Министерство науки и высшего образования Российской Федерации
**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ ГОСУДАРСТВЕННАЯ ПУБЛИЧНАЯ НАУЧНО-ТЕХНИЧЕСКАЯ БИБЛИОТЕКА СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК (ГПНТБ СО РАН)**

УДК 001.3
Рег. № НИОКР АААА-А19-119121090008-8
Рег. № ИКРБС

УТВЕРЖДАЮ
директор ГПНТБ СО РАН
Канд. техн. наук А. Е. Гуськов
«17» января 2020 г.

# ОТЧЕТ
# О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

**Наукометрический анализ публикационного потока российских исследователей и факторов его трансформации, изучение способов и методов повышения публикационной активности, развития российской научной периодики**
(промежуточный за 2019 г.)

Номер темы в ИП «Система управления НИР»
Министерства науки и высшего образования Российской Федерации 0334-2019-0006

**Приоритетное направление** IV.38. Проблемы создания глобальных и интегрированных информационно-телекоммуникационных систем и сетей, развитие технологий и стандартов GRID

Руководитель НИР
директор ГПНТБ СО РАН
канд. техн. наук А. Е. Гуськов

Новосибирск, 2020

## СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель

кадн. техн. наук, директор 17.01.2020 _____ А.Е. Гуськов

Исполнители

м.н.с. 17.01.2020 _____ А.В. Багирова

канд. пед. наук, с.н.с. 17.01.2020 _____ В.Н. Гуреев

н.с. 17.01.2020 _____ Д.В. Косяков

канд. техн. наук, в.н.с. 17.01.2020 _____ Н.А. Мазов

м.н.с. 17.01.2020 _____ И.В. Селиванова

м.н.с. 17.01.2020 _____ Л.Л. Садовская

доктор ист. наук, с.н.с. 17.01.2020 _____ Ю.П. Холюшкин

## РЕФЕРАТ

**Отчет промежуточный** 93 с., 26 рисунков, 23 таблицы, 38 источников, 1 приложение.

БИБЛИОМЕТРИЯ, НАУКОМЕТРИЯ, НАУЧНЫЕ ЖУРНАЛЫ, НАУЧНАЯ ПЕРИОДИКА, РЕДАКЦИОННАЯ КОЛЛЕГИЯ, ПУБЛИКАЦИОННАЯ АКТИВНОСТЬ, ТЕМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ, SCOPUS

**Целью исследований** являлось улучшение качества наукометрических данных; выявление стратегий и драйверов повышения публикационной активности российских исследователей; исследование критериев успешности и авторитета российской научной периодики, основанное на изучении редакционной коллегии журналов.

**Важнейшие результаты.** Исследование географического распределения и публикационной активности членов редколлегий журналов – новый наукометрический подход к оценке периодики и соответствующих отраслей науки. Анализ российских журналов по информационно-библиотечным наукам показал: расположение крупных экспертных сообществ в Москве, Санкт-Петербурге, Новосибирске, а также Самаре и Воронеже; относительно низкую долю и формальный характер участия зарубежных представителей в работе журналов; сосредоточение членов редколлегий в вузах (47 %); существенное влияние публикационной активности членов редколлегий на рейтинговые показатели журнала; отдельные случаи нарушения публикационной этики членами редколлегий для повышения рейтинга журналов; размытость ядра журналов и его смещение в область изданий с невысокими библиометрическими показателями.

Полученные данные позволили не только отразить с новых позиций состояние соответствующих дисциплин, но и сформировали методологическую базу для планируемой разработки нового библиометрического показателя оценки качества журналов на основе работы редколлегии.

**Наиболее значимые результаты:**

1. Исследована возможность классификации научных текстов на основе метода компрессии их аннотаций. Идея метода состоит в том, что алгоритмы компрессии типа PPM (Prediction by partial matching) сжимают терминологически близкие тексты существенно лучше, чем далекие. Выявлено, что построение обучающего ядра на основе высокоцитируемых публикаций уменьшает количество ошибок классификации до 12% против 32% при случайной выборке. На качество классификации влияет изначальное количество категорий: чем меньше категорий и чем больше терминологическое различие между ними, тем выше ее качество.

2. На основе анализа детализированных библиографических записей о публикациях российских исследователей за 2000-2018 гг. в базе данных Scopus показано, что ошибки идентификации приводят к наличию дублирующих профилей у 77% организаций и 24% авторов. Разобраны причины образования дублирующих профилей. Выполнена оценка влияния этих ошибок на количественные показатели результативности организаций, в среднем теряющих около 17% публикаций; отдельных авторов, для них потери составляют 11%. Эти результаты позволяют учитывать погрешность в оценке результативности организаций и отдельных авторов. Детальный разбор причин образования ошибок, учитывающий российскую специфику доведен до сведения специалистов Elsevier и может послужить повышению качества и достоверности библиографической информации в этой базе данных.

3. Показано, что в основе роста российских публикаций лежит расширение количества публикующихся (активных) авторов, сопровождаемые изменениями в секторальном и географическом распределении таких авторов. Основной вклад внесен расширением количества научно-педагогических работников вузов, начавших публиковать результаты своих исследований в изданиях, индексируемых международными БД. В результате в 2017 г. на них пришлось более 58% вклада в публикационный поток российских ученых, рассчитанного методом фракционного счета. На фоне увеличения количества публикаций растет также роль регионов, прежде всего, Западной Сибири, Урала, Поволжья.

В 2019 г. было опубликовано 10 статей в рецензируемых научных журналах, в том числе 4 – в журналах, индексируемых в Web of Science Core Collection и Scopus, 5 статей в сборниках материалов конференций, в том числе 4 – в сборниках, индексируемых в Scopus.

# СОДЕРЖАНИЕ

Введение ..... 7

1 Разработка технологии улучшения библиографических данных публикаций БД Scopus с устранением неоднозначностей и ошибок, связанных с некорректной атрибуцией авторов и аффилиаций. ..... 10

1.1 Типизация ошибок идентификации в базе данных Scopus и их влияние на оценку результативности научной деятельности ..... 10

1.2 Классификация научных текстов на основе компрессии аннотаций публикаций (на примере Scopus) ..... 18

1.3 Позиционирование работ по улучшению библиографических баз данных и проведения наукометрических исследований в деятельности научных библиотек. ..28

2 Разработка методов структурно-динамического анализа публикационного потока, выполнение анализа полученного массива данных с выявлением изменений в секторальном, тематическим и территориальном аспектах, обнаружение и сопоставление эволюционных и аномальных изменений в публикационном потоке с влиянием стимулирующих факторов. ..... 32

2.1 Данные о публикационном потоке ..... 32

2.2 Особенности национальной научной политики ..... 33

2.3 Взрывной рост количества публикаций российских ученых ..... 34

2.4 Секторальные и региональные диспропорции роста публикационной активности ...37

2.5 Низкое качество данных государственной статистики по кадровому потенциалу российской науки ..... 40

2.6 Наукометрический подход ..... 48

2.7 Аберрации в количественных показателях публикационной активности и их причины ..... 50

2.8 Международная синхронная мобильность и ее влияние на результативность российских исследователей ..... 53

2.9 Национальная оценка результативности научных организаций – проблемы и недостатки ..... 60

2.10 Обоснование необходимости перехода на фракционный счет при оценке публикационной активности организаций ..... 63

3 Анализ территориального распределения и библиометрическое исследование публикационной активности членов редколлегий научных журналов в области информационных и библиотечных наук, истории и археологии. ..... 68

3.1 Анализ территориального распределения и библиометрическое исследование публикационной активности членов редколлегий научных журналов в области информатики, библиотековедения и наук о Земле ..... 68

Заключение ..... 85

Список использованной литературы ..... 89

Приложение. Список публикаций по теме ..... 92

# ВВЕДЕНИЕ

В последние 10-15 лет в России выстраивается новая система государственного управления сектором исследований и разработок, основанная на активном применении наукометрических индикаторов. Требования к показателям публикационной активности присутствуют в структуре государственных заданий научных организаций и конкурсной документации научных фондов, а также являются предметом регулярных дискуссий относительно путей развития российской науки. Сейчас можно констатировать, что реформа фундаментальной науки внесла значительные изменения не только в институциональную структуру управления исследованиями, но и в систему мотивации ученых. В конечном итоге это привело к изменению объема и содержания российского «валового научного продукта».

Одной из главных целей реформирования является увеличение количества публикаций российских авторов в ведущих мировых научных журналах, индексируемых в Web of Science или Scopus. С 2013 года действительно можно наблюдать рост доли российских публикаций в мировом потоке, который за несколько лет поднялся с 1.8% до 2.3-2.6% (по разным оценкам). Вместе с тем, фиксируются многочисленные (а, иногда, и системные) негативные эффекты такой системы мотивации: увеличивается доля публикаций в хищнических журналах; расцветает рынок откровенно мусорных конференций и изданий; одной публикацией могут отчитываться много раз в разных проектах и организациях; в погоне за количеством падает качество научных статей; статус монографии (завершенный научный труд) долгое время остается неопределенным.

При этом за последние годы проведен ряд реформ, имевших целью улучшить российский научный ландшафт: запуск ФЦП "Исследования и разработки", Проект повышения конкурентоспособности ведущих российских университетов среди ведущих мировых научно-образовательных центров, создание Российского научного фонда, поддержка "мегагрантов", реформа государственных академий наук. Однако системный анализ фактического влияния этих изменений не проводился, что не позволяет в настоящее время уверенно говорить об их успешности, а в будущем - обоснованно вводить новые научные политики с прогнозируемыми результатами.

Этот проект посвящен изучению фактического влияния перечисленных событий на индикаторы российской науки, причем речь идет не только о количественных, но и о качественных показателей. Первым объектом изучения является массив публикаций российских ученых с 2000 года в международной базе данных Scopus, который содержит основные результаты фундаментальных исследований, признанных на мировом уровне.

Вторым объектом изучения является сама российская научная периодика, состояние которой также является характерным отражением положения российской фундаментальной науки.

На 2019 г. поставлены три задачи:

1. Разработка технологии улучшения библиографических данных публикаций БД Scopus с устранением неоднозначностей и ошибок, связанных с некорректной атрибуцией авторов и аффилиаций.

Предварительный анализ показал, что библиографические данные БД Scopus содержат большое количество неоднозначностей и ошибок, связанных с некорректной атрибуцией авторов и организаций несмотря на систему персональных и институциональных профилей, поддерживаемых в этой БД. Результаты наукометрического анализа с использованием этих данных могут быть значительно искажены, поэтому требуется оценка величины этих искажений, а также разработка методики устранения неоднозначностей, сводящей количество ошибок до приемлемого уровня. В общем виде эта проблема пока не решена. Решение проблемы потребует обогащения данных из БД косвенной информацией о российских научных организациях и вузах и ученых из других источников.

2. Разработка методов структурно-динамического анализа публикационного потока, выполнение анализа полученного массива данных с выявлением изменений в секторальном, тематическим и территориальном аспектах, обнаружение и сопоставление эволюционных и аномальных изменений в публикационном потоке с влиянием стимулирующих факторов.

Разные темпы и характер изменений объемов и качества публикационного потока в разных секторах, отраслях науки и географических регионах часто не позволяют выявить аномалии роста. Необходим многомерный анализ динамики объемных показателей публикационного потока. В связи с большим количеством измерений требуется разработка аналитических инструментов и методов для полуавтоматического анализа.

3. Анализ территориального распределения и библиометрическое исследование публикационной активности членов редколлегий научных журналов в области информационных и библиотечных наук, истории и археологии.

Как было показано в ряде работ, наукометрический анализ публикационной активности членов редколлегии позволяет выполнить оценку качества и перспектив научного журнала. Выполнение этого анализа по авторской методике потребует сбора базы данных членов редколлегий российских журналов в выбранных отраслях знаний, их

научных публикаций на основании сайтов журналов и библиографических баз данных Web of Science, Scopus и РИНЦ.

Исследование опирается на библиометрические и наукометрические методы количественного и качественного анализа данных расширенных библиографических описаний публикаций российских ученых, данных государственной статистики, государственных и ведомственных информационных систем.

По результатам исследования в 2019 г. было опубликовано 10 статей в рецензируемых научных журналах, в том числе 4 – в журналах, индексируемых в Web of Science Core Collection и Scopus, 5 статей в сборниках материалов конференций, в том числе 4 – в сборниках, индексируемых в Scopus.

# 1 РАЗРАБОТКА ТЕХНОЛОГИИ УЛУЧШЕНИЯ БИБЛИОГРАФИЧЕСКИХ ДАННЫХ ПУБЛИКАЦИЙ БД SCOPUS С УСТРАНЕНИЕМ НЕОДНОЗНАЧНОСТЕЙ И ОШИБОК, СВЯЗАННЫХ С НЕКОРРЕКТНОЙ АТРИБУЦИЕЙ АВТОРОВ И АФФИЛИАЦИЙ.

## 1.1 Типизация ошибок идентификации в базе данных Scopus и их влияние на оценку результативности научной деятельности

Библиографические базы данных (ББД) в настоящее время активно используются не только в задачах научного поиска, но и при оценке результативности научной деятельности как отдельных сотрудников, так и целых организаций. Наукометрические показатели, получаемые из ББД, рассматриваются как истина в последней инстанции, но при работе с ними не учитываются ошибки, влияющие на эти оценки.

Качество ББД изучалось в некоторых зарубежных работах [1–3]. В статьях F. Franceschini, D. Maisano, L. Mastrogiacomo, посвященных выявлению основных ошибок, встречающихся в *Web of Science* и *Scopus*, выделено два типа подобных ошибок [4,5]: ошибки, сделанные авторами / издателями / редакторами при подготовке списка литературы, и ошибки, возникшие в БД при связывании (т.е. установлении идентифицирующих связей между объектами в одной или разных системах, например, публикация-автор, автор-аффилиация) статей. В работе [6] рассматриваются ошибки в написании имен итальянских авторов в четырех базах данных: Web of Science, Scopus, PubMed, CrossRef.

В работе [7] оценивается общее влияние качества массива данных на проведение библиометрического анализа при составлении авторских рейтингов и регрессионном анализе, где в качестве зависимых переменных выступает число публикаций. Результаты исследования показали, что ранжирование авторов можно проводить только при очень высоком качестве данных. Хотя авторы рассматривали только несоответствие имен автора, они приходят к выводу, что и ошибки в названиях организаций тоже будут оказывать очень большое влияние при составлении различных рейтингов (например, рейтингов университетов). Для регрессионного же анализа с наборами библиометрических данных должны проверяться систематические различия в ошибках, которые коррелируют с независимыми переменными (например, если использовать в качестве независимой переменной страну) и которые могут влиять на достоверность результатов регрессии.

Несмотря на схожую природу, ошибки в русскоязычных статьях имеют свою специфику, которая до сих пор никем систематически не изучалась. Она связана как с транслитерацией букв кириллического алфавита, не имеющих аналогов в латинице (например, *е, ю, ш, щ, ъ, ы, ь*), так и с иерархией подчиненности научных организаций в России (имеется в виду принадлежность организаций к государственным академиям наук РФ, либо к их региональным отделениям), а также с ее изменениями в последние годы. Эти ошибки служат причиной неверной привязки публикаций, что в свою очередь ведет к появлению профилей-дублей авторов и организаций.

С целью однозначного определения фамилий авторов предложена система уникальных идентификаторов *ORCID*, в последние годы широко признанная научным сообществом и принятая многими издателями научных журналов. Ранее компанией *Thomson Reuters* был предложен универсальный идентификатор автора *ResearcherID*, который не получил широкого признания в связи с тесной связкой с базой данных *WoS*. Компания *Elsevier* в системе *Scopus* использует уникальные идентификаторы авторов *Scopus AuthorID*. Все эти идентификаторы часто применяются в различных информационно-аналитических системах научных институтов при интеграции данных о публикационной активности их сотрудников, но при связывании данных все равно отмечается необходимость их визуальной проверки. К сожалению, сами авторы плохо отслеживают свои публикации, что приводит к неполноте их профилей в ББД, и, соответственно, к некорректным результатам связывания при использовании систем идентификации, что подтверждает актуальность этой задачи.

Для нашего анализа была выбрана БД *Scopus*, в которой у каждого автора и организации может быть свой профиль с идентификатором. Данные о публикациях российских авторов были получены с помощью *Scopus Search API* и *Scopus Abstract Retrieval API* в 2017 г.

Наличие профиля-дубля у российских организаций проверялось на случайной выборке из 400 профилей. Это количество было получено из формулы В.И. Паниотто с тем условием, что доверительная вероятность P=0,954. Для указанной доверительной вероятности коэффициент доверия $t = 2$, дисперсия качественного альтернативного признака принимается максимально возможной, т.е. соответствующей доле 0,5:

$$n = \frac{t^2 p(1-p)}{\Delta^2} = \frac{2^2 \cdot 0{,}5(1-0{,}5)}{\Delta^2} = \frac{1}{\Delta^2} = \frac{1}{0{,}05^2} = 400, \qquad (1)$$

где $n$ – объем выборки; $p$ – генеральная доля единиц, обладающих значением признака, относительно которого рассчитывается ошибка выборочной доли; $t$ – доверительный коэффициент; $\Delta$ – предельная ошибка выборки.

Для того чтобы анализ был более точным, введено ограничение на количество публикаций в профиле организации: в 2017 г. их должно быть не менее 5. При меньшем количестве публикаций в список попадали организации, реальное название которых сложно было определить. Например, организация с *Affiliation ID*=119911371 называлась *Berdsk*, получившееся из-за перехода в название части адреса. Также были обнаружены явные дубли других организаций. В качестве примера приведем организацию с *Affiliation ID*= 109886056, в названии которой указано *Romsk Polytechnical University*. Можно предположить, что этот профиль, скорее всего, является дублем профиля Томского политехнического университета (ТПУ). Подтвердить это предположение невозможно, так как в этом профиле только одна публикация, в аффилиациях авторов которой указана только эта организация. В основном профиле ТПУ этих авторов обнаружить не удалось.

*Алгоритм анализа профилей организаций:*

1. Для организации формируем список из 10 аффилированных с ней авторов с наибольшим числом публикаций: $A_i = \{a_1, a_2, \ldots, a_M\}$, где M = 10.
2. Для каждого автора $a_i$, i=1,2, …,M формируем полный список его аффилиаций $OA_i = \{oa_{i1}, oa_{i2}, \ldots\}$.
3. Формируем общий список аффилиаций 10 самых продуктивных авторов из организации $GL = \bigcup OAi$ .
4. В списке *GL* выявляем дубликаты – переводы названий, сокращенные и схожие названия.

Алгоритм применялся для каждой из 400 организаций и позволил выявить дублирующие профили аффилиаций. При сопоставлении названий дополнительно изучались сайт организации, «Википедия», а также профиль организации в Научной электронной библиотеке eLibrary.ru.

Наличие профиля-дубля у автора проверялось по формуле (1) на случайной выборке из 400 активных российских авторов (не менее двух публикаций в 2017 г.). Ограничение на количество публикаций было введено из-за того, что при отсутствии этого фильтра в случайную выборку попадало слишком много профилей авторов с одной публикацией, что не позволяло проанализировать наличие у этих авторов дублирующих профилей. В списках российских авторов из Scopus было обнаружено около 47% таких профилей.

*Алгоритм поиска профиля-дубля автора:*

Обозначим множество авторов за $A = \{a_1, a_2, \ldots, a_N\}$, где $N = 400$.

1. Для каждого автора $a_i$, $i$=1, …, N формируем полный список аффилиаций $OA_i = \{oa_{i1}, oa_{i2}, \ldots\}$, т. е. для каждого из авторов – полный список организаций, о которых есть упоминание в его публикациях.

2. Для каждой организации $oa_{ij}$ формируем полный список авторов $C_{ij} = \{c_{ij}^1, c_{ij}^2, \ldots\}$, т. е. авторов, работающих в одной организации $o_{ij}$ с автором $a_i$, включая и самого автора $a_i$.

3. Формируем общий список коллег автора $a_i$ из всех организаций $C_i = \bigcup_j C_{ij}$

4. Из списка $C_i$ находим все дубликаты – авторы с ФИО автора $a_i$, которое либо полностью совпадает с его ФИО, либо похоже на него. Анализ проводился вручную, и совпадения проверялись экспертным путем (при автоматизированном поиске необходимо использовать различные меры близости).

Для того чтобы исключить возможность определения профиля полного тезки автора как профиля-дубля исследуемого автора, дополнительно анализировались списки соавторов, тематики исследования, организаций, полные тексты статей, списки авторов на eLibrary.ru и аффилиационная история. Мы полагаем, что если у нескольких профилей с одинаковыми ФИО совпадают места работы, соавторы и тематика, то это профили одного и того же автора.

### 1.1.1 Оценка количества и определение причин возникновения профилей-дублей организаций

Анализ профилей организаций показал следующие результаты.

***A.*** 19% (75 из 400) организаций дублей не имеет.

***B.*** 5% (20 из 400) профилей организаций содержат ошибки (иностранные организации, разные адреса и др.), что затрудняет определение дублей.

***C.*** 76% (305 из 400) организаций имеют хотя бы один дубль.

Рассмотрим подробнее ошибки, обнаруженные в случаях B и C.

*Случай B1. Вместо названия организации указаны профили: РАН, СО РАН и др.*

Ошибка возникает из-за неверной привязки организации, в названии которой встречается, например, Сибирское отделение Российской академии наук (СО РАН). Такая организация может быть отнесена к профилю как СО РАН, так и РАН, при этом в последнем случае публикация будет приписана Москве, а не Новосибирску или другому сибирскому научному центру. Это может повлиять на результаты некоторых исследований, например, по академической мобильности ученых (*Affiliation ID* профиля СО РАН в *Scopus* – 60017604, РАН – 60021331).

Для оценки количества публикаций, которые может потерять организация из-за отнесения к профилям РАН, РАМН или их региональным отделениям, была сделана

случайная выборка 400 профилей российских авторов и посчитана доля публикаций для случая B1. Результаты показали, что из-за неверного отнесения публикаций авторов к государственным академиям наук РФ в среднем для одного автора организация теряет до 14% публикаций.

Случаи, когда автор действительно может указывать аффилиацию Российской академии наук или ее региональных отделений, являются редкими.

