Язык и письмо	16.21.11
16.21.33	Лингвистика текста	Без изменения
16.21.35	Уровни языка. Язык как система	16.25
16.21.37	Фонология. Фонетика. Акцентология	16.25.13
16.21.39	Общие проблемы грамматики	16.25.15
16.21.41	Морфология	16.25.19
16.21.43	Синтаксис	16.25.21
16.21.45	Словообразование	16.25.19
16.21.47	Лексикология. Терминоведение	16.25.23
16.21.49	Фразеология	16.25.29
16.21.51	Семантика	Без изменения
16.21.55	Стилистика. Лингвистическая поэтика. Риторика	Без изменения
16.21.61	Культура речи и языковая норма	Без изменения
16.21.63	Диалектология и лингвистическая география	Без изменения
16.21.65	Лексикография	Без изменения
16.31	Прикладное языкознание	Без изменения
16.31.02	Общие проблемы прикладного языкознания	16.31
16.31.21	Автоматическая обработка текста. Автоматический перевод. Автоматическое распознавание речи	Исключить коды
16.31.31	Информационные и формализованные языки	Без изменения
16.31.41	Лингвистические вопросы перевода	Без изменения
16.31.51	Лингвистические вопросы преподавания языков	Без изменения
16.31.61	Приложение языкознания к другим наукам	Без изменения
16.41 и нижестоящие		Исключить коды
39.29	Топонимика	16.25.25

Приложение 19. Указатель языков и языковых групп в НИРЯЗ

Отдельные языки

Аккадский
Алеутский
Английский

Арабский
Арамейский
Армянский
Башкирский
Болгарский
Бурятский
Велатру
Вепский
Водский
Гилянский
Гренландский
Греческий
Грузинский
Даргинский
Древнегреческий
Древнеегипетский
Древнерусский
Ижорский
Инуитов Аляски
Инуитов Канады
Иранский
Зырянский
Калмыцкий
Карельский
Керекский
Китайский
Коми
Коптский
Корейский
Латинский
Манде
Мазандеранский
Маньчжурский
Марийский
Монгольский
Мордовский
Нганасанский
Немецкий
Ненецкий
Олонхо
Осетинский
Пермский (пермяцкий)
Персидский
Пинежский диалект
Русский
Саамский
Самодийский

Санскрит
Сирийский
Согдийский
Тангутский
Татарский
Тибетский
Уйгурский
Чеченский
Чжурчжэньский
Чувашский
Чукотский
Шамерзади
Церковнославянский
Эвенский
Эфиопский
Юкагирский
Язык Командорских островов
Якутский
Японский

Генеалогические (генетические) группы

Балто-славянские
Иберо-романские
Индоевропейские
Иранские
Кавказские
Коми
Монгольские
Прибалтийско-финские
Самодийские
Славянские
Тюркские
Тюркские миноритарные
Угорские
Уральские
Финно-угорские
Чукотско-камчатские
Эскимосско-алеутские
Южнославянские
Юпикские языки Аляски

Ареальные и другие группы языков

Арабиграфические языки
Восточные языки
Древние языки
Европейские языки
Индийские языки

Ингерманландские языки
Классические языки
Мертвые языки
Центральноазиатские языки
Языки Африки
Языки Западной Африки
Языки коренных народов Арктики
Языки коренных народов Якутии
Языки народов Дагестана
Языки народов России
Языки Северной Осетии

Приложение 20.

Техническое задание на формирование Онтологии поисковых терминов по лингвистике

Онтология поисковых терминов по лингвистике (ОПТЕЛ) – информационный ресурс, создаваемый в рамках исследований по гранту РФФИ 18-00-00298 КОМФИ «Интеграция научно-информационных ресурсов учреждений РАН по гуманитарным наукам (на примере языкознания) как части единого цифрового информационного пространства РАН».

При разработке ОПТЕЛ должен использоваться опыт создания онтологической модели, разработанной в рамках предыдущего гранта.

Цель проекта

Практической первоочередной целью создания ОПТЕЛ является организация поиска по совокупности разнородных информационных ресурсов по лингвистике, которые будут интегрированы и / или агрегированы в рамках Центра лингвистических ресурсов (ЛИР) РАН.

Перспективной целью ОПТЕЛ является создание основы для полноценной онтологии по лингвистике (языкознанию), которая должна представлять собой концептуальную схему знаний в области лингвистики как часть Единого российского электронного пространства знаний или Единого цифрового пространства научных знаний.

Состав ОПТЕЛ

ОПТЕЛ должна включать лексику и парадигматику, а также дополнительные сведения, представленные в информационно-поисковых языках и системах метаданных, используемых в информационных ресурсах по языкознанию, которые используются в модели Центра ЛИР. К этим ИПЯ и системам метаданных относятся:

- информационно-поисковый тезаурус по языкознанию ИНИОН РАН;
- рубрикатор по языкознанию ИНИОН РАН;
- словари библиографических данных БД по языкознанию ИНИОН РАН;
- государственный рубрикатор научно-технической информации;
- универсальная десятичная классификация;
- библиотечно-библиографическая классификация;

- теги разметки (признаки и значения признаков) комплекса корпусов, входящих в состав Национального корпуса русского языка;
- метаданные (реквизиты и значения реквизитов) лексикографических БД, формируемых в ИРЯ РАН;
- реестры языков.