*Случай B2. Иностранные организации*

Ошибка возникает из-за неверного распознавания названия города. Например, университет Айдахо расположен в городе *Moscow*, штат Айдахо, США. Но в *Scopus* один из профилей университета относится к России (*Affiliation ID*=100804052, количество публикаций = 106). При этом в список из 400 организаций попал еще один профиль университета Айдахо, также отнесенный к российским организациям (*Affiliation ID*=110070192, количество публикаций = 73).

*Случай B3. Сложносоставные организации с разными подразделениями*

К данному случаю относятся как университеты, где авторы иногда указывают только название факультета, и такая публикация может попасть в другой профиль, так и организации, подразделения которых находятся в разных городах. В качестве примера приведем организацию с *Affiliation ID*=100465640. В ее названии указано *Chemistry Department*. При этом в адрес организации попало полное название *Lomonosov MSU*, *Moscow*.

В таблице 1 указаны причины возникновения дублей профилей российских организаций (результат C), а также частота их встречаемости на примере тех профилей, которые имеют только один дубль (общим количеством 66).

Данные в ББД *Scopus* активно обновляются и их качество растет, как в связи с усилиями специалистов *Elsevier*, так и с обращениями пользователей системы. Результаты нашего исследования показали, что из 66 дублирующихся профилей на начало 2017 г. почти 41% оказались объединенными к марту 2019 г.

Максимальное количество – 26 дублей обнаружено в *Scopus* в названии Федерального государственного бюджетного учреждения «Национальный медицинский исследовательский центр психиатрии и наркологии имени В.П. Сербского» Министерства здравоохранения Российской Федерации, сокращенно – ФГБУ «НМИЦПН им. В.П. Сербского» Минздрава России, старые названия которого: «Государственный научный центр социальной и судебной психиатрии им. В.П. Сербского», «Федеральный медицинский исследовательский центр психиатрии и наркологии имени В. П. Сербского.

**Таблица 1 – Анализ профилей с одним дублем**

| Случай | Причина возникновения дублей | Кол-во профилей | Доля в общем количестве, % |
|---|---|---|---|
| C1 | Старое название организации | 1 | 2 |
| C2 | Сокращенное название организации, включая<br>• частичную аббревиацию названия (например, вместо *the Siberian Branch of the RAS* указано *SB RAS*);<br>• сокращенное название, добавленное к полному | 13 | 20 |
| C3 | Отсутствие части названия | 17 | 26 |
| C4 | Отсутствие имени/части имени ученого, присвоенного организации, его инициалов | 3 | 5 |
| C5 | Попадание части названия организации в адрес | 2 | 3 |
| C6 | Неверный/неполный адрес организации (включая неверное указание страны или города) | 4 | 6 |
| C7 | Разный порядок слов в названии | 1 | 2 |
| C8 | Полное совпадение названий (например, у *Tver State Medical University* имелся дубль с таким же названием) | 4 | 6 |
| C9 | Разная транслитерация названия | 2 | 3 |
| C10 | Разный перевод названия (включая перевод не только на английский язык) | 9 | 14 |
| C11 | Ошибки в названии (опечатки, лишние пробелы и знаки препинания) | 10 | 15 |
| C12 | Зависимость от регистра | 0 | 0 |

### 1.1.2 Оценка количества и определение причин возникновения профилей-дублей авторов

Анализ профилей авторов показал следующие результаты.

A. У 75% (300 из 400) авторов не было обнаружено профилей-дублей.

B. У 1% (3 из 400) не удалось установить наличие дубля.

Это связано, например, с тем, что в профиле организации в *Scopus* встречается несколько авторов с одним и теми же фамилией и инициалами, но в профиле организации, как и на сайте института, такого сотрудника нет. В профилях этих авторов не совпадают также и соавторы. В другом случае имя и фамилия автора имеют вид Alexander R., т. е. определить фамилию этого автора становится невозможным.

C. У 24% (97 из 400) были обнаружены профили-дубли.

Необходимо отметить, что с начала 2017 г. к марту 2019 г. 9% (36 из 400) профилей авторов объединены в *Scopus* полностью, 2% (7 из 400) – частично, а 13,5% (54 из 400) профилей так и остаются необъединенными.

Рассмотрим причины и частоту возникновения 61 профиля-дубля из тех, которые полностью (54 профиля) или частично (7 профилей) не объединены в *Scopus*.

***Случай 1.*** *Разная транслитерация (41 из 61, 67%)*

Разная транслитерация фамилии, имени или отчества автора появляется из-за наличия в них букв *е, ю, я, х, ц, ш, ь* и др. (например, имя одного автора транслитерируется в одном случае как Yury, в другом – Yrii).

Примером автора, в дублирующих профилях которого наблюдается разная транслитерация и в имени, и в фамилии, является Bakaeva Natalia Vladimirovna (*AuthorID* = 57195920506). В ее профиле-дубле (*AuthorID* = 56826095700) имя указано как Bakayeva, Natalya. Для этого примера отметим еще один важный момент: обоим профилям дается один и тот же *ORCID*, но профили не объединены, т. е. при объединении профилей в *Scopus* механизм объединения по совпадающему *ORCID* либо отсутствует, либо срабатывает не всегда.

***Случай 2.*** *Ошибка в фамилии (10 из 61, 16%).*

- Лишние буквы, опечатки, ошибки распознавания, похожие по внешнему виду буквы (например, заглавная I и строчная l), строчные и заглавные буквы, ошибки в самой публикации.

  Одной из иллюстраций этого типа ошибок является профиль автора (основной *AuthorID* = 7202813119) Костромского государственного университета – Smirnova, N.A. Второй ее профиль (*AuthorID* = 56568452600) обозначен как Smimova, N. A., т.е. буквы rn в фамилии были распознаны как m. В ее третьем профиле (*AuthorID* = 55480627000) в качестве ФИО указано: SMlRNOVA, N. A., т.е. буква I была распознана как l.

  В другом примере ошибочное написание фамилии автора возникло уже в исходной публикации. Среди профилей авторов Научно-исследовательского института фармакологии им. В.В. Закусова были обнаружены два автора: Seredenin, Sergey B. (*AuthorID* = 7004680975) и Seredin, S. B. (*AuthorID* = 6504406983). При этом в списке авторов на сайте eLibrary.ru автора с фамилией Середин обнаружено не было. В полном тексте статьи из профиля автора Seredin, S. B. также была указана фамилия Середин. Через eLibrary.ru было выяснено, что статья является переводной версией

русскоязычной публикации, в которой автором указан Середенин С.Б. О том, что эта статья принадлежит именно Середенину, также свидетельствует наличие общих соавторов с основным профилем, список литературы, где практические все работы принадлежат Середенину.

- К окончанию фамилии приписана буква от аффилиации.

  Примером может послужить профиль автора из Института общей и неорганической химии имени Н.С. Курнакова РАН – Fedorchenko, Irina Valentinovna (*AuthorID* = 24474356600). У нее есть профиль-дубль с одной публикацией, где в имени профиля написано Fedorchenkob, I.V. Детальный анализ полного текста этой публикации показал, что *b* в конце фамилии появилась из-за обозначения этой буквой ссылки на организацию автора; в полном тексте публикации данной ошибки нет.

***Случай 3.*** *Разный вариант представления имени (8 из 61, 13%).*

- Вместо полного имени/отчества указаны инициалы; отчество отсутствует полностью.

  Рассмотрим профиль автора из Института общей физики имени А.М. Прохорова РАН Kozlov, D.N. (*AuthorID* = 55407907300). В профиле у автора 75 публикаций. Для этого ученого был найден еще один профиль: Kozlov, Dimitry N. (*AuthorID* = 26650432300) с одной публикацией. Соавторы и тематики исследований в профилях совпадают.

- Перепутаны местами имя, отчество и фамилия.

  У автора из Университета ИТМО – Leonov Mikhail Yu – два профиля: в первом (*AuthorID* = 36604774600) в качестве фамилии указано: Leonov, имени: Mikhail Yu; во втором (*AuthorID* = 57192376476) фамилия у автора уже Leonov Mikhail, в имени указаны лишь инициал отчества Yu. В основном профиле у автора 39 публикаций, в дубле – одна.

  Другим подобным примером является профиль автора из Института гидродинамики имени М.А. Лаврентьева СО РАН, в основном профиле которого (*AuthorID* = 6603685354) указаны фамилия: Liapidevskii, имя: Valery Yu. В одном из его профилей-дублей (*AuthorID* = 57196437043) в фамилию также вошел инициал от отчества: Yu Liapidevskii, а имя осталось верным: V.

***Случай 4.*** *Разные периоды публикаций (1 из 61, 2%).*

Единственным примером, относящимся к этому случаю, является профиль автора из Рязанского государственного медицинского университета. В одном профиле (*AuthorID* = 6505788719) охват публикаций с 1985 по 1993 гг., в другом (*AuthorID* = 57195313658) – с 2017 по 2018 гг. Профиль автора был найден также в eLibrary.ru, где охват публикаций с

1986 по 2018 гг. С профилем в eLibrary.ru совпадают списки статей и большинство соавторов как из первого, так и из второго профиля. Стоит отметить, что на eLibrary.ru возможны, хотя и редки, случаи неверного объединения профилей авторов, но в этом случае из-за совпадения большинства соавторов профиль объединен верно. Дополнительного третьего профиля в *Scopus*, в котором могли бы находиться публикации автора за неотраженный период (с 1994 по 2016 гг.), обнаружить не удалось.

### 1.1.3 Оценка доли публикаций в профилях-дублях от общего количества

На третьем этапе исследования для оценки количества публикаций, которые могут пропасть из основного профиля, был введен показатель $L = \frac{S(D)}{S(O+D)}$, где $S(D)$ – суммарное число публикаций во всех профилях-дублях, а $S(O+D)$ –общее число публикаций в основном профиле и его профилях-дублях.

Анализ показал, что для организации средние потери публикаций могут составлять 17% (максимальное значение – 83% достигается для упомянутого *Serbsky Federal Medical Research Centre for Psychiatry and Narcology*, *Affiliation ID* основного профиля = 116873944), для авторов – 11% (максимальное значение – 55%, *AuthorID* основного профиля = 8630661000).

## 1.2 Классификация научных текстов на основе компрессии аннотаций публикаций (на примере Scopus)

В последние десятилетия в связи с экспоненциальным ростом количества информации проблема классификации текстов становится особенно актуальной. При этом классифицируются как целые художественные или поэтические тексты, так и короткие текстовые сообщения, например, СМС или твиты. Но единого метода классификации, который показал бы сразу и высокую эффективность, и низкие трудозатраты, до настоящего момента найдено не было.

Одной из областей, где эта задача имеет особую важность, является классификация научных документов. Классификация крайне важна для задач наукометрического анализа, связанного с оценкой вклада организации и страны в развитие науки, потенциала и перспектив развития. Наиболее авторитетными ББД для ученых являются Web of Science (WoS) и Scopus. Они служат источником данных для множества наукометрических исследований, различных рейтингов университетов, а также используются при оценке публикационной активности исследователей, организаций и стран.

Тем не менее, классификация публикаций в этих библиографических базах данных вызывала большое количество вопросов. Одним из главных ее недостатков являлось то, что

как в WoS, так и в Scopus определение тематики публикации происходило только на уровне журнала, в котором она издана [8,9]. Особенно негативно это сказывалось на статьях из мультидисциплинарных журналов. Публикация, которая могла иметь только одно направление, была отнесена ко всем тем же предметным рубрикам, что и этот журнал. Такое явление создавало «замусоренность» научных направлений публикациями, которые, возможно, не имели к ним никакого отношения.

Для улучшения качества классификации в 2017 году в Scopus введена новая система классификации Topic Prominence in Science, которая была построена подобно методике L. Waltman и N. J. Van Eck, описанной в работе [10]. Авторы предложили способ присвоения публикациям отдельных категорий, который состоял из трех этапов. На первом этапе происходило определение связанности публикаций путем использования прямого цитирования. На втором этапе происходила иерархическая кластеризация публикаций. Третий этап был посвящен названию получившихся кластеров метками на основе терминов заголовков и аннотаций публикаций, находящихся в этой группе.

Ранее в работах Б. Я. Рябко и соавторов был предложен метод классификации научных текстов, основанный на сжатии информации. Он показал более чем 90%-эффективность и низкие трудозатраты при классификации англоязычных текстов архива научных публикаций arXiv.org. Одним из недостатков этих работ являлось применение метода только к полнотекстовым документам. Полные тексты статей, в виду их объема, зачастую содержат много лишних фраз, что может привести к некоторым ошибкам при классификации. Аннотации публикаций, наоборот, должны содержать только ключевые моменты, используемые в статье, что может облегчить автоматическую классификацию научных работ. Более того, во многих научных библиографических базах данных не всегда удается получить полный текст статьи и часто доступна только ее аннотация.

Опишем суть метода, основанный на алгоритмах сжатия. Пусть есть $n$ научных областей $X_1, \ldots, X_n$, для каждой из которых определен характерный для нее набор текстов (файлы, с которыми составляют «ядро» это области). Также есть некоторый тестовый файл $y$, тематику которого нужно определить, и архиватор $\varphi$, который может быть применен для сжатия любого множества текстов.

Работа метода состоит в том, что тестовый файл $y$ начинает последовательно сжиматься с каждым из n ядер при помощи архиватора $\varphi$. В итоге, область тестового файла $y$ определяется тем ядром, с которым он имеет наилучшее сжатие.

Для работы метода были выбраны следующие параметры:

- Архиватор - WinRAR при максимальном значении памяти 128 Мбайт, алгоритм сжатия – PPMd [6, 9];

- Объем ядра – 100 файлов. Этот объем является рекомендованным в работе [9], т.к. при большем количестве файлов в ядре качество классификации практически не улучшается, а время работы алгоритма увеличивается.

Источником информации для настоящего исследования была ББД Scopus. Процесс извлечения данных состоял из трех этапов:

1. Извлечение информации о названии журнала, eid публикации, цитировании, категории через Scival - аналитический инструмент компании Elsevier, основанный на данных Scopus. Данные по 30 категориям были выбраны за период 2009-2018 гг. (Таблица 2)
2. Формирование файлов аннотаций путем получения их текстов через Scopus Abstract Retrieval API
3. Удаление файлов аннотаций с отсутствующим текстом

   Всего было выгружено 15 000 файлов (по 500 для каждой категории).

Для проверки того, насколько эффективно метод работает на аннотациях публикаций, индексируемых в ББД Scopus, классификация проводилась для:

1. Тестовых файлов с одной категорией
2. Тестовых файлов с несколькими категориями

В первом случае осуществлялась проверка работы метода путем классификации файлов с одной категорией, а также подбор оптимальных по составу ядер. Во втором классификация проводилась для тестов с несколькими категориями, ядра для которой были подобраны на первом этапе. Этот случай полезен при проверке качества классификации публикаций из мультидисциплинарных изданий.

При классификации тестовых файлов *с одной категорией* введены три типа ошибок определения категории:

- Ошибки I типа связаны с неправильным определением категории внутри научного направления. Например, вместо категории «Aquatic Science» определилась «Plant Science» из той же области «Agricultural and Biological Sciences».
- К ошибкам II типа отнесены те тестовые файлы, у которых определилась другая область при одном научном направлении. Например, вместо нужной категории «Cell Biology» из области «Biochemistry, Genetics and Molecular Biology» определилась категория «Pharmocology» из области «Pharmacology, Toxicology and Pharmaceutics». При этом общее научное направление «Life Sciences» сохранилось.

**Таблица 2 – Исследуемые области наук по уровням классификации**

| Направление | Область | Категория |
|---|---|---|
| Life Sciences | Agricultural and Biological Sciences | Animal Science and Zoology |
| Life Sciences | Agricultural and Biological Sciences | Aquatic Science |
| Life Sciences | Agricultural and Biological Sciences | Plant Science |
| Social Sciences | Arts and Humanities | History |
| Social Sciences | Arts and Humanities | Literature and Literary Theory |
| Life Sciences | Biochemistry, Genetics and Molecular Biology | Cell Biology |
| Life Sciences | Biochemistry, Genetics and Molecular Biology | Endocrinology |
| Social Sciences | Business, Management and Accounting | Marketing |
| Physical Sciences | Chemical Engineering | Catalysis |
| Physical Sciences | Chemistry | Inorganic Chemistry |
| Physical Sciences | Chemistry | Organic Chemistry |
| Physical Sciences | Computer Science | Artificial Intelligence |
| Physical Sciences | Computer Science | Computer Vision and Pattern Recognition |
| Physical Sciences | Computer Science | Hardware and Architecture |
| Physical Sciences | Earth and Planetary Sciences | Geology |
| Physical Sciences | Earth and Planetary Sciences | Oceanography |
| Physical Sciences | Mathematics | Algebra and Number Theory |
| Physical Sciences | Mathematics | Geometry and Topology |
| Physical Sciences | Mathematics | Logic |
| Physical Sciences | Mathematics | Numerical Analysis |
| Physical Sciences | Mathematics | Statistics and Probability |
| Health Sciences | Medicine | Ophthalmology |
| Health Sciences | Medicine | Surgery |
| Life Sciences | Pharmacology, Toxicology and Pharmaceutics | Pharmocology |
| Physical Sciences | Physics and Astronomy | Astronomy and Astrophysics |
| Physical Sciences | Physics and Astronomy | Condensed Matter Physics |
| Physical Sciences | Physics and Astronomy | Nuclear and High Energy Physics |
| Social Sciences | Psychology | Social Psychology |
| Social Sciences | Social Sciences | Library and Information Sciences |
| Social Sciences | Social Sciences | Sociology and Political Science |

- III тип ошибок является самым серьезным. К нему отнесены те тестовые файлы, у которых ошибка в классификации произошла на самом верхнем уровне. Например, вместо научного направления «Physical Sciences» определилось «Social Sciences».

Классификация осуществлялась с использованием ядер двух типов:

- **произвольные ядра** были составлены из аннотаций публикаций, отобранных случайным образом.

- **подобранные ядра** были составлены из 100 самых высокоцитируемых публикаций в каждой области наук. Такой подход, предположительно, увеличивает качество ядра в связи с тем, что публикации из той же области, цитирующие отобранные в ядро статьи, с высокой долей вероятности наследуют и характерную лексику.

В обоих случаях использовался один и тот же набор произвольных тестовых файлов, для каждой из 30 категорий было отобрано по 20 тестовых файлов, суммарно 600 тестов.

Для классификации тестовых файлов *с несколькими категориями* использовались только ядра с самыми высокоцитируемыми публикациями.

Отбор 20 тестов осуществлялся произвольным образом из публикаций, у которых по меньшей мере 2 категории совпадало с категориями из таблицы 1. Суммарно, как и в предыдущих случаях, было отобрано 600 тестов.

Введем следующие группы результатов:

1. У тестового файла верно определилось не менее 50% указанных категорий. Например, у теста было указано 4 категории: «Algebra and Number Theory», «Numerical Analysis», «Geometry and Topology», «Discrete Mathematics and Combinatorics». В число исследуемых нами категорий входят только первые 3. Соответственно, чтобы попасть в эту группу, у тестового файла должны определиться минимум 2 из 3-х первых. Верно определенными считаются категории, у которых «нормированный коэффициент сжатия» (процент сжатия за вычетом минимального процента сжатия по всем категориям), меньше, чем минимальный процент сжатия со всеми категориями*0,50% (при большем пороге количество ошибок почти не изменялось). В качестве примера такого расчета рассмотрим два тестовых файла с категориями «Algebra and Number Theory» и «Geometry and Topology» (таблица 3). Минимальный процент сжатия у первого теста определился с категорией «Algebra and Number Theory». При этом если в качестве порогового значения выбирать не только минимальный процент сжатия, а минимальный процент сжатия со всеми категориями*0,50%, то с этим тестом правильно будет определена и вторая указанная категория: «Geometry and Topology». У второго же тестового файла была определена только категория «Geometry and Topology».
2. У тестового файла определилась хотя бы одна из указанных категорий
3. Все категории тестового файла определились неверно. К этому случаю будут отнесены те тестовые файлы, у которых определилась какая-то другая категория. При этом один тестовый файл может относиться только к одной из этих групп.

**Таблица 3 – Пример расчета нормированного коэффициента сжатия.**

| **Области теста** | **Algebra and Number Theory** | **Geometry and Topology** | **мин** | **Algebra and Number Theory** | **Geometry and Topology** |
|---|---|---|---|---|---|
| **Algebra and Number Theory, Geometry and Topology** | ***26,70%*** | 26,80% | 26,70% | OK | OK |
| **Algebra and Number Theory, Geometry and Topology** | 33,25% | ***32,85%*** | 32,85% | missed | OK |

*Результаты классификации при произвольных ядрах*

Результаты классификации тестовых файлов с одной категорией при произвольных ядрах по типам ошибок приведены в таблице 4.

Доля ошибочно определенных тестовых файлов составила 32% от общего количества (192 из 600 тестов). Чаще всего определение неверного научного направления происходит из-за категорий, близких по терминологии (например, «Aquatic Science» и «Oceanography»).

**Таблица 4 – Результаты классификации тестовых файлов при произвольных ядрах с различным числом цитирования**

| Категория | Общее кол-во тестов | Кол-во ошибок | Ошибки I типа | Ошибки II типа | Ошибки III типа |
|---|---|---|---|---|---|
| Algebra and Number Theory | 20 | 7 | 5 | 2 | 0 |
| Animal Science and Zoology | 20 | 1 | 0 | 1 | 0 |
| Aquatic Science | 20 | 12 | 11 | 0 | 1 |
| Artificial Intelligence | 20 | 8 | 6 | 2 | 0 |
| Astronomy and Astrophysics | 20 | 1 | 1 | 0 | 0 |
| Catalysis | 20 | 5 | 0 | 5 | 0 |
| Cell Biology | 20 | 1 | 0 | 0 | 1 |
| Computer Vision and Pattern Recognition | 20 | 3 | 3 | 0 | 0 |
| Condensed Matter Physics | 20 | 11 | 2 | 4 | 5 |
| Endocrinology | 20 | 3 | 1 | 2 | 0 |
| Geology | 20 | 0 | 0 | 0 | 0 |
| Geometry and Topology | 20 | 11 | 9 | 2 | 0 |
| Hardware and Architecture | 20 | 0 | 0 | 0 | 0 |
| History | 20 | 8 | 1 | 7 | 0 |
| Inorganic Chemistry | 20 | 13 | 4 | 1 | 8 |
| Library and Information Sciences | 20 | 12 | 2 | 8 | 2 |
| Literature and Literary Theory | 20 | 11 | 5 | 5 | 1 |
| Logic | 20 | 2 | 0 | 1 | 1 |
| Marketing | 20 | 2 | 0 | 2 | 0 |
| Nuclear and High Energy Physics | 20 | 4 | 4 | 0 | 0 |
| Numerical Analysis | 20 | 0 | 0 | 0 | 0 |
| Oceanography | 20 | 11 | 0 | 1 | 10 |
| Ophthalmology | 20 | 9 | 7 | 0 | 2 |
| Organic Chemistry | 20 | 3 | 0 | 2 | 1 |
| Pharmocology | 20 | 18 | 0 | 18 | 0 |
| Plant Science | 20 | 20 | 13 | 7 | 0 |
| Social Psychology | 20 | 1 | 0 | 1 | 0 |
| Sociology and Political Science | 20 | 8 | 0 | 7 | 1 |
| Statistics and Probability | 20 | 6 | 0 | 2 | 4 |
| Surgery | 20 | 1 | 0 | 0 | 1 |
| **Общее количество** | 600 | 192 | 74 | 80 | 38 |
| **Доля от общего количества** | | 32% | 12% | 13% | 6% |

### 1.2.1 Классификация при подобранных ядрах

В таблице 5 приведены результаты классификации тестовых файлов с одной категорией при ядрах, в состав которых входят самые высокоцитируемые публикации. Использование подобранных ядер улучшило результаты классификации на 20%. Число ошибок III типа уменьшилось в 3 раза. В основном такие ошибки возникали из-за находящихся в разных научных направлениях категорий или файлов, в которых применяются схожие термины. Например, у теста из категории «Sociology and Political Science» неверно определилась категория «Aquatic Science», но в тексте этой аннотации применяется много терминов, используемых в категории «Aquatic Science».

**Таблица 5 – Результаты классификации тестовых файлов при ядрах с самыми высокоцитируемыми публикациями**

| Категория | Общее кол-во тестов | Кол-во ошибок | Ошибки I типа | Ошибки II типа | Ошибки III типа |
|---|---|---|---|---|---|
| Algebra and Number Theory | 20 | 8 | 7 | 1 | 0 |
| Animal Science and Zoology | 20 | 7 | 6 | 1 | 0 |
| Aquatic Science | 20 | 2 | 2 | 0 | 0 |
| Artificial Intelligence | 20 | 5 | 2 | 2 | 1 |
| Astronomy and Astrophysics | 20 | 0 | 0 | 0 | 0 |
| Catalysis | 20 | 4 | 0 | 4 | 0 |
| Cell Biology | 20 | 4 | 1 | 3 | 0 |
| Computer Vision and Pattern Recognition | 20 | 3 | 3 | 0 | 0 |
| Condensed Matter Physics | 20 | 2 | 0 | 2 | 0 |
| Endocrinology | 20 | 1 | 0 | 1 | 0 |
| Geology | 20 | 0 | 0 | 0 | 0 |
| Geometry and Topology | 20 | 3 | 3 | 0 | 0 |
| Hardware and Architecture | 20 | 0 | 0 | 0 | 0 |
| History | 20 | 4 | 2 | 1 | 1 |
| Inorganic Chemistry | 20 | 3 | 0 | 3 | 0 |
| Library and Information Sciences | 20 | 3 | 1 | 1 | 1 |
| Literature and Literary Theory | 20 | 2 | 1 | 1 | 0 |
| Logic | 20 | 3 | 2 | 1 | 0 |
| Marketing | 20 | 1 | 0 | 1 | 0 |
| Nuclear and High Energy Physics | 20 | 3 | 3 | 0 | 0 |
| Numerical Analysis | 20 | 0 | 0 | 0 | 0 |
| Oceanography | 20 | 4 | 0 | 0 | 4 |
| Ophthalmology | 20 | 0 | 0 | 0 | 0 |
| Organic Chemistry | 20 | 1 | 0 | 0 | 1 |

| Категория | Общее кол-во тестов | Кол-во ошибок | Ошибки I типа | Ошибки II типа | Ошибки III типа |
|---|---|---|---|---|---|
| Pharmocology | 20 | 2 | 0 | 2 | 0 |
| Plant Science | 20 | 2 | 1 | 1 | 0 |
| Social Psychology | 20 | 2 | 0 | 2 | 0 |
| Sociology and Political Science | 20 | 2 | 1 | 0 | 1 |
| Statistics and Probability | 20 | 1 | 0 | 1 | 0 |
| Surgery | 20 | 0 | 0 | 0 | 0 |
| **Общее количество** | 600 | 72 | 35 | 28 | 9 |
| **Доля от общего количества** | | 12% | 6% | 5% | 2% |

Некоторые неверно определенные тесты, возможно, связаны с неверной изначальной классификацией. Так, в случае категории «Library and Information Sciences» тестовый файл отнесся к категории «Artificial Intelligence», при этом текст аннотации содержит значительно количество терминов, характерных для области «Computer Science».

### 1.2.2 Классификация тестовых файлов с несколькими категориями

Результаты классификации тестовых файлов с несколькими категориями приведены в таблице 6. 23% (140 из 600) тестовых файлов определились ошибочно. При этом неверно определилось научное направление у 6 % (37 из 600) тестов. Стоит отметить, что в некоторых случаях эта ошибка возникала из-за категорий, близких по терминологии, но находящихся в разных научных направлениях. Например, категория «Aquatic Science» из направления «Life Sciences» и категория «Oceanography» из «Physical Sciences».

**Таблица 6 – Результаты классификации файлов с несколькими категориями**

| Группа теста | Количество тестов | Доля от 600 тестов |
|---|---|---|
| Определилось не менее 50% категорий | 413 | 69% |
| Определилась хотя бы одна категория | 47 | 8% |
| Все категории определились неверно | 140 | 23% |

### 1.2.3 Влияние количества ядер на качество классификации

На рисунке 1 изображена зависимость количества ошибок от количества ядер, участвующих в классификации. Результаты показывают, что при расширении числа категорий также увеличивается и число ошибок.

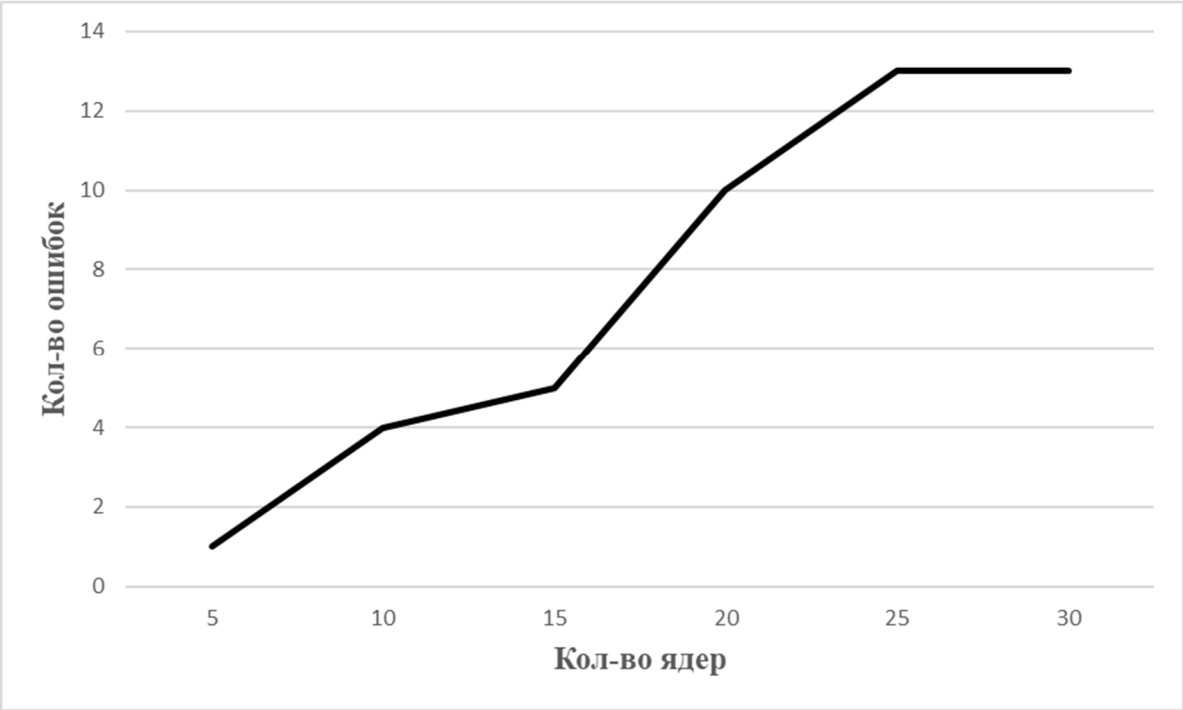


**Рисунок 1 – Зависимость качества классификации от количества ядер**

Если при 5 ядрах ошибочно определился только один тест из категории «Algebra and Number Theory», то при 30 категориях количество ошибок возросло до 13. При этом как при 5, так и при 30 ядрах безошибочно определялись тесты из категории «Surgery».

Таким образом, изначальная выборка количества и состава категорий влияет на точность классификации. Чем меньше категорий участвуют при классификации и чем больше терминологическое различие между ними, тем выше ее качество.

Исследование показало, что метод классификации научных текстов, основанный на алгоритмах сжатия информации, показывает разную эффективность в зависимости от условий, в которых проводится классификация. При произвольных аннотациях публикаций, индексируемых в ББД Scopus, в ядре количество ошибок классификации составило 32%. Подбор ядер из аннотаций высокоцитируемых публикаций позволил сократить их количество до 12%. Это связано, предположительно, с тем, что в таких аннотациях используется более характерная для области наук лексика, которая наследуется и остальными публикациями.

Помимо состава ядер на качество классификации влияет и изначальное количество категорий: чем меньше категорий участвуют при классификации и чем больше терминологическое различие между ними, тем выше ее качество.

Тем не менее, исходная классификация статей была сделана на основе журнальной классификации Scopus, которая также имеет ошибки, и их количество крайне трудно измерить. Это вносит свою объективную погрешность в наши исследования.

## 1.3 Позиционирование работ по улучшению библиографических баз данных и проведения наукометрических исследований в деятельности научных библиотек.

Как известно, отставание нашей страны в развитии пятого технологического уклада очень сильно повлияло на место научных и научно-технически библиотек в исследовательском процессе. До начала 1990-ых они по праву считались уникальными точками доступа к самому актуальному научному знанию, российскому и зарубежному. Однако в настоящее время, значение традиционной библиотеки для исследователя на разных этапах научной работы существенно снизилось: старые технологии оказываются невостребованными, а новые требуют коренной перестройки существующих процессов.

Проблема позиционирования научных и научно-технических библиотек в условиях цифровой трансформации модели научных исследований и коммуникаций широко обсуждается в зарубежной печати. В качестве ответа на эти вызовы рассматривается кардинальное изменение места и роли библиотек в процессе научного исследования, активное участие библиотечных специалистов практически во всех этапах исследования. Особое внимание уделяется участию библиотек в развитии информационных грамотности (Information Literacy) научных сотрудников и соответствующих компетенций, управлении исследовательскими данными (Research Data Management), поддержке инфраструктуры и практик открытой науки (Open Science), включая открытый доступ к результатам исследований, а также работу с наукометрическими и библиографическими данными. Это объясняется тем, что информационная эпоха поставила перед современным научным сообществом вызовы, имеющие системный характер.

***Лавинообразный рост объемов научных материалов.*** Научные публикации не только ушли в цифровую форму, их объемы многократно возросли, причем настолько, что ни один эксперт уже не в состоянии уследить за потоком новых документов в своей дисциплине.

***Усиление формализации и стандартизации научной деятельности.*** Исследователи все больше времени вынуждены тратить на поиск финансирования, подготовку отчетов, написание публикаций, актуализацию своих профилей в базах данных и другие околонаучные задачи. На саму науку времени остается все меньше и меньше.

***Необходимость соответствовать международному уровню, рост интереса к междисциплинарным исследованиям и новым направлениям.*** Принцип «Publish or perish!» заставляет больше публиковаться (желательно, в самых рейтинговых журналах), а развитие коллабораций – осваивать новые, не всегда смежные научные дисциплины. Современный ученый должен учиться, учиться и еще раз учиться.

***Появление новых средств научной коммуникации.*** Web of Science, Scopus, Dimensions, Google Scholar, Pubmed, ResearchGate, Mendeley, FigShare, Arxiv.org, Altmetrics, Publons, ReadCube, VIVO, Pure, dSpace, GitHub… За последние 10 лет появилось так много инструментов, облегчающих научные коммуникации, что это кажется слишком сложным. Разобраться в современном ландшафте технологий научных коммуникаций и сформировать собственную экосистему для эффективной поддержки своей научной деятельности сможет далеко не каждый кандидат наук.

Текущая ситуация располагает библиотеки к поиску таких ниш научно-информационной деятельности, которые:

- освободят высококвалифицированных и дорогостоящих специалистов от всего, что не связано с исследовательским процессом;
- сформируют и будут развивать информационную среду для повышения производительности ученых и обеспечения доступа к необходимым ресурсам.

На основе опыта работы российских научных и вузовских библиотек, анализа зарубежных публикаций по теме Library and Information Science, общения с коллегами и результатах стратегических сессий были сформулированы 10 задач, в которых существенным образом могут участвовать библиотеки, а также необходимые для этого ресурсы и компетенции (рис.2).

Одной из таких задач является поддержка процессов **мониторинга развития научной сферы.** Управление сферой научных исследований и разработок все больше основывается на принципах цифровизации, когда каждый объект исследовательской инфраструктуры (организация, ученый, публикация, патент, проект, научное оборудование) имеет цифровую тень – набор данных, характеризующих исходный объект. Информационные системы и базы данных, содержащие эти сведения, необходимы для проведения наукометрических исследований и экспертизы, которые лежат в основе принятия решений в области развития научной политики.

Обеспечение корректного и полноценного информационного наполнения в таких системах является актуальной задачей. Исследования, в том числе и выполненные в рамках этого проекта, показывают, что качество данных в библиографических базах данных Web of Science, Scopus и РИНЦ оставляют желать лучшего, а их связывание с другими базами данных и вовсе является нерешенной задачей. Основная сложность – это отсутствие глобальных систем идентификации, которые могли бы обеспечить автоматическое связывание (DOI и ORCID уже массово используются, однако еще далеки от полного охвата всего потока научных публикаций). Кроме того, алгоритмы связывания по косвенным признакам, таким как совпадение ФИО, аффилиации, тематики, соавторов, все равно дают

заметное число ошибок, а большое количество записей затрудняет обработку силами операторов. Эти сложности преодолимы, и для этого необходимо создание системы сбора информации о научных исследованиях, в которой определенно есть место экспертам, умеющим обрабатывать большие потоки документов.

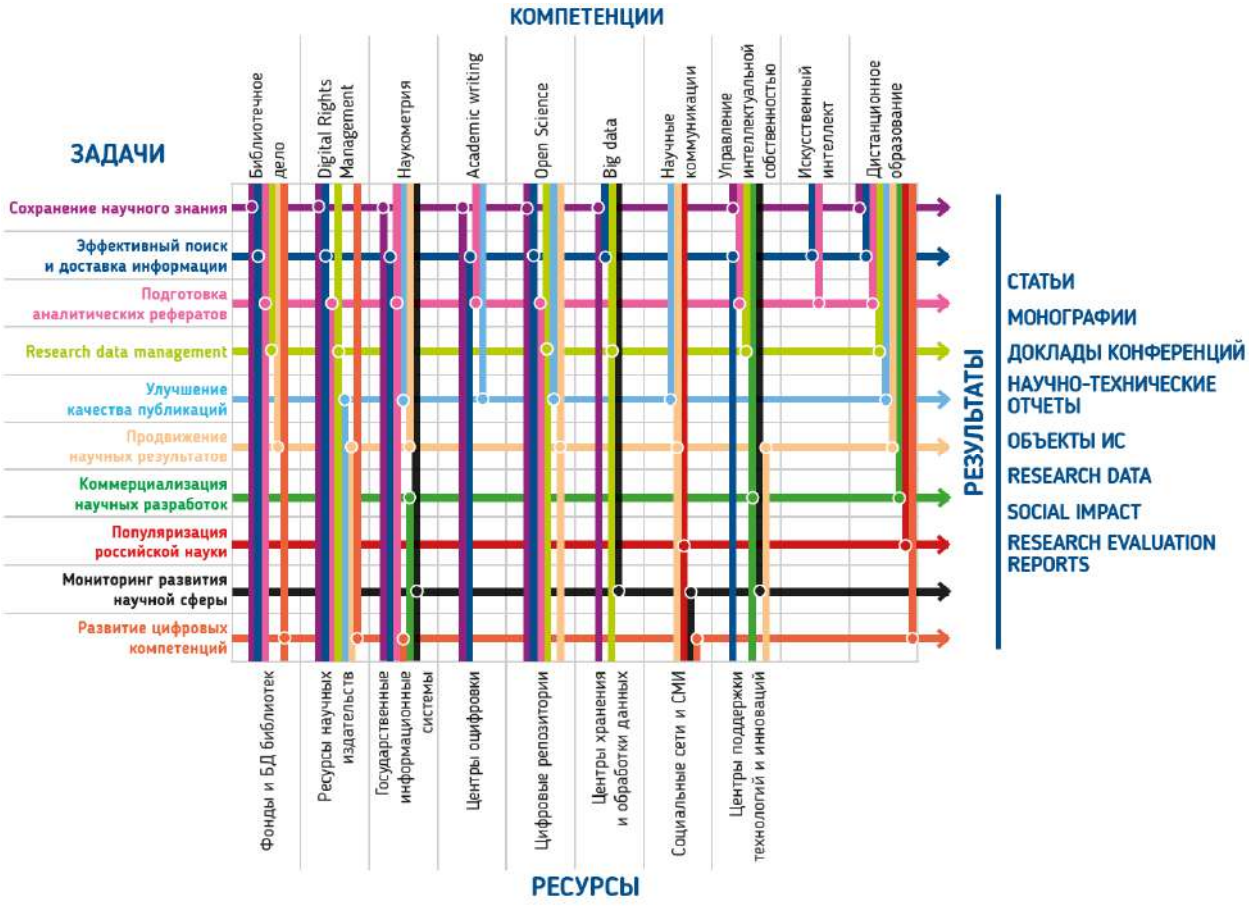


**Рисунок 2 – Матрица, иллюстрирующая компетенции и ресурсы, необходимые для решения задач информационного сопровождения научных исследований.**

Второй важной задачей является наукометрический анализ с целью изучения трендов развития научной сферы и разработки рекомендаций по изменению научной политики. Несмотря на обилие данных о научных исследованиях, их интерпретация является научно-исследовательской работой, которая под силу лишь специалистам с соответствующей квалификацией. Задачами таких исследований могут быть:

- изучение количественных и качественных характеристик публикационной активности;
- оценка результативности научных исследований;
- анализ цитируемости публикаций;
- изучение фронтов исследований в мировой и отечественной науки;

- анализ кадрового потенциала российских исследований;
- развитие научных репозиториев и другой инфраструктуры открытой науки;
- изучение влияния научных исследований на социально-экономическое развитие регионов;
- анализ международных и внутрироссийских коллабораций, академической мобильности;
- изучение взаимодействия академического, вузовского и коммерческого секторов исследований и разработок.
- анализ проблем трансфера технологий и инновационных циклов;
- разработка индикаторов развития науки и технологий.

Для решения этих задач необходимы специализированные постоянно действующие лаборатории, имеющие доступ к широкому кругу источников научно-технической информации, владеющие компетенциями в области наукометрии, интеллектуальной собственности, big data, социальных сетей и СМИ.

Совершенно очевидно, что в рамках одной организации едва ли возможно собрать все необходимое для решения перечисленных задач. Текущая ситуация в научных библиотеках такова, что часть компетенций уже приобретена (например, библиотечное дело и наукометрия), некоторые находятся в процессе освоения и переосмысления (научные коммуникации, открытая наука), а наиболее прорывные (большие данные и искусственный интеллект) имеют слишком высокий порог вхождения. С ресурсами ситуация тоже неоднородная – из-за недофинансирования в научных библиотеках в основном присутствует то, что не требует больших капиталовложений – собственные фонды, генерируемые базы данных, а также доступ к ресурсам научных издательств, предоставляемый по программам национальной и централизованной подписки.

Тем не менее, именно библиотеки являются наилучшими кандидатами для решения перечисленных задач. За редкими исключениями в России нет иных организаций, которые были бы ориентированными на решение прикладных задач информационного обеспечения и демонстрировали бы серьезные успехи. Ряд университетов, научных учреждений и других организаций способны внести значительный вклад в систему информационного обеспечения исследований, поскольку имеют отсутствующие в библиотеках центры хранения и обработки данных, научные лаборатории, которые специализируется на извлечении и интеллектуальной обработке информации и ряд других ресурсов. Это означает, что умелое выстраивание взаимодействия между несколькими участниками, позволит решить многие задачи при небольших затратах.

# 2 РАЗРАБОТКА МЕТОДОВ СТРУКТУРНО-ДИНАМИЧЕСКОГО АНАЛИЗА ПУБЛИКАЦИОННОГО ПОТОКА, ВЫПОЛНЕНИЕ АНАЛИЗА ПОЛУЧЕННОГО МАССИВА ДАННЫХ С ВЫЯВЛЕНИЕМ ИЗМЕНЕНИЙ В СЕКТОРАЛЬНОМ, ТЕМАТИЧЕСКИМ И ТЕРРИТОРИАЛЬНОМ АСПЕКТАХ, ОБНАРУЖЕНИЕ И СОПОСТАВЛЕНИЕ ЭВОЛЮЦИОННЫХ И АНОМАЛЬНЫХ ИЗМЕНЕНИЙ В ПУБЛИКАЦИОННОМ ПОТОКЕ С ВЛИЯНИЕМ СТИМУЛИРУЮЩИХ ФАКТОРОВ.