Данный список не является исчерпывающим и закрытым, он должен пополняться по мере развития Центра ЛИР.

На первом этапе в состав ОПТЕЛ включается лексика ИПЯ, создаваемых интеллектуально. Лексика индексов, словариков, глоссариев, онтологий и других словарей ИПЯ, автоматически создаваемых на основе обработки текста, в ОПТЕЛ не включается.

Общая структура ОПТЕЛ

ОПТЕЛ включает две сущности: лексические единицы (ЛЕ) и источники ЛЕ (ИСТ). Каждая ЛЕ входит в набор ИСТ. ИСТ представляет собой адресуемый и поименованный массив данных.

ЛЕ и ИСТ имеют собственные наборы реквизитов.

Структура сведений о ЛЕ:

- ЛЕ;
- набор идентификаторов ИСТ;

(Все следующие реквизиты относятся к одному ИСТ; для каждого ИСТ сведения повторяются.)

- тип ЛЕ (задается списком – дескриптор, аскриптор, рубрика, определитель, элемент авторитетного файла, тег разметки, признак, значение признака и т.д.);
- поле (множество полей) БД, для поиска по которому используется ЛЕ;
- синонимы, члены синсетов, аскрипторы;
- пометы, реляторы для ЛЕ;
- кодовая форма ЛЕ (например, для классификаторов);
- сокращенная форма ЛЕ;
- эквивалент ЛЕ на иностранных языках (набор пар: язык – перевод ЛЕ);
- вышестоящие ЛЕ (включая тематические поля, рубрики, признаки для значений признаков и др.);
- список нижестоящих ЛЕ;
- список ассоциативно связанных ЛЕ (вкл. ссылку «см. также»);
- другие связи ЛЕ (набор пар: вид связи – ЛЕ);
- дефиниции ЛЕ.

Структура сведений об ИСТ:

- наименование ИСТ;
- идентификатор ИСТ;
- тип ИСТ (задается списком – тезаурус, рубрикатор, БД, корпус, другой);
- адрес ИСТ в Интернете.

Для иерархически организованных ресурсов:

- ИСТ, являющиеся частью данного ИСТ;
- ИСТ, включающий данный ИСТ.

Функции ОПТЕЛ

Навигация:

- просмотр ЛЕ по алфавиту с возможностью просмотра всей информации о ЛЕ;
- просмотр перечня наименований ИСТ с возможностью просмотра всего ИСТ.

Поиск

(Индекс формируется только по полю ЛЕ.)

Поиск ЛЕ на полное совпадение или на совпадение фрагмента ЛЕ с выдачей:

- всей информации о ЛЕ;
- списка ИСТ, приписанных данной ЛЕ;
- всей информации о ЛЕ, имеющейся в данном ИСТ;
- информации о ЛЕ, имеющейся в данном поле данного ИСТ;
- информацию о ЛЕ, имеющейся в данном поле всех ИСТ.

При просмотре показывать только заполненные поля ЛЕ.

Редактирование

Включение новых ЛЕ с ручным заполнением всех или некоторых полей:

- добавление значений ЛЕ в данное поле
- удаление ИСТ;
- удаление ЛЕ;
- удаление значения любого поля ЛЕ.

Пакетная загрузка новых ИСТ в установленном формате.

Добавление новых полей (реквизитов).

Сохранение истории изменений.

А.Б. Антопольский

**НАУЧНАЯ ИНФОРМАЦИЯ
И ЭЛЕКТРОННОЕ ПРОСТРАНСТВО ЗНАНИЙ**

Под научной редакцией Д.В. Ефременко

Оформление обложки И.А. Михеев
Техническое редактирование
и компьютерная верстка К.Л. Синякова
Корректор Я.А. Кузьменко

Гигиеническое заключение

№ 77.99.6.953.П.5008.8.99 от 23.08.1999 г.

Подписано к печати 24 / XI – 2020 г.

Формат 60х84/16 Бум. офсетная № 1 Печать офсетная

Усл. печ. л. 16,25 Уч.-изд. л. 11,5

Тираж 300 (1–100 экз. – 1-й завод) Заказ № 98

**Институт научной информации
по общественным наукам Российской академии наук (ИНИОН РАН)**

Нахимовский проспект, д. 51/21, Москва, 117418

<http://inion.ru>, https://instagram.com/books_inion

**Отдел маркетинга и распространения
информационных изданий**

Тел.: +7 (925) 517-36-91, +7 (499) 134-03-96

e-mail: shop@inion.ru

Отпечатано по гранкам ИНИОН РАН

ООО «Амирит»,

410004, Саратовская обл., г. Саратов,

ул. Чернышевского, д. 88, литера У