## 2.1 Данные о публикационном потоке

В качестве источника данных для анализа публикационного потока российских исследователей была выбрана база данных Scopus. Основными преимуществами этой БД являются более широкий охват российских публикаций по сравнению с БД Web of Science, наличие программного интерфейса (API), доступ к которому включен в национальную подписку, использование уникальных идентификаторов для авторов и их аффилиаций. Все это значительно упрощает извлечение и использование библиографических данных для целей наукометрического анализа. Как было сказано выше, база данных Scopus не лишена некоторых ошибок и неоднозначностей, наличие которых может искажать результаты анализа. Однако, комбинирование улучшения данных с использованием контролируемых словарей для уменьшения синонимичности и количества дублирующихся сущностей с разными идентификаторами в сочетании с рассчитанной оценкой возможных ошибок и пониманием их природы позволяет значительно минимизировать это влияние, а в ряде случаев – полностью исключить.

Отбор библиографических данных выполнялся посредством Scopus Search API по запросу вида «AFFILCOUNTRY(“Russian Federation”)» с ограничением по годам публикаций. Были выгружены метаданные всех публикаций российских авторов за 2000-2018 гг. Из-за технологических особенностей Scopus Search API метаданные публикаций с числом авторов более 100 неполны, в связи с этим для уточнения таких записей использовалась выгрузка полных метаданных публикации через Scopus Abstract Retrieval API. В результате была сформирована база данных из 948 568 записей. Последовательностью обработок записи были нормализованы для облегчения последующего анализа, частично исправлены и обогащены дополнительной информацией из других источников. Дальнейший анализ выполнялся с помощью разработанных программных скриптов и с использованием Microsoft Excel.

При расчетах для структурно-динамического анализа использовался частичный фракционный метод счета, при котором публикация «разделялась» между российскими организациями пропорционально количеству авторов, указавших разные аффилиации. При этом суммирование по всем организациям дает общий публикационный поток Российской Федерации.

## 2.2 Особенности национальной научной политики

Развитие российской науки в последние два десятилетия в значительной мере определяется изменениями в национальной научной политике, выразившимися, прежде всего, в сдвиге приоритетов финансирования фундаментальной науки от научно-исследовательских институтов государственных академий наук к высшим учебным заведениям, сопровождавшимся, с одной стороны, демонтажем системы управления академической наукой в процессе реформы государственных академий и созданием сети федеральных, национальных исследовательских и, впоследствии, опорных вузов, целевой поддержкой развития вузовской науки (табл. 7), прежде всего в рамках проекта повышения конкурентоспособности российских университетов 5/100. Особый импульс развитию науки придал переход на количественные целевые показатели публикационной активности по данным международных баз данных научного цитирования (МНБД), заданный в «майских» указах президента РФ в 2012г. и подтвержденный в указе 2018 г.

С 2012 года одним из основных показателей развития российской науки является количество публикаций, проиндексированных в международной базе данных Web of Science, а в последние годы – и в Scopus. Несмотря на обильную критику «наукометрических» методов управления, следует признать, что научное сообщество приняло новые правила игры. Это подтверждается устойчивым ежегодным ростом количества публикаций на 10-20%, который обгоняет среднемировые темпы.

**Таблица 7 – Государственные программы поддержки науки** [11]

| Programme | Years | Amount | Winners |
|---|---|---|---|
| Statement № 89 “On support measures for universities performing innovative educational | 2006-2008 | $ 1.5 B | Universities |
| FTOP `Educational Development´ | 2006-2010 | $ 85 mln | Educational system in total |
| | 2011-2015 | $ 4.5 B | |
| | 2016-2020 | $ 2 B | |
| Federal Universities | 2007-2009 | $ 2 B | Universities |
| FTOP ‘Scientific and Scientific- Teaching Personnel of Innovative Russia’ | 2009-2013 | $ 3 B | Universities |
| | 2014-2020 | | |
| National research universities, National `Education´ project | 2009-2014 | $ 1.5 B | Universities |
| Statement № 218 on developing cooperation between universities and high-tech enterprises | 2010-2012 | $ 0.6 B | Universities |
| Statement №220 on attracting foreign scientists | 2010-… | $ 5 M per grant | Universities |
| FTOP ‘Innovative Infrastructure’ | 2010-2012 | $ 270 M | Universities |
| `5-100-2020´Competitiveness Improvement Programme | 2013-2019 | $ 0.7 B (2013-2016) | Universities |

## 2.3 Взрывной рост количества публикаций российских ученых

Взрывной рост публикаций российских ученых, индексируемых в базе данных Scopus был отмечен проф. Хэнком Мудом в докладе на конференции НЭИКОН в 2017 г. Позже Муд, Маркусова и Акоев проанализировали этот феномен в статье в журнале Scientometrics [12]. Среди основных причин резкого роста после 2012 г. авторы выделили расширение охвата материалов российских конференций, включение новых российских журналов в индексацию и общее повышение публикационной активности российских исследователей как в тех журналах, в которых они публиковались ранее, так и в новых для них журналах. Авторы предположили также, что часть этих новых статей опубликована в индексируемых в БД Scopus журналах вместо публикаций в российских журналах, не индексируемых в международных базах данных. Также было замечено, что количественные характеристики публикационной активности российских исследователей зависят от используемой базы данных и заметно различаются между Scopus и Web of Science.

Наши наблюдения показывают, что данные различных систем различаются, но характер изменений наблюдается и в Scopus, и в Web of Science (рис. 3). Ускоренный рост количества публикаций наблюдается с 2012 г. По количеству публикаций Россия вышла на 11 место в мире, обойдя Нидерланды, Бразилию, Южную Корею и Испанию (рис. 4). В процентном отношении публикации с российскими аффилиациями составили в 2018 г. 3,23% от мирового объема. Выше в рейтинге расположились Австралия (3,4%), Канада (3,6%), Италия (3,85%), Франция (3,9%) и Япония (4,2%). Их доля в последние годы не увеличивается, поскольку эти страны демонстрируют рост публикаций на среднемировом уровне и ниже. Если такая динамика сохранится, то даже ежегодный 10-процентный прирост публикаций позволит России к 2024 году выйти на уровень 180 тысяч публикаций. И этого может оказаться достаточно, чтобы занять пятое место в научном рейтинге Scimago Country Rank.

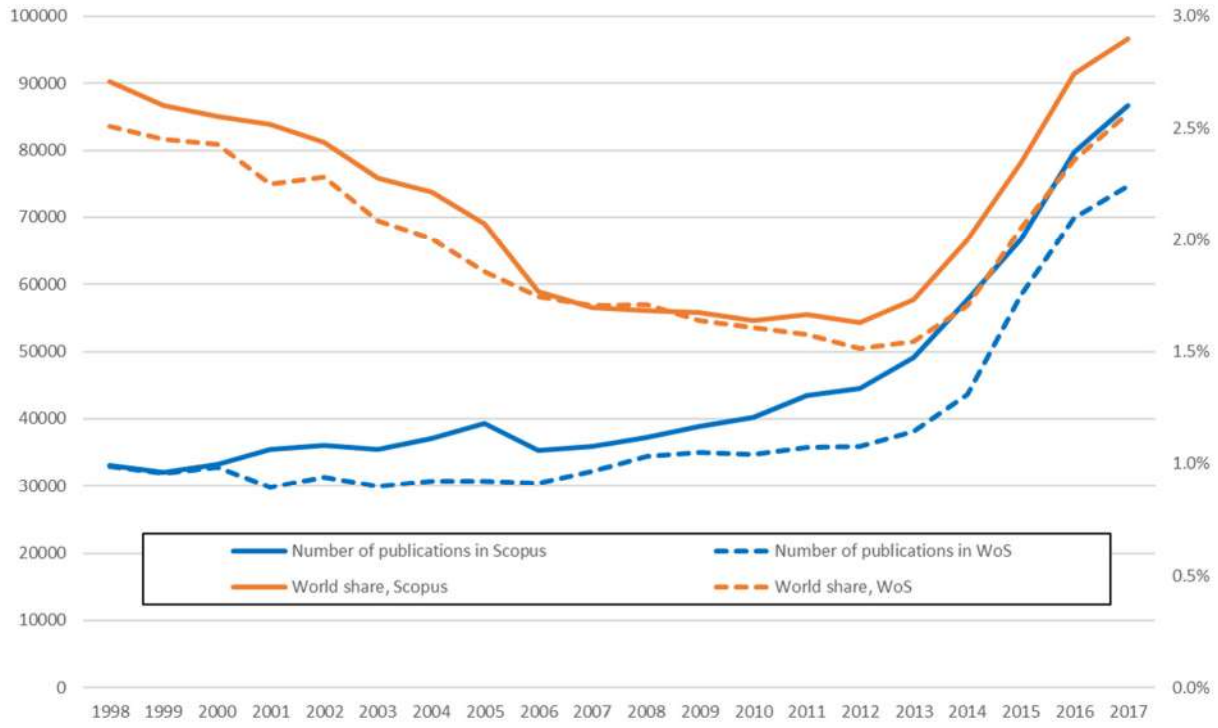


**Рисунок 3 – Количество публикаций (левая шкала) и доля в мировом потоке (правая шкала) российских исследователей по данным Scopus и Web of Science за 1998-2017 гг.**

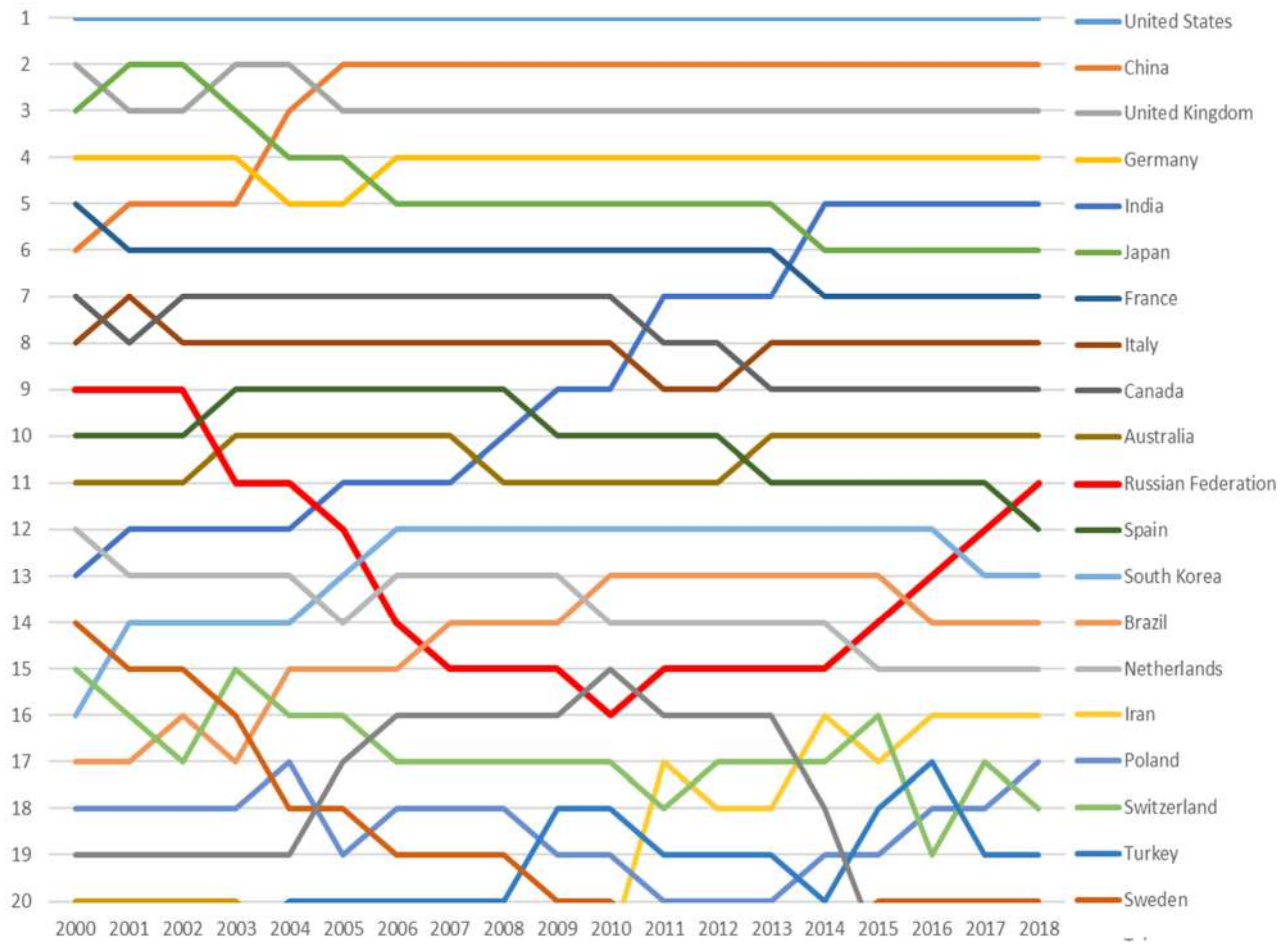


**Рисунок 4 – Изменение позиций лидеров мирового рейтинга по количеству публикаций всех типов в БД Scopus за 2000-2018 гг.**

Очевидно, что основной импульс этому росту был дан в 2012 г. поставленной в указе президента №599 целью «Обеспечить увеличение к 2015 году доли публикаций российских исследователей в общем количестве публикаций в мировых научных журналах, индексируемых в базе данных «Сеть науки», до 2,44 %». В достижение этой цели были вовлечены через различные механизмы практически все российские научные организации и учреждения высшего профессионального образования. Это обеспечило резкое расширение круга активных авторов – российских ученых, публикующих результаты своих исследований в индексируемых в Scopus изданиях (рис. 5). Причем количество активных авторов росло опережающими темпами и только в два последних года наблюдается постепенный рост средней результативности. Из рис. 5 также видно, что заметная часть – более трети активных авторов – это «новые» авторы, то есть ученые, ранее не публиковавшие результаты своих исследований в индексируемых в Scopus изданиях.

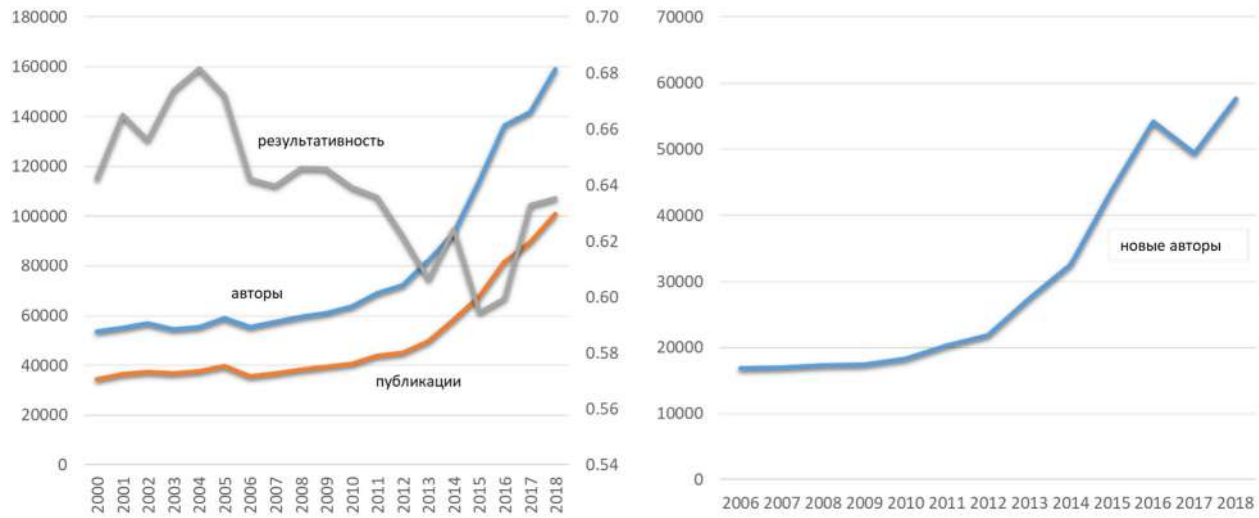


**Рисунок 5 – Увеличение количества публикаций российских исследователей, индексируемых в БД Scopus, количества авторов этих публикаций, изменение средней результативности одного автора и рост числа «новых» авторов, ранее не публиковавших статьи в индексируемых в Scopus журналах.**

## 2.4 Секторальные и региональные диспропорции роста публикационной активности

Анализ показал, что наблюдаемый рост публикационной активности неравномерен как по секторам сферы исследований и разработок, так и в региональном разрезе. В середине 1990-х годов основной вклад в национальный научный продукт вносили исследовательские институты государственных академий наук, что сказывалось и на доле их публикаций, индексируемых в международных базах данных. Так, в 1996 г. около 44% публикаций российских ученых в Scopus были написаны сотрудниками академических институтов (рис. 6). Как уже замечалось выше, под воздействием государственной научной политики в 2010х годах стал активно развиваться сегмент вузовской науки, перетянувший на себя первенство и в 2017 г. обеспечивший около 54% всех публикаций. Количественно продуктивность академической науки выросла (рис. 7), но доля ее в национальном научном продукте упала до 24% в 2017 г.

Изменилось также и региональное распределение публикационного вклада. Центральный экономический регион, включающий Москву и Московскую область, в котором сосредоточено основное количество ведущих научно-исследовательских организаций и крупных вузов заметно потерял в доле публикаций, упав с 60,5% в 1996 г. до 43% в 2017 г. Северо-Западный регион с кластером НИИ и вузов г. Санкт-Петербурга потерял незначительно. Заметно выросла доля Западной Сибири, Уральского, Поволжского, Восточно-Сибирского, Волго-Вятского регионов (рис. 8).

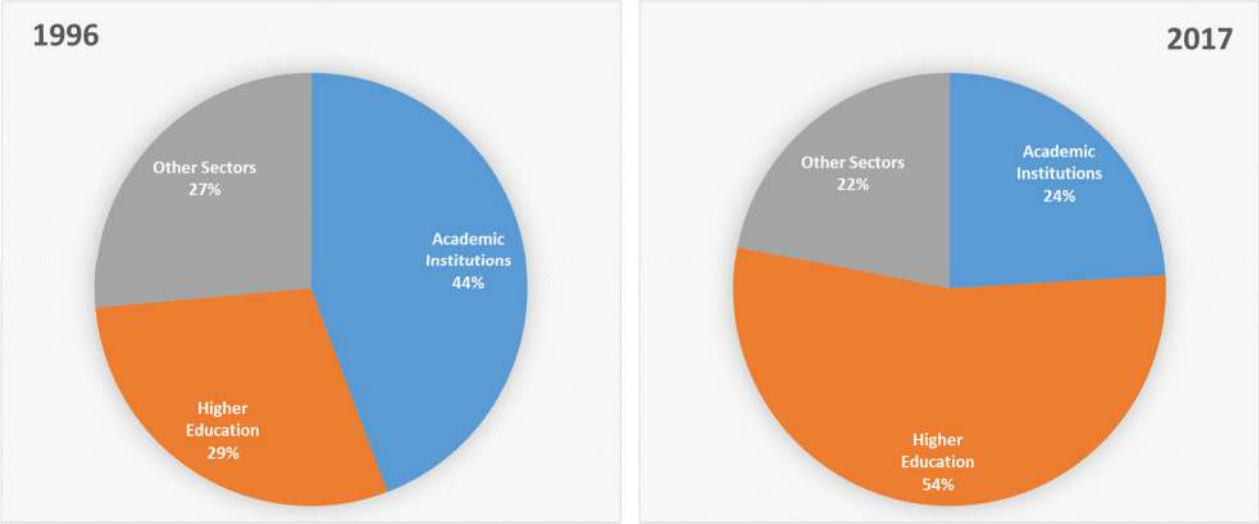


**Рисунок 6 – Структура российского публикационного потока в секторальном разрезе в 1996 и 2017 гг.**

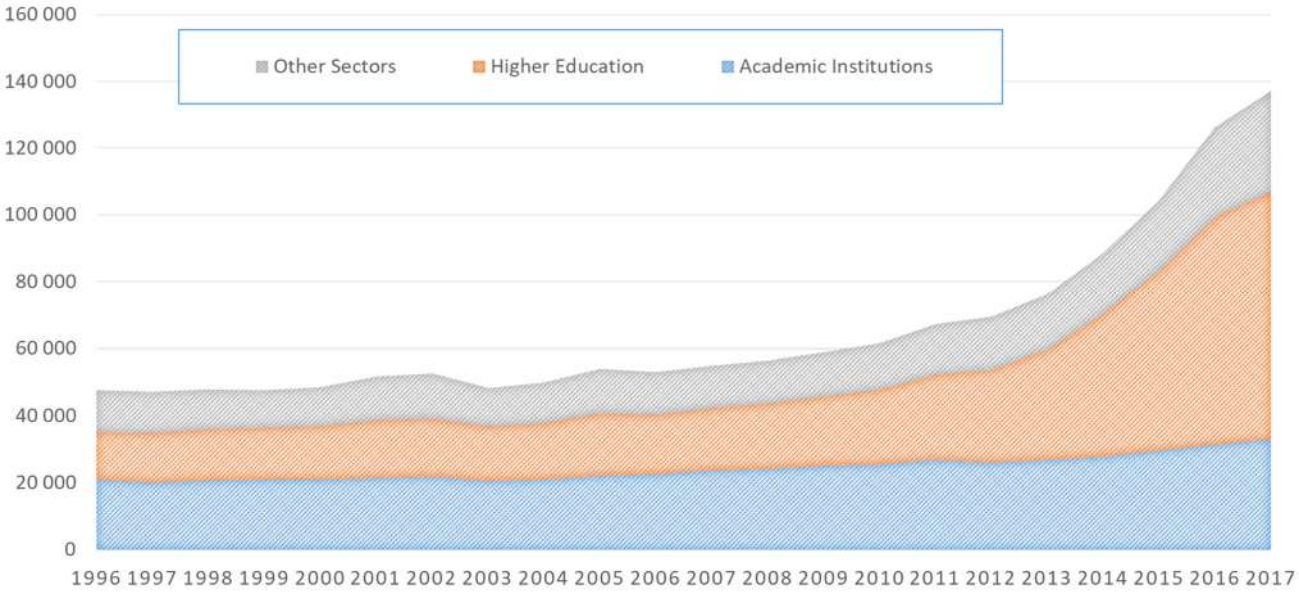

**Рисунок 7 – Динамика структуры российского публикационного потока в секторальном разрезе с 1996 по 2017 гг.**

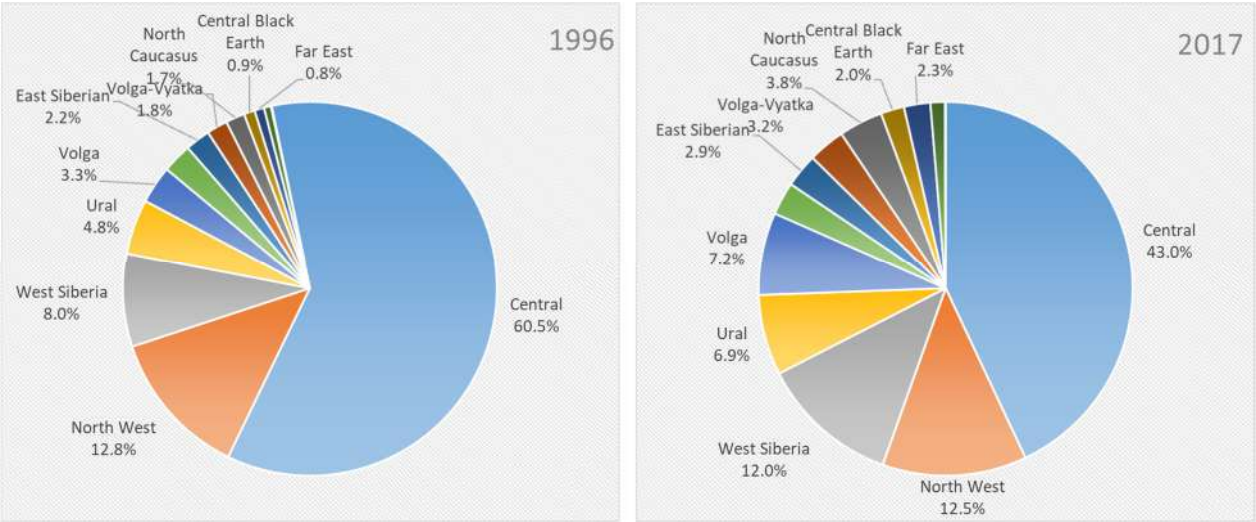


**Рисунок 8 – Структура российского публикационного потока в региональном разрезе в 1996 и 2017 гг.**

Таким образом, можно заметить, что наиболее активный рост продуктивности происходил в сегменте высшего профессионального образования и в регионах. В сочетании с данными о снижении средней результативности российских авторов и увеличении их общего числа нетрудно сделать вывод, что основной рост пришелся на те сектора и регионы, которые обладали наибольшим кадровым потенциалом – научными сотрудниками и преподавателями, не публиковавшими свои результаты в индексируемых изданиях или вообще не занимавшимися наукой. Этот вывод также подтверждается данными о росте количества организаций и городов, указываемых российскими авторами индексируемых в Scopus публикаций (рис. 9).

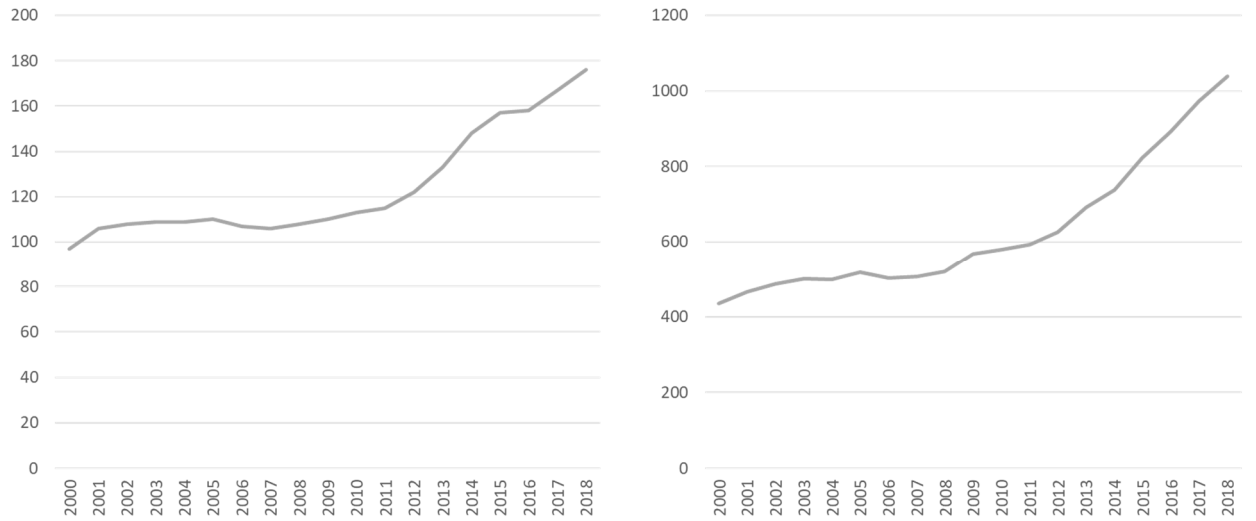

**Рисунок 9 – Рост количества городов (левый график) и организаций (правый график), указанных российскими авторами публикаций, проиндексированных в Scopus. Выбраны города и организации с числом публикаций не менее 10 шт. в год.**

## 2.5 Низкое качество данных государственной статистики по кадровому потенциалу российской науки

Важность целенаправленного и системного развития научных кадров высшей квалификации очевидна для государства, делающего ставку на инновационную экономику. Однако тяжелое наследие массовой утечки мозгов в 1990-х гг., государственные реформы в области образования и науки, а также социально-экономические потрясения последних лет требуют пристального внимания к сохранности и развитию этого ценнейшего ресурса.

Основным источником информации о научных кадрах являются формы государственной статистики, которые ежегодно заполняются научными организациями. Главная форма статистического мониторинга исследований «2-наука» была разработана в 1994 г. и с тех пор изменялась практически ежегодно в целях отслеживания новых тенденций. Например, в 2008 г. в форме стали отдельно учитываться научные исследования и разработки, связанные с нанотехнологиями, а с 2014 г. в табличных формах суммарные данные об организациях высшего образования были разделены на «ведущие классические университеты» – Московский государственный университет (МГУ) и Санкт-Петербургский государственный университет (СПбГУ), федеральные университеты (и отдельно их филиалы), национальные исследовательские университеты (и отдельно их филиалы) и другие образовательные организации. При этом раздел 1 «Персонал, занятый научными исследованиями и разработками» содержит лишь сведения о количестве исследователей по областям наук, возрасту, полу и наличию ученой степени, а также о движении персонала, занятого научными исследованиями и разработками. Очевидно, что на основании этих данных невозможно провести глубокий анализ качества научных кадров, например, выяснить их соответствие мировому уровню (о котором так часто говорится в контексте формирования национальной научной политики), или отследить потоки научной миграции. Представить объективную картину происходящего, основанную на фактических данных, тем более важно, поскольку статистические подходы все чаще не отражают реального положения вещей. Это происходит по следующим причинам.

### 2.5.1 Методики статистического учета не всегда отражают реальную ситуацию

Государственная статистика фиксирует задокументированные факты, однако реальное положение дел может отличаться от «нарисованного» на бумаге, причем достаточно сильно. Причины искажения могут быть различными: от невозможности

корректно фиксировать фактическое состояние дел до манипулирования показателями. Яркий пример приводят Н. Мкртчян и Ю. Флоринская, показывая, что методика учета эмиграции, применяемая Росстатом, неадекватна реальной ситуации. Российская статистика выбытий фиксирует в основном окончание срока временной регистрации в России у граждан постсоветских стран, а не эмиграцию российских граждан в зарубежные страны. Одновременно российская эмиграционная статистика недоучитывает реальных эмигрантов с российским гражданством, выезжающих в развитые страны, так как она включает только тех людей, отъезд которых сопровождается снятием с миграционного учета, а большинство эмигрантов сохраняют свой статус и жилье в России и с учета не снимаются. Последний вывод подтверждается сравнением российских данных и данных национальной статистики стран эмиграции: они различаются в разы [13].

Что касается примеров манипулирования, то в данном случае достаточно упомянуть многострадальный вопрос сравнения среднестатистической и реальной зарплаты профессорско-преподавательского состава вузов. Обсуждения этой темы начались после того, как в 2012 г. Минобрнауки России провело мониторинг и опубликовало данные о средней заработной плате штатных работников профессорско-преподавательского состава федеральных вузов за октябрь 2012 г. [14]. Как оказалось, в рамках майских указов ректорам в качестве одного из показателей эффективности их работы установили средний уровень заработной платы преподавателей вузов, что при дефиците финансирования не могло не привести к разного рода искажениям [15,16]. В 2017–2018 гг. эта тема активно обсуждалась в организациях, подведомственных ФАНО России, где реальность фактического выполнения указа президента о доведении средней заработной платы научных сотрудников до 200 % средней заработной платы по региону также ставилась под сомнение [17]. Есть многочисленные неофициальные свидетельства о том, что научным организациям и вузам приходилось снижать количество штатных научных и педагогических сотрудников путем перевода их на неполную ставку. Очевидно, эти события также внесли заметные искажения в статистические данные.

Так, по данным, приводимым Высшей школой экономики в сборнике «Индикаторы науки» [18], количество исследователей в целом в последние годы менялось незначительно, а в научно-исследовательских учреждениях даже сокращалось. При этом, количество активных авторов статей в Российском индексе научного цитирования, а особенно в международных базах данных (БД), в частности в Scopus, – росло [19]. Если обратить внимание на сектор высшего образования, то количество исследователей в нем в 2016 г. составило почти 45 тысяч человек; в то же время количество авторов статей,

индексируемых в (БД) Scopus и указавших университетские аффилиации, составило почти 70 тысяч (рис. 10).

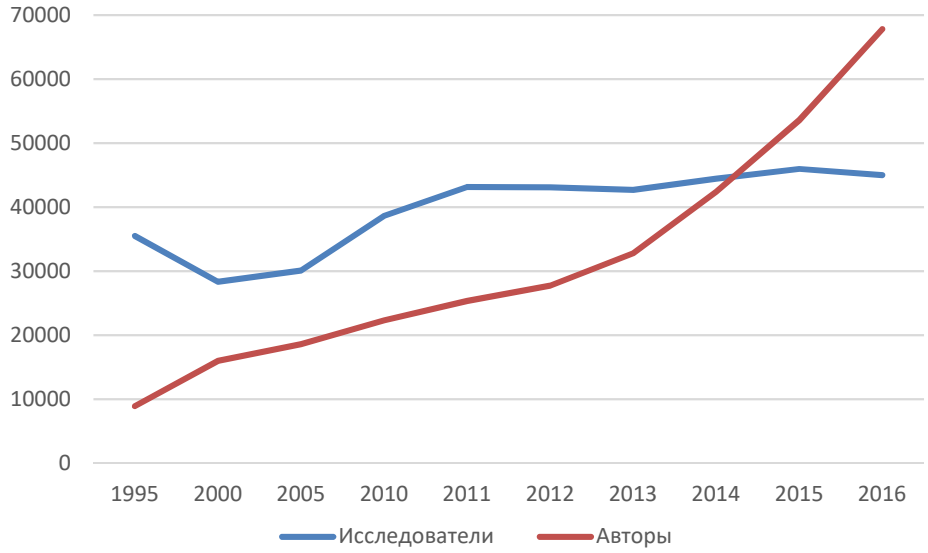


**Рисунок 10 – Количество исследователей, занятых в секторе высшего образования, и количество авторов индексируемых в Scopus публикаций, указавших университетские аффилиации**

### 2.5.2 Результаты статистического учета имеют высокую погрешность, причем величину погрешности сложно оценить, а результаты – проверить

Нельзя утверждать, что заполнению форм статистического учета в научных и образовательных организациях уделяется достаточно внимания. Общее отношение к ним характеризуется словосочетанием «неизбежное зло» – еще одна форма отчетности, смысл подготовки которой, возможно, и существует, но постоянно ускользает из области понимания заполняющих ее сотрудников. Следовательно, возникает довольно небрежное отношение к указываемым в этих формах показателям, что часто делает их недостаточно достоверными. При этом степень достоверности меняется и нередко зависит от самого показателя: новые и сложно вычисляемые показатели могут заметно отличаться от реальных.

В качестве примера рассмотрим ежегодные отчеты, собираемые Федеральной системой мониторинга результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы (ФСМНО). Один из показателей отчета – «количество положительных и нейтральных упоминаний организации в федеральных печатных изданиях, радио- и телевизионных СМИ». При этом не указывается никакой методики подсчета,

общепринятых практик в этой области также не появилось. Таким образом, каждый заполняющий форму, будет изобретать собственный способ расчета, что сделает этот показатель бессмысленным.

Подсчет числа исследователей, работающих в организации, является классической проблемой для любого отдела кадров. Дело в том, что не существует однозначной трактовки понятия «исследователь» (а также «работник, выполняющий исследования и разработки»), следовательно, возможно несколько вариантов расчета их общего количества. Например, в расчете могут присутствовать или отсутствовать преподаватели, инженеры, лаборанты, сотрудники патентных и информационных служб, библиотекари, технические специалисты и т. д., соответственно, результаты могут отличаться в разы. Кроме того, возможно несколько различных вариантов расчета их количества. В практике кадровиков и экономистов и в формах статистического наблюдения «2-наука» и «ЗП-наука» используются понятия «списочная численность», в которой сотрудник учитывается как единица вне зависимости от занимаемой работником ставки, и «среднесписочная численность», которая учитывает ставку, при этом, согласно методическим рекомендациям от 2017 г., при заполнении данных ФСМНО используется только численность «по головам», вне зависимости от занимаемой ставки. Также разночтения могут быть вызваны тем, что в значения в формах статистического наблюдения не включаются внешние совместители, в то время как в ФСМНО они учитываются. Результаты сравнения ежегодной отчетности ФСМНО и сведений, полученных из альтернативных источников, показывают расхождение указанного в отчетах количества исследователей и преподавателей с реальным количеством полноценно работающих сотрудников. При этом надо признать, что по основным показателям методические рекомендации по заполнению доведены до приемлемого уровня и не подразумевают вариантов трактовки. Совершенствуются и процедуры проверки показателей путем их сопоставления с данными из других государственных систем и подтверждения их документально или фактически – сведениями из внешних источников.

Наконец, нельзя не отметить, что территориальные подразделения Росстата в случае заметных изменений статистических показателей по сравнению с предыдущим годом запрашивают у организаций обоснование и объяснение причин таких изменений. Фактически это стимулирует организации сдавать статистические отчеты «как в прошлом году» (или с минимальными изменениями) – в этом случае трудозатраты будут минимальными и вопросов ни у кого не возникнет. В итоге различия между фактической ситуацией и ситуацией «на бумаге» с годами будут накапливаться, хотя сами отчеты будут выглядеть вполне правдоподобно, а выявить такие ошибки практически невозможно.

### 2.5.3 Статистический учет крайне негибок к изменениям параметров мониторинга

Как правило, все отчетные формы имеют фиксированную систему показателей, которые измеряют априори заданные свойства и мало пригодны для изучения чего-либо еще. Например, в форме «2-наука» есть таблица распределения исследователей по областям наук и таблица распределения исследователей по возрасту. Однако построить таблицу распределения исследователей по возрасту в конкретной области наук уже не представляется возможным. Если же новые задачи статистического мониторинга выходят на уровень государственной важности – вводятся новые показатели и таблицы. Так, в 2008 г. в форме появился новый отдельно стоящий показатель «численность исследователей, выполнявших научные исследования и разработки, связанные с *нанотехнологиями*», а в 2009 г. – форма «4-инновации», которая на начало 2018 г. имеет 12 разделов и около 200 (!) строк-показателей. Отследить аналогичные данные до 2009 г. уже невозможно.

Потребность вести мониторинг новых аналитических срезов и введение новых статистических показателей приводят к нелинейному росту сложности отчетных форм, что также снижает их достоверность. Так, появление новых отчетных показателей нередко происходит из-за использования различных классификаторов областей наук (ГРНТИ, УДК, классификаторы Web of Science или Scopus), между которыми нет однозначного соответствия. Для небольших организаций может оказаться, что сдать электронные таблицы с первичными данными проще, чем следующие из них сводные данные.

### 2.5.4 Статистический учет может лишь выявить, но не объяснить очевидные аномалии

Согласно индикаторам науки [18] с 2006 по 2015 г. в России ежегодно работали 75,6–83,5 тысячи исследователей со степенью кандидатов наук и 23,9–28,0 тысячи докторов наук. По данным, приведенным на информационном интернет-ресурсе «Кадры высшей научной квалификации», среднее количество защит кандидатских диссертаций в эти годы составляло почти 23 тысячи в год, докторских – почти 3 тысячи (рис. 11). Фактически статистика говорит о том, что за 10 лет был более чем полностью воспроизведен корпус докторов наук, а кандидатов наук защитилось почти втрое больше, чем их всего трудоустроено в российском научном секторе! На этом фоне прирост исследователей со степенями кандидатов и докторов наук за эти годы был незначительным, то есть большое число высококвалифицированных исследователей оказались потерянными для российской науки.

По каким причинам это происходило? Их несколько: переход на другую работу вне научного сектора, переезд за рубеж, выход на пенсию, естественная убыль. Но долю

влияния этих факторов еще только предстоит установить, и для этого требуются иные методы.

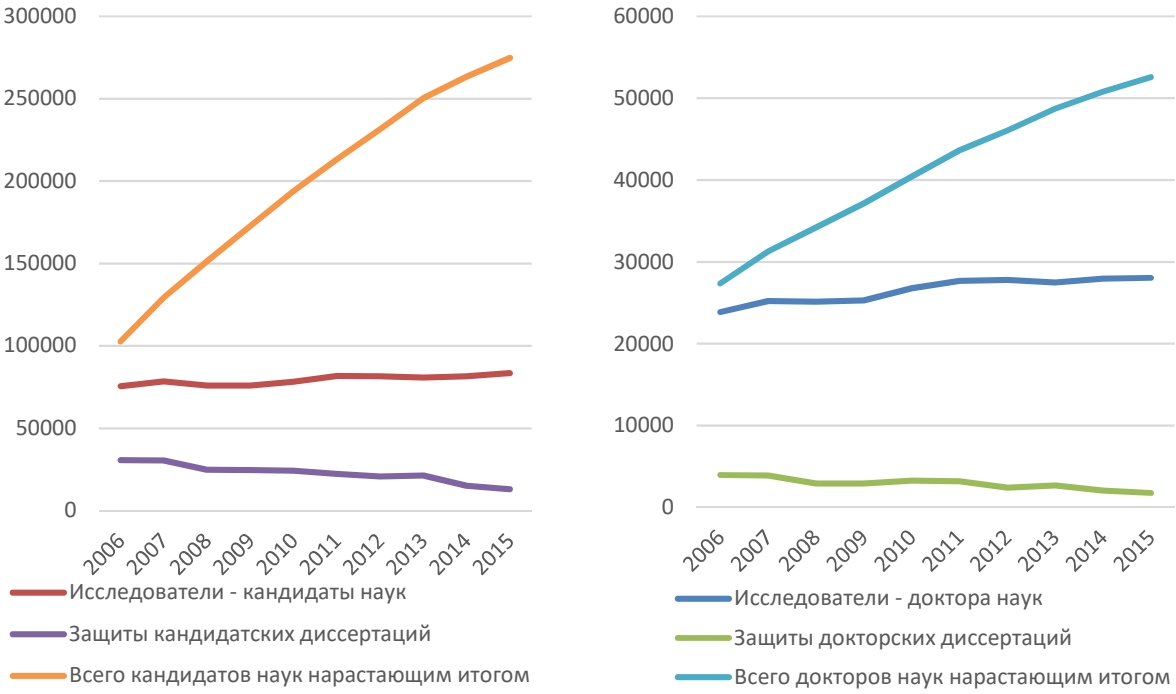


**Рисунок 11 – Динамика количества исследователей – кандидатов и докторов наук, защит кандидатских и докторских диссертаций и накопленное количество кандидатов и докторов наук с 2006 г.**

### 2.5.5 Статистический учет не позволяет проводить качественную оценку научных кадров

Смещение акцента в государственной политике за последние 10 лет на наукометрические показатели привело к заметному росту публикационной активности российских ученых [7, 10]. Однако еще только предстоит ответить на вопрос, за счет чего он произошел. Не привело ли увеличение количества публикаций к понижению качества, как это нередко бывает? Как изменилась доля российских статей в высокорейтинговых научных журналах? Как часто российские публикации попадают в число высокоцитируемых и входят в ядро новых исследовательских фронтов, которые постоянно появляются на переднем крае науки? Привела ли государственная политика к желаемому эффекту: увеличению числа ученых-лидеров и появлению новых сильных научных коллективов? Или же рост был достигнут за счет увеличения общей массы исследователей, публикации которых попали в международные библиографические БД Web of Science и

Scopus? Подобные сведения отсутствуют в формах государственной статистики, что означает их непригодность для качественного анализа деятельности научных кадров.

Попытка комплексировать некачественные данные и применить их для системного анализа часто приводит к неадекватным результатам, аномалиям и искажениям. Такой пример дан в нашей работе [20], где представлено сравнение средних показателей результативности по областям наук, рассчитанных с помощью разных методик и наборов данных. Согласно расчету на основе данных ФСМНО, наиболее результативными являются исследователи в области филологических (2,74 публикации в Web of Science или Scopus на человека) и философских (2,17 публикации) наук. Среди естественнонаучных направлений лидируют математики (1,46 публикации), у физиков – 1,36, у химиков от 0,79 до 1,27. Эти данные входят в глубокое противоречие с общепризнанными наукометрическими исследованиями как на мировом, так и на российском уровне.

### 2.5.6 Данные статистического учета не позволяют строить обоснованные прогнозы

Статистика фиксирует показатели, сформировавшиеся в результате уже случившихся событий. В задачах прогнозирования, как правило, рассматриваются новые ситуации, которые еще не возникали. Это означает, что для описания таких ситуаций нужны совершенно иные данные, которые позволят смоделировать новое пространство вероятных событий. Так, в России впервые в новейшей истории произошла реформа системы диссертационных советов, которая, как уже было отмечено, привела к заметному снижению числа защит диссертаций с 2014 г. Если влияние указанных выше факторов сохранится, не приведет ли это к снижению численности высококвалифицированных кадров из-за того, что скорость их «воспроизводства» будет ниже, чем «потери»?

Другие примеры существенных изменений в национальной научной политике: перенос «центра тяжести» финансирования исследований из академической науки в вузовскую, реформа Российской академии наук, а также неравномерная оплата труда ученых в столице и регионах, вызванная выполнением в 2017–2018 гг. майских указов президента РФ. Как они скажутся на российском научном ландшафте?

Аргументированный ответ на подобные вопросы часто невозможен без построения математических моделей, описывающих национальный сектор исследований и разработок, а для этих моделей необходимы детализированные и достоверные данные о научных кадрах. Отсутствие таких данных у лиц, принимающих решения, в современном мире

означает отсутствие обоснований для принятия решений. Ведь «если у вас нет данных, то все что у вас есть – лишь мнение»[1].

Применение данных государственного статистического учета вполне оправдано, когда не существует иных способов получения данных. При этом необходимо отдавать себе отчет, что эти сведения имеют невысокий уровень достоверности, предназначены для решения фиксированных, заранее определенных задач и плохо подходят для глубоких аналитических исследований, особенно прогнозного характера. Наиболее распространенная альтернатива – социологические опросы, которые дают широкую свободу в формулировке исследовательских задач и, при тщательной подготовке, демонстрируют неплохую точность результатов.

Однако и у них есть свои недостатки. Один из главных – опросы могут фиксировать лишь мнение респондентов, а не факты. Кроме того, опрос на маленькой выборке респондентов будет иметь низкую точность, тогда как опрос с достаточно репрезентативной выборкой является уже весьма ресурсоемким мероприятием, но подобные опросы не могут проводиться слишком часто.

### 2.5.7 Новые возможности

Так или иначе, за неимением реальных альтернатив статистический учет до сих пор остается основной формой государственного управления, основанного на данных (data-driven management). Но всеобщая информатизация и цифровая революция предлагают новые подходы и принципы управленческого анализа и поддержки принятия решений. Какие изменения к этому привели?

- Массивы первичной информации переведены в электронную форму, пригодную для автоматизированной обработки.
- Профессиональными сообществами разработаны стандарты передачи открытых данных, позволяющих собирать первичную информацию в одном месте.
- Появились технологии идентификации сущностей, позволяющие проводить связывание данных, полученных из различных источников.
- Стали доступны вычислительные мощности, позволяющие обрабатывать и анализировать большие объемы данных.
- Разработаны математические модели и программные алгоритмы, выявляющие ошибки в больших объемах данных, обнаруживающие аномалии, скрытые связи и характерные шаблоны, существенно улучшающие их качество.

[1] “Without data, you're just another person with an opinion” (W. Edwards Deming)

В результате те данные, которые раньше были доступны лишь для изучения «на местах», теперь могут быть легко перемещены и обработаны в любом месте и с относительно небольшими трудозатратами.

Безусловно, переход от сведения статистический форм к анализу первичной информации является тектоническим сдвигом системы государственного управления и важнейшим шагом к так называемой цифровизации экономики. При этом нередко можно наблюдать частичные решения, когда устаревшие подходы пытаются перевести на новые «цифровые рельсы», продолжая через новые системы электронного документооборота собирать все те же агрегированные данные с низким уровнем достоверности. Это приводит к увеличению скорости сбора данных и снижению трудозатрат на обработку, но не решает остальные проблемы.

Важность кадрового научного потенциала для социально-экономического развития нашей страны делает необходимым исследование его текущего состояния и действующих трендов, основанное на первичных (то есть на фактических) данных с высокой степенью достоверности результатов.

## 2.6 Наукометрический подход

Наукометрический подход предполагает формирование, обработку и анализ массива данных, описывающих относительно полный массив сведений об исследуемом объекте научной деятельности. Чаще всего таким массивом оказывается выборка библиографических описаний научных публикаций (для анализа русскоязычного научного сегмента), полученная из систем Web of Science, Scopus или Российский индекс научного цитирования (РИНЦ). Преимуществом такого подхода является более высокая достоверность (исследователь контролирует все этапы анализа, начиная с получения первичных данных[2]), воспроизводимость (исследование всегда можно повторить с использованием тех же данных и методик) и полнота охвата (которая лимитируется лишь ограничениями библиографических БД).

Для решения задачи мониторинга научных кадров необходимо определить методику сбора и формирования рабочего массива первичных данных. Его ядром должен стать реестр активных российских ученых. Затем для каждого исследователя должен быть сформирован

[2] В этом исследовании под первичными данными понимаются неагрегированные сведения, характеризующие исследовательский процесс. К ним относятся метаданные о публикации (название, издание, авторы и пр.), об исследователях (ФИО, место работы, должность и пр.) и так далее.

его научный цифровой след – набор данных о его научных результатах, на основе которого можно будет оценить его вклад в науку.

Наиболее полными и достоверными данными о российских исследователях обладают научные и научно-образовательные организации. Однако получить информацию из их кадровых БД не представляется возможным (в частности, из-за закона о персональных данных, различий в используемых форматах хранения данных и административных барьеров), а их веб-сайты не позволяют извлечь ее в унифицированной форме и в необходимом объеме. В будущем эта проблема могла бы быть решена за счет создания межведомственной системы кадрового учета, которая в автоматическом режиме собирает эту информацию, что одновременно сделает бессмысленным периодическое заполнение форм кадровой отчетности.

Фактически наиболее полноценным из доступных источником информации об активных исследователях, работающих в фундаментальной науке, являются библиографические БД публикаций, содержащих данные об опубликованных результатах научной деятельности. И если раньше главной задачей таких систем было формирование каталогов публикаций, отвечающих определенным признакам (например, общая тематика), то в современных системах есть устойчивый тренд «from cataloguing to catalinking» [13] – выделение сопутствующих вторичных сущностей и их связывание с внешними источниками данных. Одной из таких сущностей являются исследователи – авторы публикаций. Следовательно, БД индексов цитирования могут выступать источником информации и о количестве исследователей, и об объеме их вклада в фундаментальную науку.

Три реферативные БД, наиболее пригодные для проведения наукометрических исследований, относящихся к России, – Web of Science (владелец – Clarivate Analytics), Scopus (Elsevier) и РИНЦ (ООО «Научная электронная библиотека») – имеют каждая свою историю создания, охват источников, структуру данных и регламенты работы с ними. В 2018 г. было объявлено о запуске еще одной системы – Dimensions (Digital Science) – с большим международным охватом, однако ее применимость для наукометрических исследований еще предстоит подтвердить. На основе этих БД для каждого исследователя будет определена его публикационная и аффилиационная история, анализ которой позволит, например, выявить факты академической мобильности – смену места работы, в частности сопровождающуюся переездом в другой город или страну, изменение сектора науки (вузы, академические и отраслевые НИИ). Изменение тематики публикаций позволит установить факт смены области научных интересов исследователя.

Следует отметить, что в Scopus недостаточно полно представлены российские исследования в ряде областей наук: сельском хозяйстве, медицине, гуманитарных направлениях. В связи с этим представляется целесообразным, преодолев организационные и технические барьеры, сопоставить полученную картину с данными РИНЦ.

Еще одним ценным источником данных о карьерах российских ученых является массив сведений о защитах кандидатских и докторских диссертаций, который публикуется в Единой государственной информационной системе учета результатов научно-исследовательских, опытно-конструкторских и технологических работ гражданского назначения (ЕГИСУ НИОКТР) с 1991 г. Логично предположить, что за 27 лет подавляющее большинство активных российских ученых оставило след в этой БД в качестве соискателя, научного руководителя или оппонента хотя бы одной диссертации. Таким образом, появляется задача связывания массива авторов публикаций (в различных областях наук) с массивом диссертаций. Ее решение вплотную подводит нас к ответу на такие важные вопросы: каким образом меняется публикационная активность исследователя после защиты им диссертации в различных областях наук, какая часть защитившихся российских ученых мигрировала за рубеж, как меняется эта динамика в разные годы и в разных областях наук.

## 2.7 Аберрации в количественных показателях публикационной активности и их причины

Количественные показатели публикационной активности организаций подвержены заметным искажениям, связанным с разного рода недобросовестными и сомнительными практиками. Широко известен феномен хищнических издателей и журналов, размещающих статьи авторов без должной процедуры рецензирования за деньги. Авторы недавнего исследования [21] оценивают долю публикаций российских авторов в хищнических журналах в 10% общего публикационного потока. Значительный интерес наукометристов вызывает практика недобросовестного авторства, особенно самый свежий его вариант – покупное авторство. В 2019 г. специалисты Clarivate Analytics обратили внимание на сайт организации «Международный издатель», продающей авторство в уже принятых в печать статьях на аукционах [22]. Влияние этой практики оценить сложно, но с учетом утверждений компании о том, что они уже продали авторство 10 тыс. «ученых» оно может быть значительным. Заметным может быть также влияние плагиата и самоплагиата, в том числе в форме повторных публикаций одной и той же статьи [23].

В немалой степени выдающийся результат по количеству публикаций в Scopus достигнут за счет материалов конференций – за последние годы их доля увеличилась с 16%

до 27%. Если рассчитывать рейтинг только по документам типа Article и Review, то позиции России окажутся скромнее – 13-ое место. Большая часть этих конференций проводится российскими организациями, многие из них являются мультидисциплинарными, в труды попадают сотни статей, причем размер этих статей небольшой, качество часто вызывает вопросы к уровню рецензирования материалов.

### 2.7.1 Академическая мобильность и миграция как важный фактор развития российской науки. Феномен синхронной мобильности.

В последние годы феномен мобильности ученых привлекает значительное внимание исследователей в различных областях знаний, что вызвано возросшими потоками научной миграции, процессами глобализации, затронувшими в том числе научные исследования, а также увеличением числа междисциплинарных и политематических работ. В международном масштабе мобильность приносит положительные результаты для генерации, накопления и распространения научных знаний [24,25], расширения международных коллабораций [26] и карьерного роста ученых. Мобильность способствует созданию новых научных центров притяжения. Таким образом, она является саморегулируемой системой [27]. При этом на региональных уровнях могут наблюдаться отрицательные эффекты [28]. По данным масштабного исследования, охватывающего 16 ведущих стран [29], до 40% ученых – иммигранты, еще 40% поддерживают связи с родными странами, способствуя передаче передовых научных наработок. Участие мобильных ученых в международных проектах выше на 13,8%, а у вернувшихся на родину – на 7,4% в сравнении с немобильными исследователями.

Мобильности противопоставляется академическая немобильность, для обозначения которой в англоязычной литературе используется понятие инбридинг (термин заимствован из биологии и означает скрещивание близкородственных форм в пределах одной популяции) [30]. Применительно к научным исследованиям инбридинг – работа исследователя в организации, в которой он получил образование [31]. Отсутствие мобильности в основном оценивается негативно, в отдельных случаях наблюдается положительный эффект для карьерного роста исследователей.

В целом у немобильных ученых отмечается более низкая научная продуктивность, а в организациях, не поощряющих мобильность, – частичное нарушение принципов прозрачности проведения исследований, неполное соответствие учебных курсов международным стандартам, более частое возникновение сфер влияния отдельных исследователей. Негативный эффект может быть обусловлен более низким доходом и большей научной и преподавательской нагрузкой немобильных ученых, выполнением ими административных заданий [30].

Немобильные ученые в меньшей степени мотивированы на совместную работу с зарубежными коллегами, в их публикациях существенно меньше зарубежных соавторов по сравнению с мобильными, в том числе вернувшимися из-за рубежа, а статьи, как правило, публикуются в журналах с более низкими импакт-факторами [32].

Многое из перечисленного характерно для российских исследователей. Выявлено, что за последние десять лет 70% российских ученых не меняли место работы [28]. До 45% научных сотрудников экономических факультетов 28 вузов Санкт-Петербурга одобряют практику, когда на работу принимают собственных выпускников, и лишь 12% респондентов согласны предпочесть мобильных исследователей [30]. За инбридинг выступают преимущественно немобильные ученые, тогда как мобильность считают лучшей моделью мобильные.

Согласно показателям научной продуктивности, немобильные ученые в основном публикуются в университетских периодических изданиях (мобильные – в более авторитетных журналах, издаваемых организациями Российской академии наук); для опубликования работ они чаще используют имеющиеся неформальные связи; у них значительно более узкий круг профессионального общения. По сравнению с мобильными исследователями, немобильные в меньшей мере склонны обсуждать профессиональные вопросы, касающиеся методов и организации проведения исследований.

Примечательно, что немобильные ученые чаще всего нацелены на карьеру внутри своей организации и не стремятся приобрести широко признанный авторитет. Стоит указать отдельно на очень ограниченное взаимодействие российских ученых с бывшими соотечественниками, работающими в зарубежных научных организациях (в отличие, например, от ученых Китая и Индии). Необходимость в такой коллаборации ясно осознается представителями структур в области научной политики [33,34].

Анализ источников по этой теме позволил выделить четыре основных подхода к типизации мобильности (табл. 8).

Перечисленные типы академической мобильности могут пересекаться. Например, переезд ученого в другую страну может сопровождаться должностным повышением, сменой направлений исследований и носить временный характер. В этом случае мобильность будет международной, временной, дисциплинарной и вертикальной.

Учитывая повышенный интерес к мобильности ученых в последние годы, исследователи разрабатывают различные подходы к ее анализу. На данный момент преобладают социологические подходы. Многообещающими представляются подходы с использованием наукометрических методов – из-за относительной простоты их применения, прозрачности и объективности. В наибольшей мере они применимы к

исследованию связи мобильности с карьерным продвижением ученых, динамикой их публикационной активности, с созданием научных коллабораций и к выявлению маршрутов научной миграции ученых.

**Таблица 8 – Типы мобильности и их основные признаки.**

| Подход к исследованию | Тип мобильности | Основные признаки |
|---|---|---|
| Географический | Международная | Переезд ученого из родной страны за рубеж |
| | Внутригосударственная | Смена организации ученым в своей стране или городе |
| Временной | Временная | Смена организации ученым на временной основе |
| | Постоянная | Смена организации ученым на постоянной основе |
| | Синхронная | Одновременное трудоустройство ученого в нескольких организациях в родной стране или в нескольких странах |
| Предметный | Отраслевая (дисциплинарная) | Смена ученым направления исследования или научной дисциплины |
| | Межсекторальная | Переход ученого из научной организации или университета в наукоемкие, производственные организации и пр. |
| Профессиональный | Горизонтальная | Смена ученым организации – без карьерного продвижения |
| | Вертикальная | Смена ученым организации – с повышением в должности |

Детальное исследование академической мобильности позволяет сформировать реальную картину кадрового потенциала страны и проанализировать влияние на нее различных факторов: государственной политики в области образования и науки, социально-экономической ситуации и др.

Отдельный интерес вызывает эффект, связанный с международной синхронной академической мобильностью – явлением, связанным в значительной степени с многочисленными программами привлечения ведущих зарубежных ученых в российские организации.

## 2.8 Международная синхронная мобильность и ее влияние на результативность российских исследователей

Важность международных коллабораций для повышения результативности и качества научных исследований широко признана в научном мире [35,36]. Курс на интеграцию российской науки в международное сообщество, усиление международных

коллабораций, в том числе за счет привлечения в российские организации ведущих зарубежных ученых закреплен программными документами и указами президента Российской Федерации.

На фоне быстрого роста общего количества публикаций, количество российских публикаций в международных коллаборациях росло значительно медленней, а их доля в общем потоке стабильно уменьшалась, начиная с 2007 года несмотря на все усилия, предпринимаемые государством для интенсификации международного научного сотрудничества [37] (рис. 12). Программы привлечения ведущих зарубежных ученых привели к появлению в российских университетах и исследовательских институтах иностранных исследователей. В большинстве случаев зарубежные исследователи привлекаются на частичную занятость и временные позиции, в публикациях они при этом указывают несколько аффилиаций. Некоторые публикации этих ученых выходят без участия российских соавторов, при этом в них также указана в числе прочих российская аффилиация этого ученого. Встречается и обратная ситуация – когда российский автор указывает в качестве дополнительной зарубежную аффилиацию, при этом все остальные авторы указывают только российские аффилиации. На основе формальных критериев такие публикации рассматриваются как написанные в международных коллаборациях с российским участием, хотя вряд ли такими являются, так как не могут рассматриваться как результат сотрудничества исследователей из России и других стран.

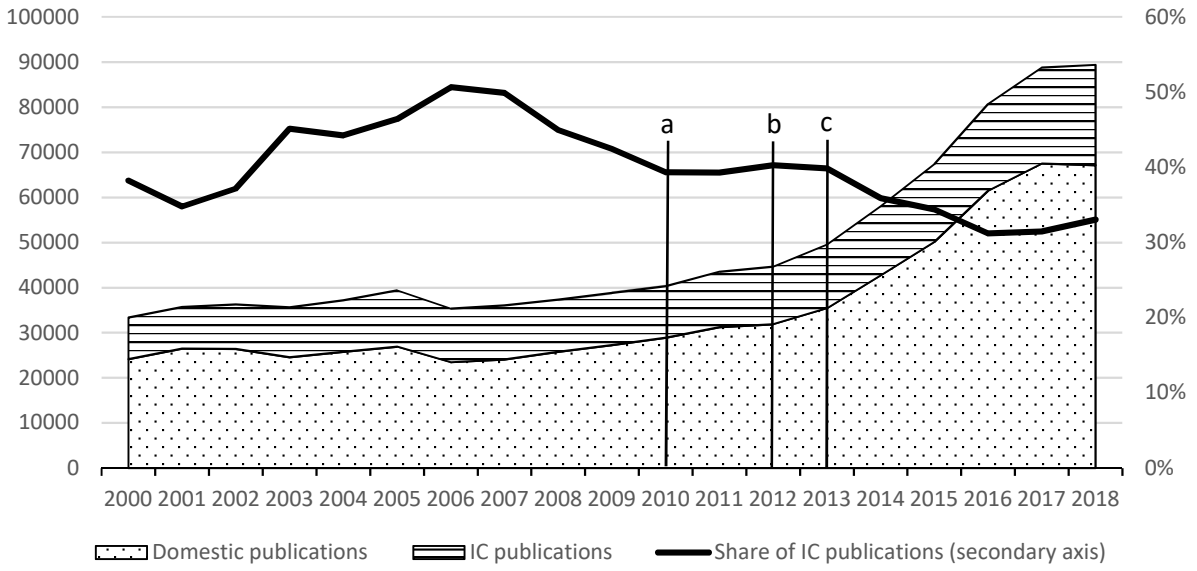

**Рисунок 12 – Общее количество российских публикаций в Scopus, количество публикаций в международных коллаборациях (левая шкала) и доля таких публикаций в общем потоке (правая шкала).**

В контексте этого исследования нас прежде всего интересует оценка влияния феномена множественных аффилиаций на количественные показатели международной научной коллаборации российских исследователей и оценка эффективности программ привлечения ведущих зарубежных ученых, выявление и попытка категоризации «российских» авторов, указывающих множественные аффилиации.

Нами был проанализирован массив российских публикаций в международных коллаборациях, в которых хотя бы один автор указал одну или несколько российских аффилиаций и одну или несколько – зарубежных. Публикации были классифицированы по следующему принципу:

- Если у всех авторов публикации присутствует как российская, так и зарубежная аффилиация, такая публикация относится к типу «SP» (Synchronous paper).
- Если публикация не относится к типу «SP» и у всех авторов публикации присутствует минимум одна российская аффилиация, такая публикация относится к типу «RP» (Russian Paper).
- Если публикация не относится к типу «SP» и у всех авторов присутствует хотя бы одна зарубежная аффилиация, такая публикация относится к типу «FP» (Foreign Paper).
- У остальных публикаций есть как авторы только с российскими публикациями, так и авторы с только иностранными публикациями. Такие публикации относятся к типу «IP» (International Paper).

Аналогичным образом была проведена классификация авторов, для чего пришлось выполнить дополнительное извлечение данных о полностью зарубежных публикациях этих авторов:

- Авторы, указывавшие в части публикаций только российские аффилиации и ни разу – только зарубежные, были отнесены к типу «RA» (Russian Authors).
- Авторы, указывавшие в части публикаций только зарубежные аффилиации и ни разу – только российские, были отнесены к типу «FA» (Foreign Authors).
- Авторы, указывавшие, как только зарубежные, так и только российские аффилиации в части своих публикаций были отнесены к типу «MA» (Migrating Authors).

- Авторы, во всех публикацях которых за рассматриваемый период были указаны как российская, так и зарубежная аффилиации были отнесены к типу «SA» (Synchronous Authors). У большинства таких авторов в выборке присутствует только одна публикация, что позволяет предположить, что в этом случае мы имеем дело с дублем авторского профиля.

В таблице 9 приведены данные по количеству российских публикаций в международной коллаборации в соответствии с описанными выше принципами классификации. Можно заметить, что доля полноценных коллабораций снизилась с 80% в 2009 до 70% к 2018 за счет роста количества публикаций с российскими соавторами, некоторые из которых указали дополнительную зарубежную аффилиацию (RP) и публикаций с зарубежными соавторами, некоторые из которых указали дополнительную российскую аффилиацию (FP). Доля последних выросла почти до 20% от всех российских публикаций в международной коллаборации и почти 4,5% от общего количества российских публикаций, что также хорошо видно на рис. 13. Как уже было сказано выше, количество российских публикаций в международных коллаборациях более чем удвоилось за рассматриваемый период, но на фоне почти трехкратного роста общего количества публикаций это привело к снижению доли таких публикаций в общем количестве.

**Таблица 9 – Количество российских публикаций в соответствии с предложенной классификацией**

| Годы | SP | RP | FP | IP | Всего |
|---|---|---|---|---|---|
| 2000 | 409 | 277 | 1604 | 6935 | 9225 |
| 2001 | 398 | 271 | 1535 | 7012 | 9216 |
| 2002 | 411 | 291 | 1420 | 7700 | 9822 |
| 2003 | 348 | 303 | 1922 | 8505 | 11078 |
| 2004 | 370 | 287 | 1928 | 8817 | 11402 |
| 2005 | 388 | 403 | 1838 | 9857 | 12486 |
| 2006 | 398 | 408 | 1687 | 9404 | 11897 |
| 2007 | 329 | 358 | 1659 | 9648 | 11994 |
| 2008 | 349 | 388 | 1719 | 9135 | 11591 |
| 2009 | 387 | 399 | 1465 | 9308 | 11559 |
| 2010 | 392 | 404 | 1529 | 9060 | 11385 |
| 2011 | 414 | 435 | 1687 | 9739 | 12275 |
| 2012 | 415 | 601 | 1816 | 9998 | 12830 |
| 2013 | 412 | 730 | 1991 | 10989 | 14122 |
| 2014 | 468 | 829 | 2447 | 11595 | 15339 |
| 2015 | 511 | 1089 | 2871 | 12780 | 17251 |
| 2016 | 593 | 1365 | 3296 | 13934 | 19188 |
| 2017 | 569 | 1575 | 3694 | 15414 | 21252 |

| | | | | | |
|---|---|---|---|---|---|
| 2018 | 611 | 1842 | 4331 | 15411 | 22195 |

**Рисунок 13 – Количество публикаций с зарубежными соавторами, некоторые из которых указали дополнительную российскую аффилиацию (FP) и доля таких публикаций в общем количестве российских публикаций в международных коллаборациях.**

Классификация авторов по предложенным принципам дала результаты, представленные в таблице 10. Рассматриваются только те авторы, которые хотя бы в части публикаций указывали как российскую, так и зарубежную аффилиацию. Указано количество активных авторов, то есть имеющих хотя бы одну публикацию в данный год. Можно заметить, что среднее по году количество приглашенных зарубежных авторов (FA) колебалось в пределах 600-300 достигнув 322 в 2010, начиная с 2012 демонстрирует устойчивый рост. Мы предполагаем, что этот рост связан с эффектом программ по привлечению ведущих зарубежных ученых и стимулирования публикационной активности, отражаемой в международных базах данных.

**Таблица 10 – Количество авторов с международной синхронной мобильностью в предлагаемой классификации**

| Годы | FA | MA | RA | SA | Всего |
|---|---|---|---|---|---|
| 2000 | 599 | 2420 | 3192 | 20 | 6231 |
| 2001 | 553 | 2521 | 3316 | 19 | 6409 |
| 2002 | 655 | 2571 | 3530 | 13 | 6769 |
| 2003 | 390 | 2438 | 3505 | 14 | 6347 |
| 2004 | 344 | 2465 | 3678 | 24 | 6511 |
| 2005 | 452 | 2511 | 3912 | 22 | 6897 |
| 2006 | 427 | 2551 | 3900 | 10 | 6888 |
| 2007 | 544 | 2573 | 3991 | 18 | 7126 |
| 2008 | 296 | 2633 | 4124 | 18 | 7071 |
| 2009 | 364 | 2644 | 4196 | 10 | 7214 |
| 2010 | 322 | 2697 | 4367 | 6 | 7392 |
| 2011 | 448 | 2834 | 4470 | 15 | 7767 |
| 2012 | 391 | 2866 | 4510 | 24 | 7791 |
| 2013 | 436 | 2938 | 4744 | 19 | 8137 |
| 2014 | 606 | 3041 | 4945 | 38 | 8630 |
| 2015 | 777 | 3146 | 5155 | 45 | 9123 |
| 2016 | 1072 | 3236 | 5202 | 82 | 9592 |
| 2017 | 1146 | 3208 | 5266 | 94 | 9714 |
| 2018 | 1186 | 3077 | 5043 | 131 | 9437 |

Дополнительно были проанализированы российские организации, которые указывали «зарубежные» авторы (FA) в качестве дополнительной аффилиации. В качестве базы для анализа были выбраны публикации за 2014-2018 так как в этот период уже в полной мере действовали все факторы, связанные с государственной научной политикой по привлечению ведущих зарубежных авторов. В целях снижения ошибок, связанных с неточностью данных в Scopus, были выбраны авторы с более чем одной публикацией за весь период 2000-2018. Мы насчитали 2253 таких автора, публиковавшихся с российскими аффилиациями в 2014-2018, в то время как общее количество таких авторов за весь период 2000-2018 составило 7284. Наиболее активные в привлечении зарубежных авторов организации показаны в таблице 11.

**Таблица 11 – Топ-20 организаций по количеству публикаций FP в 2018 г.**

| Организация | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|---|---|---|
| ИТМО | 4 | 11 | 7 | 44 | 152 | 254 | 299 | 394 | 438 |
| РУДН | 3 | 1 | 3 | 1 | 6 | 7 | 13 | 108 | 252 |
| СПбГУ | 63 | 78 | 87 | 99 | 109 | 155 | 179 | 198 | 236 |
| МГУ | 152 | 180 | 207 | 226 | 216 | 199 | 220 | 206 | 198 |
| ТГУ | 11 | 13 | 20 | 17 | 92 | 146 | 180 | 147 | 198 |
| КФУ | 18 | 24 | 25 | 37 | 69 | 142 | 145 | 185 | 192 |
| ВШЭ | 8 | 8 | 28 | 30 | 64 | 91 | 129 | 162 | 179 |
| УрФУ | 2 | 5 | 9 | 16 | 32 | 56 | 79 | 102 | 162 |
| ТПУ | 4 | 1 | 7 | 7 | 43 | 51 | 68 | 124 | 156 |
| МИФИ | 9 | 16 | 20 | 21 | 29 | 54 | 103 | 146 | 152 |
| ОИЯИ | 77 | 87 | 121 | 123 | 121 | 117 | 94 | 117 | 147 |
| МИСИС | 4 | 4 | 13 | 15 | 56 | 85 | 121 | 125 | 141 |
| 1 Мед | 2 | 2 | 2 | 2 | 3 | 2 | 10 | 31 | 140 |
| МФТИ | 13 | 13 | 21 | 51 | 77 | 85 | 77 | 88 | 123 |
| Сколково | 0 | 0 | 5 | 17 | 44 | 61 | 68 | 64 | 115 |
| ННГУ | 10 | 9 | 8 | 33 | 63 | 73 | 92 | 76 | 72 |
| НГУ | 4 | 5 | 7 | 10 | 19 | 42 | 59 | 40 | 71 |
| ФТИ им. А.Ф. Иоффе | 104 | 107 | 110 | 110 | 93 | 90 | 94 | 79 | 63 |
| ФИАН | 66 | 62 | 62 | 72 | 41 | 65 | 49 | 57 | 61 |
| ИППИ РАН | 20 | 27 | 24 | 28 | 38 | 43 | 26 | 46 | 53 |

Проведенный анализ позволяет предположить, что мы имеем дело с довольно сложным и разноплановым явлением, оказывающим заметный эффект на результативность и перспективы развития российской науки. Анализ наиболее активных авторов, публикующих свои результаты с указанием российской аффилиации как дополнительной (FA), представленный в таблице 12, показывает, что необходимо более детальное изучение этого явления. Большей глубины понимания может помочь достичь международный контекст будущего анализа.

**Таблица 12 – Топ-10 наиболее результативных зарубежных авторов (FA) за 2014-2018 гг.**

| Зарубежная аффилиация | Российская аффилиация | Автор | Публикации | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | IP | FP | RP | SP | Ито-го | Доля FP |
| University of Cambridge | Special Astrophysical Observatory, Russian Academy of Sciences | Stolyarov V. | 85 | 3 | | | 88 | 3% |
| University of Atlanta | Institute of Physiologically Active Compounds of the Russian Academy of Science | Aliev G. | 41 | 28 | 3 | 1 | 73 | 38% |
| INRS-Energy, Materials and Telecommuni-cations | Saint Petersburg National Research University of Information Technologies, Mechanics and Optics University ITMO | Morandotti R. | 4 | 66 | | | 70 | 94% |
| University of Houston | Saratov State Universit | Larin K. | 4 | 63 | 1 | 1 | 69 | 91% |
| University of the Free State, South Africa | Saint Petersburg National Research University of Information Technologies, Mechanics and Optics University ITMO | Finkelstein M. | 1 | 64 | 1 | 1 | 67 | 96% |
| Ospedale Vito Fazzi, Italy | Moscow Institute of Physics and Technology | Tinelli A. | 11 | 36 | 2 | 9 | 58 | 62% |
| Panepistimio Thesalias, Greece | Institute of High Temperature Electrochemistry of the Ural Branch of the Russian Academy of Sciences | Tsiakaras P. | 14 | 18 | 25 | 1 | 58 | 31% |
| Karolinska Institutet, Sweden | Lomonosov Moscow State University | Zhivotovsky B. | 9 | 18 | 28 | 2 | 57 | 32% |
| The University of Warwick, UK | Pulkovo Observatory of the Russian Academy of Sciences | Nakariakov V. | 12 | 35 | 2 | 4 | 53 | 66% |
| Carleton University, Canada | Tomsk State Universit | Ernst R.E. | 10 | 40 | 1 | | 51 | 78% |

## 2.9 Национальная оценка результативности научных организаций – проблемы и недостатки

В 2017-2018 гг. авторы участвовали в разработке методики, проведении расчетов и анализа результатов количественного этапа ведомственной оценки результативности научных организаций, подведомственных ФАНО России [20]. Разработанные модификации в методику оценки позволили обоснованно уточнить пороговые значения, учесть специфику деятельности академических научно-исследовательских институтов при выборе профиля, по которому проводилась оценка и добиться более устойчивых результатов (рис. 14).

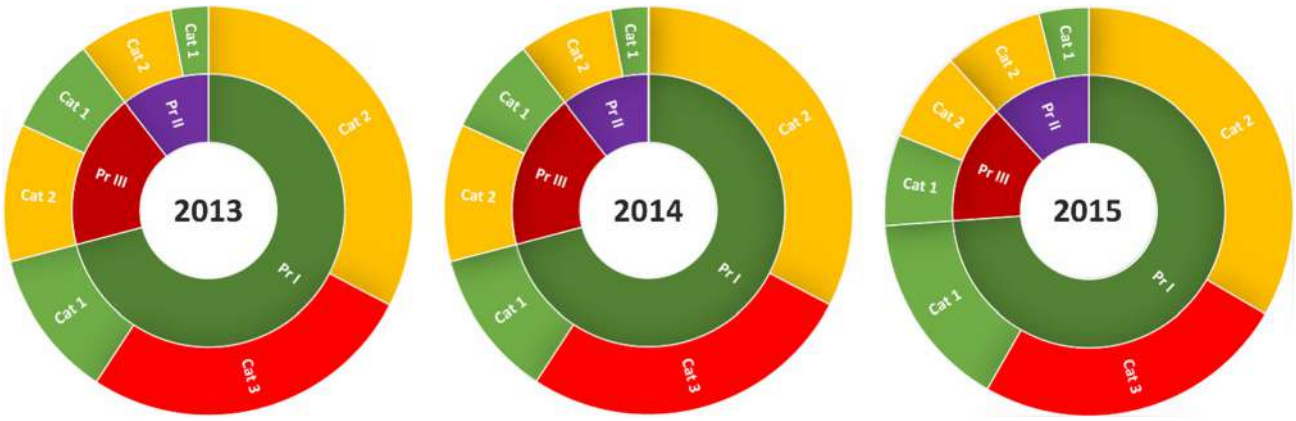


**Рисунок 14 – Распределение оцениваемых организаций по профилям деятельности и категории результативности по результатам количественной оценки.**

В 2017 г. был завершен экспертный этап оценки результативности, что позволило нам сравнить полученные результаты с результатами количественного этапа. Общее распределение оцениваемых организаций по профилям деятельности и категориям результативности показано на рис. 15.

За счет экспертной оценки произошло заметное перераспределение организаций по категориям результативности, что видно в таблице 13. Нетрудно заметить, что эти изменения носили разнонаправленный характер в разных направлениях наук. Так, в естественных, социальных и гуманитарных науках у значительной части организаций на основе экспертной оценки была повышена категория результативности, а в технических – понижена.

**Рисунок 15 – Итоговое распределение оцениваемых организаций по профилям деятельности и категории результативности по результатам ведомственной оценки.**

**Таблица 13 – Сопоставление результатов количественного этапа и финальной оценки по распределению организаций по категориям в разных отраслях наук.**

| Направления | 2013 - 2015 среднее распределение по категориям | | | Финальное распределение по категориям | | |
|---|---|---|---|---|---|---|
| | Cat 1 | Cat 2 | Cat 3 | Cat 1 | Cat 2 | Cat 3 |
| Естественные науки | 29% | 49% | 22% | 45% | 45% | 11% |
| Технические науки | 36% | 46% | 18% | 28% | 48% | 23% |
| Социальные науки | 14% | 44% | 42% | 20% | 49% | 31% |
| Гуманитарные науки | 11% | 49% | 40% | 56% | 41% | 4% |
| Медицина | 30% | 49% | 20% | 21% | 59% | 21% |
| Сельскохозяйственные науки | 18% | 55% | 27% | 20% | 43% | 37% |

Объемы изменений, внесенные экспертной оценкой в окончательный результат оценки, показывают, что, вероятно, выбранные данные и методика, несмотря на все попытки, недостаточно адекватно отражают результативность научных организаций. Среди основных проблем можно выделить следующие:

- На результаты оценки значительно повлияла недостаточная степень достоверности данных. Выборочный анализ показал наличие значительных искажений в показателях публикационной активности некоторых организаций при высокой чувствительности методики к таким ошибкам. Еще большее влияние на результаты оказывает искажение данных по численности научных работников. Ряд организаций предоставлял некорректные данные в связи с недостаточным уровнем понимания как самого термина «научные работники», так и необходимости указывать значения в эквиваленте полных штатных единиц. Нельзя сбрасывать со счетов также возможность сознательной манипуляции количеством штатных единиц за счет дробления ставок. Расчеты по высшим учебным заведениям показывают еще большие проблемы с этим показателем, так как научной работой занимается также профессорско-

преподавательский состав, при этом особенности форм отчетности позволяют исключить их из введенных значений.

- Значительное влияние оказывает нарастающий тренд на межорганизационные коллаборации при проведении научных исследований. Применяемый в методике целый счет дает преимущество активно сотрудничающим с внешними партенрами организациям.
- В методике применяется принцип определения количества публикаций, индексируемых в международных базах данных как максимума из значений по Web of Science и по Scopus. При этом заметное количество публикаций российских исследователей индексируется только в одной из этих систем. В связи с этим проигрывают организации, авторам которых не повезло выбрать для своих публикаций как журналы, индексируемые только в WoS, так и только в Scopus.
- Распределение организаций по референтным группам приводит к образованию ряда групп с малым числом членов, что, в связи с особенностями методики приводит к недостоверным результатам.

## 2.10 Обоснование необходимости перехода на фракционный счет при оценке публикационной активности организаций

Наблюдаемый в последние годы активный рост коллабораций между российскими научными организациями, частично выражающийся в синхронной мобильности авторов, т.е. указании ими нескольких аффилиаций в своих публикациях оказывает все возрастающее влияние на результативность отдельных организаций и российской науки в целом. Принятый при оценке результативности научных организаций целый счет статей в сочетании с растущими коллаборациями приводит к «антисинергетическому эффекту» – увеличению разницы между суммарной результативностью организаций и общей национальной результативностью (рис. 16).

Для анализа этого эффекта мы выполнили расчеты результативности ведущих российских организаций методами целого и фракционного счета и коэффициента ранговой корреляции между ними. При фракционном счете количества публикаций организации учитывались только российские аффилиации. Таким образом суммарная результативность всех российских организаций, рассчитанная таким методом, равняется национальной результативности.

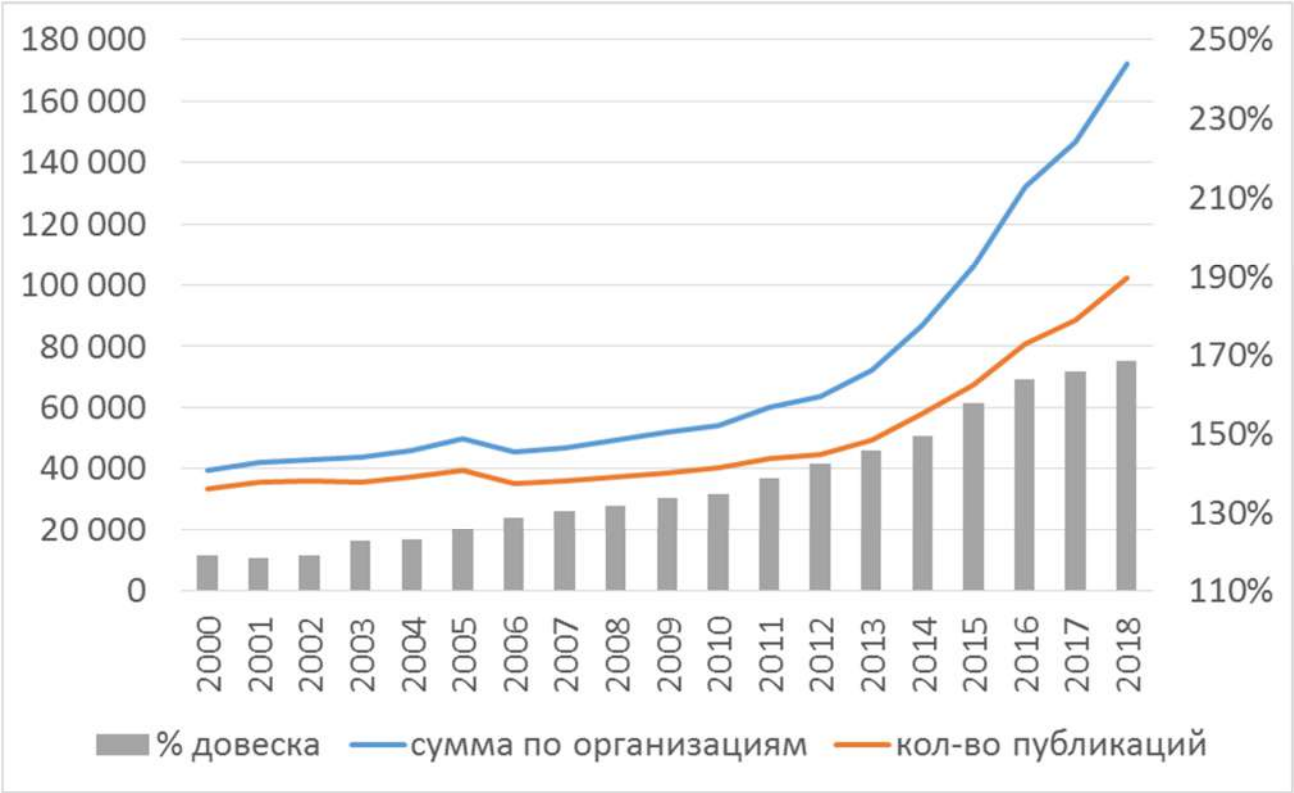


**Рисунок 16 – Увеличение разрыва между национальной результативностью российской науки и суммарной результативностью организаций при целом счете статей.**

Для наглядности была выполнена группировка ведущих организаций в три категории:

1. 5топ100 – университеты, участвующие в проекте повышения конкурентоспособности российских вузов, целью которого является вхождение 5 российских университетов в сотню ведущих мировых научно-образовательных центров 5top100. На весь проект ежегодно выделялось дополнительное финансирование около 10 млрд. руб. (140 млн евро), которые распределялись в зависимости от достигнутых результатов. По этой причине администрация и научные сотрудники этих университетов работали в условиях сильного «наукометрического прессинга». Количественный результат превзошел многие ожидания – в 2016 г. этими организациями было опубликовано в 5 раз больше проиндексированных в Scopus статей, чем в 2010-ом [38]. В 2018 г. наибольшее число таких публикаций было у Новосибирского государственного университета (2839), наименьшее – у Тюменского государственного университета (364).

2. Институты – 100 ведущих академических НИИ. После реформы государственных академий наук и смены учредителя у более чем 700 научных организаций (которым в 2013

г. стало Федеральное агентство научных организаций) драматически изменился стиль управления научными исследованиями, который стал существенно ориентироваться на количественные показатели. Так в 2015 году в государственное задание был введен обязательный показатель – количество статей, которые должны быть опубликованы в отчетном году и проиндексированы в Web of Science, Scopus или в Российском индексе научного цитирования. В 2015 году была объявлена, а 2017-2018 году – проведена оценка научной результативности с учетом наукометрических показателей. Из 454 прошедших проверку организаций к первой категории отнесено 142, ко второй — 205, к третьей — 107. Попавшие в первую категорию могли претендовать на дополнительное финансирование, а для третьей категории разрабатывались планы ликвидации или присоединения к другим организациям. «Наукометрический прессинг» для этих также присутствовал, но он был менее выражен и начался на 2-3 года позже, чем для группы ведущих университетов. В эту группу были включены 100 научных организаций, у которых в 2018 г. было наибольшее количество публикаций, проиндексированных в Scopus (min 126, max 1493).

3. Вузы. В этой группе оказались организации разного профиля: федеральные и региональные университеты технической, медицинской, педагогической, строительной и иной направленности, не включенные в группу ведущих университетов. До 2016 г. к ним не предъявлялось формальных требований к повышению наукометрических показателей, хотя высокая конкуренция за дополнительное финансирование в научных фондах к этому располагало. В 2016-2017 г. на основании конкурса были определены 33 опорных вуза, относящиеся к разным регионам и получившие дополнительную финансовую поддержку. Таким образом, эти организации позже всех и в наименьшей степени столкнулись с системным «наукометрическим прессингом». В эту группу мы включили 100 вузов, у которых в 2018 г. было наибольшее количество публикаций, проиндексированных в Scopus (min 91, max 1406).

Результаты ранговой корреляции, представленные на рисунке 17, показывают, что она падает быстрее всего у организаций, находящихся под наибольшим «наукометрическим» давлением, которое заставляет эти организации искать пути интенсификации научных исследований и повышения результативности всеми возможными способами, в том числе и за счет внешних коллабораций. В работе [38] мы уже показывали, что зачастую активно применяется также «привлечение» сторонних исследователей на условиях внешнего совместительства. Предположительно, научные исследования при этом выполняются по месту основной работы этих ученых, а дополнительная аффилиация ставится для получения денежного вознаграждения.

**Рисунок 17 – Динамика значения коэффициента ранговой корреляции между результативностью организаций в трех группах, рассчитанной методами целого и фракционного счета.**

Отношение результативности организации, рассчитанное фракционным счетом с учетом только российских аффилиаций авторов к значению, рассчитанному целым счетом изменяется в пределах от 0 до 1 и показывает коэффициент «самодостаточности» организации – в какой мере она зависит от других участников научных коллабораций. Результаты расчета этого соотношения для ведущих российских университетов показаны на рисунке 18 в динамике. Можно заметить, что значения снижаются во времени для всех университетов из этой выборки, что говорит о системной тенденции. Скорость снижения варьируется, что также связано с разными возможностями и стратегиями вхождения в коллаборации.

Самое большое снижение зафиксировано у трех университетов с ярко выраженным физическим профилем (МИФИ - National Research Nuclear University, МФТИ - Moscow Institute of Physics & Technologies, НГУ - Novosibirsk State University). Причем последний всегда тесно сотрудничал с научными институтами Новосибирского Академгородка, тогда как первые два стали активнее вступать в коллаборации после 2009 года. В данной выборке именно эти три организации внесли наибольший вклад в уменьшение корреляции между результатами целого и фракционного счета. Это говорит о необходимости проведения аналогичного анализа в разрезе областей наук.

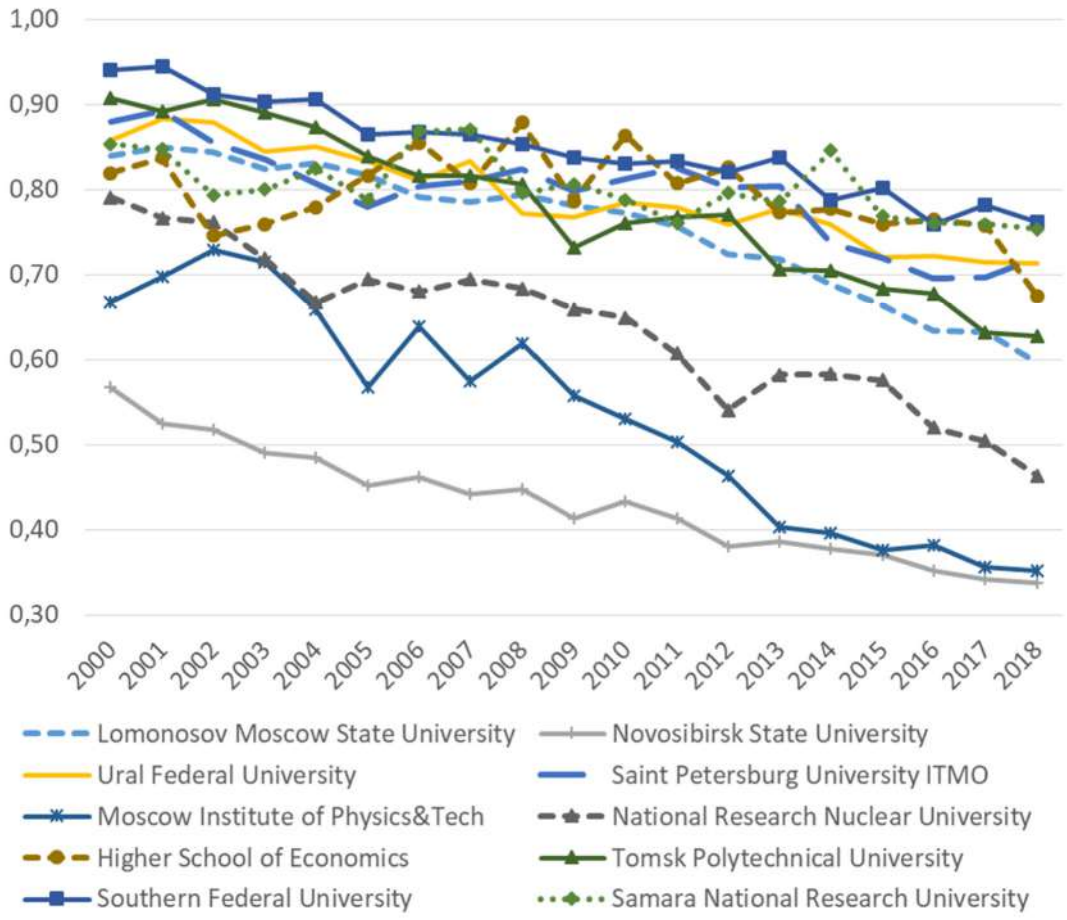


**Рисунок 18 – Динамика отношения результативности организации, рассчитанной фракционным счетом с учетом только российских аффилиаций авторов к значению, рассчитанному целым счетом для ведущих российских вузов.**

Переход на использование фракционного счета при оценке результативности, с одной стороны позволит более адекватно учесть вклад каждой организации в научные исследования, с другой – значительно изменит результаты оценки и результирующее ранжирование организаций. При этом ряд авторов, исследующих фракционный счет, высказывают сомнения в адекватности оценки таким методом вклада отдельного автора, особенно это касается статей с большим количеством авторов, для которых вклад основных авторов может быть «размыт». Фактически, фракционный счет уже применяется Министерством науки и высшего образования РФ без всестороннего обсуждения с научной общественностью и профессиональным наукометрическим сообществом. Последствия перехода на фракционный счет в оценке результативности и возможное влияние этого перехода на развитие научных организаций необходимо изучать дополнительно.

# 3 АНАЛИЗ ТЕРРИТОРИАЛЬНОГО РАСПРЕДЕЛЕНИЯ И БИБЛИОМЕТРИЧЕСКОЕ ИССЛЕДОВАНИЕ ПУБЛИКАЦИОННОЙ АКТИВНОСТИ ЧЛЕНОВ РЕДКОЛЛЕГИЙ НАУЧНЫХ ЖУРНАЛОВ В ОБЛАСТИ ИНФОРМАЦИОННЫХ И БИБЛИОТЕЧНЫХ НАУК, ИСТОРИИ И АРХЕОЛОГИИ.

Анализ состава редакционных коллегий и их научной деятельности в последние годы привлекает все большее внимание специалистов по информационному делу и участников издательского процесса. Результаты исследований по географическому распределению членов редколлегий научных журналов, сравнение этих данных с географическим распределением авторов, сопоставление показателей научной продуктивности редакторов с рейтинговыми показателями самого издания позволяют с новых позиций определять ценность периодического издания, его вовлеченность в мировое научное пространство, а также дают возможность скорректировать научную политику журнала, направленную на расширение его читательской аудитории и повышение качества публикуемых материалов. Кроме того, данные по редколлегиям представляют дополнительные критерии оценки научной отрасли в целом.

## 3.1 Анализ территориального распределения и библиометрическое исследование публикационной активности членов редколлегий научных журналов в области информатики, библиотековедения и наук о Земле

Среди разнообразных фактологических основ для исследования научных журналов и отрасли в целом, включающих отдельные кластеры публикаций, ученых или журналы в целом, оцениваемых с позиций библиометрии, научной экспертизы или с применением формализованных показателей, значительно меньшей популярностью пользуются редакционные коллегии периодических изданий, которые могут анализироваться с точки зрения географической представленности участников редколлегий, библиометрического анализа их публикационной активности, а также гендерного распределения. Значимость данного подхода обусловлена тем, что главные редакторы и члены редакционных коллегий научных журналов играют одну из ключевых ролей в развитии науки и поддержании в ней принципов добросовестности, редколлегии считаются влиятельным органом принятия решений в науке, а редакторы по праву являются наиболее авторитетными учеными.

В рамках задач по проекту в 2019 г. было проведено библиометрическое исследование отечественных журналов по библиотековедению, информатике и наукам о Земле, публикационной активности членов их редколлегий и выявлено их географическое

распределение, что в совокупности дает дополнительные эффективные подходы к общей оценки отрасли, а также может стать полезной информацией при решении вопросов редакционной политики в журналах исследуемой области.

### 3.1.1 Географическое распределение членов редколлегий журналов по информатике и библиотековедению и выявление экспертных центров

В 22 отобранных для анализа журналах по информационно-библиотечной тематике числится 478 членов редакционных коллегий. Наименьшее число участников редколлегии – 12 в журнале «Вестник Московского городского педагогического университета. Серия: информатика и информатизация образования», наибольшее – 51 в издании «Бизнес-информатика», при среднем значении по всем журналам в 25 человек. В табл. 14 приводятся основные данные по географическому распределению членов редакционных коллегий.

Члены редколлегий представляют 33 государства, где в первую пятерку входят Россия – 79,3 % (379 экспертов), Германия – 3,1 %, США – 2,3 %, Казахстан – 1,7 % и Белоруссия – 1,5 %. На долю остальных стран приходится 12,1 %. Примечательна относительно небольшая интеграция со странами СНГ, которая уступает сотрудничеству с организациями западных государств. С учетом не всегда высоких показателей российских журналов на международном уровне можно предположить, что участие многих зарубежных специалистов в работе отечественных журналов носит формальный характер. Исключительно российских представителей в редколлегиях имеют лишь четыре журнала.

Из российских 36 городов, в организациях которых работают члены редколлегий, лидирует столица (60 % экспертов), за которой следуют Санкт-Петербург (9 %), Воронеж и Новосибирск (по 7 %) и Самара (3 %). Иное распределение по значимости городов как центров сосредоточения экспертных сообществ получается при сопоставлении городов по числу членов редакционных коллегий в расчете на организацию, где наибольшая концентрация членов редколлегий наблюдается уже не в столице, а в региональных центрах (табл. 15).

Всего российские члены редакционных коллегий исследуемых журналов работают в 168 организациях, при этом половина из них является сотрудниками лишь 21 организации, которые ввиду высокого сосредоточения членов редколлегий можно считать основными организациями в отрасли (табл. 16).

**Таблица 14 – Географическое распределение участников редакционных коллегий в журналах по библиотековедению и информатике.**

| Название журнала | Число стран членов редколлегии | Число членов редколлегии | Число и доля зарубежных участников (%) |
|---|---|---|---|
| Библиография. Научный журнал по библиографоведению, книговедению и библиотековедению | 6 | 21 | 5 (23,8) |
| Библиосфера | 4 | 21 | 3 (14,3) |
| Библиотековедение | 7 | 41 | 7 (17,1) |
| Бизнес-информатика | 11 | 51 | 19 (37,3) |
| Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии | 4 | 26 | 4 (15,4) |
| Вестник Московского городского педагогического университета. Серия: Информатика и информатизация образования | 2 | 12 | 1 (8,3) |
| Вестник Новосибирского государственного университета. Серия: Информационные технологии | 7 | 27 | 6 (22,2) |
| Вычислительные технологии | 11 | 35 | 15 (42,9) |
| Информатика и ее применения | 9 | 36 | 11 (30,6) |
| Информатика и образование | 1 | 18 | 0 |
| Информационное общество | 1 | 33 | 0 |
| Информационные ресурсы России | 3 | 18 | 2 (11,1) |
| Информационные технологии и вычислительные системы | 1 | 18 | 0 |
| Научно-техническая информация. Серия 1: Организация и методика информационной работы | 6 | 14 | 5 (35,7) |
| Научно-техническая информация. Серия 2: Информационные процессы и системы | 2 | 17 | 1 (5,9) |
| Научные и технические библиотеки | 4 | 19 | 3 (15,8) |
| Онтология проектирования | 5 | 28 | 5 (17,9) |
| Прикладная информатика | 3 | 16 | 3 (18,8) |
| Проблемы информационной безопасности. Компьютерные системы | 9 | 27 | 8 (29,6) |
| Программная инженерия | 4 | 37 | 3 (8,1) |
| Системы и средства информатики | 6 | 24 | 6 (25,0) |
| Системы управления и информационные технологии | 1 | 22 | 0 |

**Таблица 15 – Географические центры с наиболее значимыми экспертными группами в области библиотековедения и информатики.**

| Город | Число | | |
|---|---|---|---|
| | членов редколлегий | организаций | членов редколлегий в расчете на организацию |
| 1. Воронеж | 27 | 7 | 3,9 |
| 2. Новосибирск | 27 | 8 | 3,4 |
| 3. Санкт-Петербург | 35 | 12 | 2,9 |
| 4. Самара | 11 | 4 | 2,8 |
| 5. Москва | 226 | 92 | 2,5 |
| 6. Прочие | 53 | 45 | 1,2 |

**Таблица 16 – Основные организации, в которых работают члены редакционных коллегий российских журналов в области библиотековедения и информатики.**

| Организация | Город | Число членов редколлегий |
|---|---|---|
| ФИЦ «Информатика и управление» РАН | Москва | 32 |
| Московский государственный университет им. М.В. Ломоносова | Москва | 19 |
| Национальный исследовательский университет «Высшая школа экономики» | Москва | 15 |
| Воронежский государственный университет | Воронеж | 13 |
| Российская государственная библиотека | Москва | 13 |
| Институт вычислительных технологий СО РАН | Новосибирск | 10 |
| Воронежский государственный технический университет | Воронеж | 9 |
| Санкт-Петербургский политехнический университет Петра Великого | Санкт-Петербург | 9 |
| Государственная публичная научно-техническая библиотека России | Москва | 8 |
| Институт проблем управления РАН | Москва | 8 |
| Московский городской педагогический университет | Москва | 7 |
| Государственная публичная научно-техническая библиотека СО РАН | Новосибирск | 6 |
| Российский государственный гуманитарный университет | Москва | 6 |
| Самарский национальный исследовательский университет им. С.П. Королева | Самара | 6 |
| Всероссийский институт научной и технической информации РАН | Москва | 5 |
| Научный и издательский центр «Наука» РАН | Москва | 5 |
| Санкт-Петербургский государственный университет | Санкт-Петербург | 5 |
| Санкт-Петербургский государственный институт культуры | Санкт-Петербург | 5 |
| Новосибирский государственный университет | Новосибирск | 4 |
| Российская академия народного хозяйства и государственной службы | Москва | 4 |
| Санкт-Петербургский институт информатики и автоматизации РАН | Санкт-Петербург | 4 |

Вызывает интерес распределение членов редколлегий по организациям в зависимости от их ведомственной принадлежности (табл. 17). В выявленном распределении видно, что сотрудники учебных заведений играют более заметную роль в работе периодических изданий, существенно опережая научно-исследовательские организации.

**Таблица 17 – Распределение членов редакционных коллегий российских журналов по библиотековедению и информатике в зависимости от ведомственной принадлежности организации – места работы экспертов.**

| Ведомство | Число членов редколлегий | Доля членов редколлегий, % |
|---|---|---|
| Министерство науки и высшего образования (вузы) | 179 | 47,2 |
| Министерство науки и высшего образования (научно-исследовательские организации) | 132 | 34,8 |
| Бюджетные организации (включая Министерство культуры РФ) | 51 (из них 29 – Минкульт) | 13,5 |
| Коммерческие организации и некоммерческие образования | 17 | 4,5 |

### 3.1.2 Библиометрический анализ публикационной активности членов редколлегий журналов по информатике и библиотековедению

На рис. 19-20 приводятся основные данные по научной продуктивности журналов и членов их редакционных коллегий и демонстрируется существенный вклад работы редакционной коллегии в развитие журналов по библиотековедению и информатике и влияние на рейтинг для большинства анализируемых изданий.

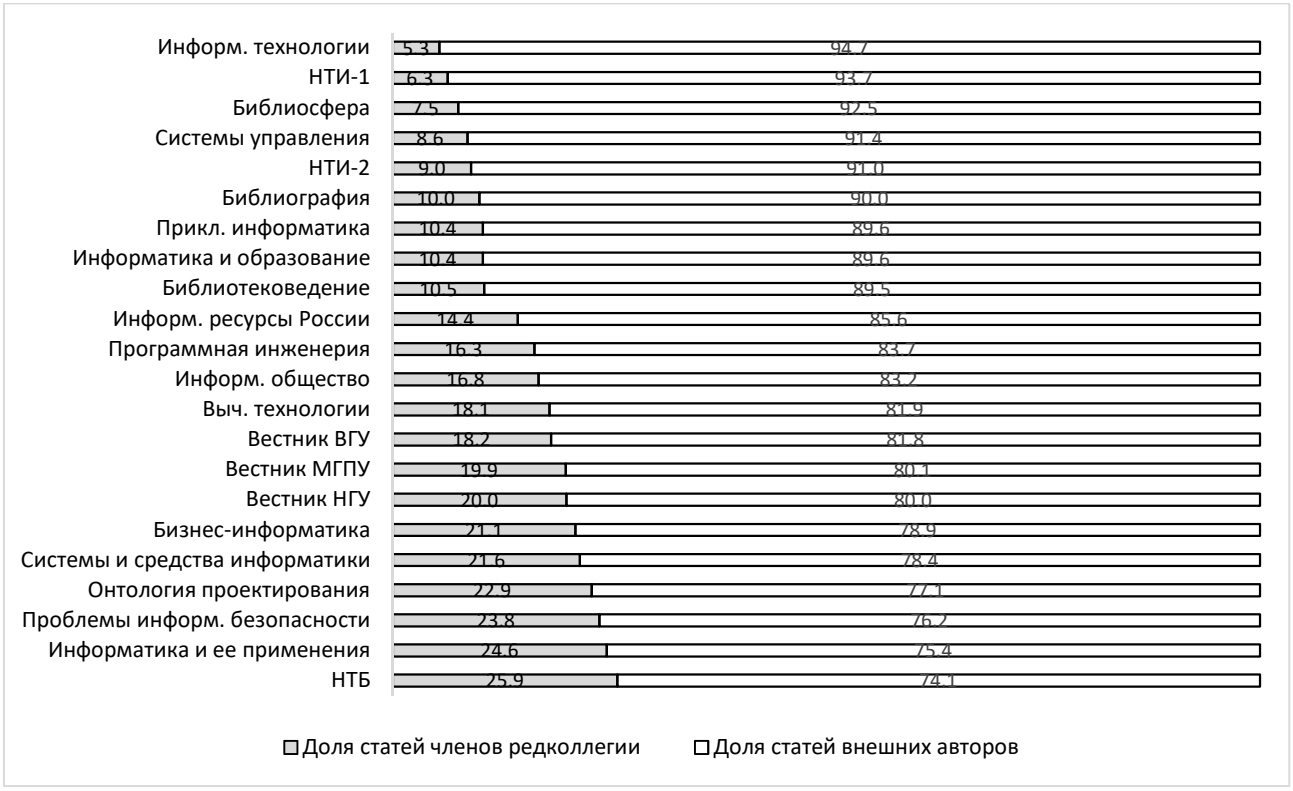


**Рисунок 19 – Соотношение доли публикаций членов редколлегии и доли публикаций внешних авторов в журналах по библиотековедению и информатике.**

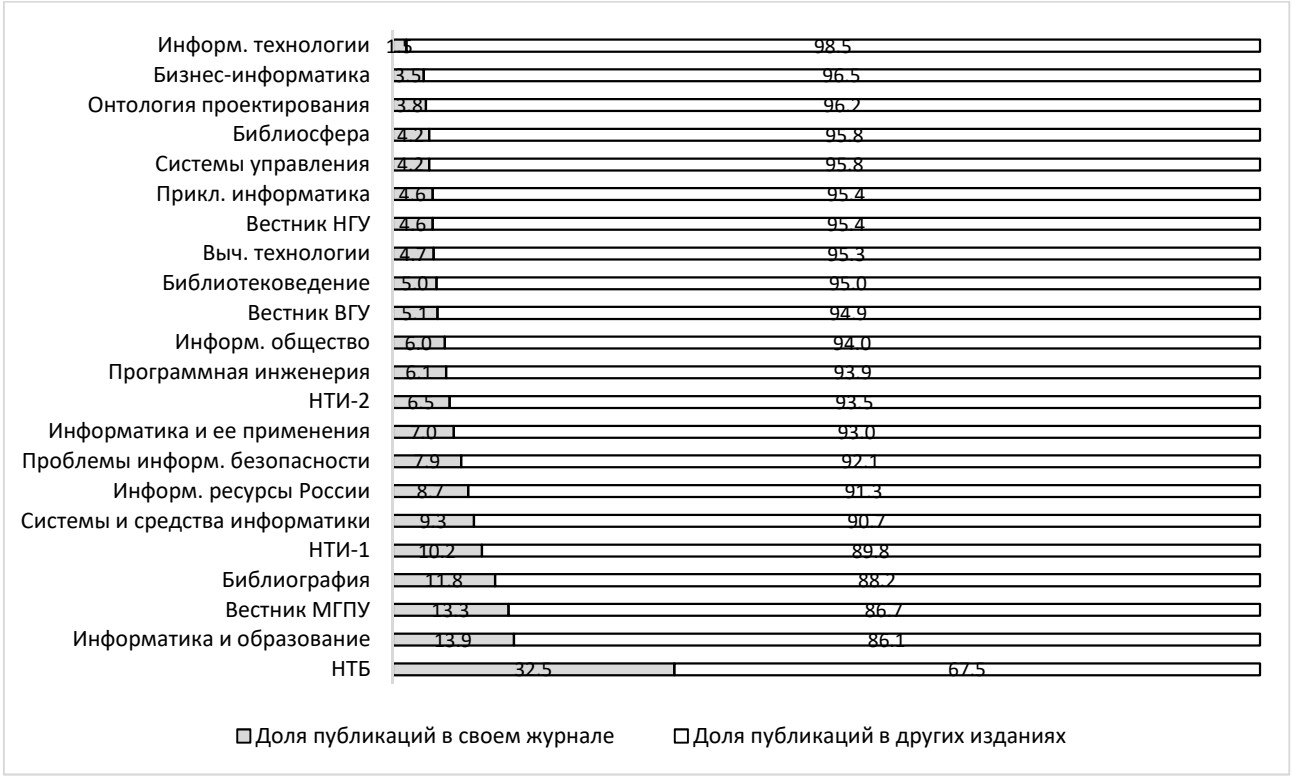

**Рисунок 20 – Соотношение доли публикаций членов редколлегий журналов по библиотековедению и информатике в собственном издании и в других источниках.**

Полученные нами данные по цитируемости статей журналов и публикаций членов редколлегии также позволяют оценить вклад редакционной коллегии в развитие и рейтинг соответствующих изданий (рис. 3 и 4). Как видно из рис. 3, редакционные коллегии могут оказывать существенное влияние на библиометрические показатели журнала и его общий рейтинг, при расчете которого цитирования играют ключевую роль. Характерен пример последних трех журналов в списке на рис. 4, где публикации членов редколлегий приносят журналам более трети, а в случае с «Вестником НГУ» – почти половину всех цитирований. При этом данные рис. 21, демонстрирующие высокую долю публикаций членов редколлегии в этих трех журналах, позволяют предположить, что это сознательная политика редакции, направленная на повышение рейтинга издания.

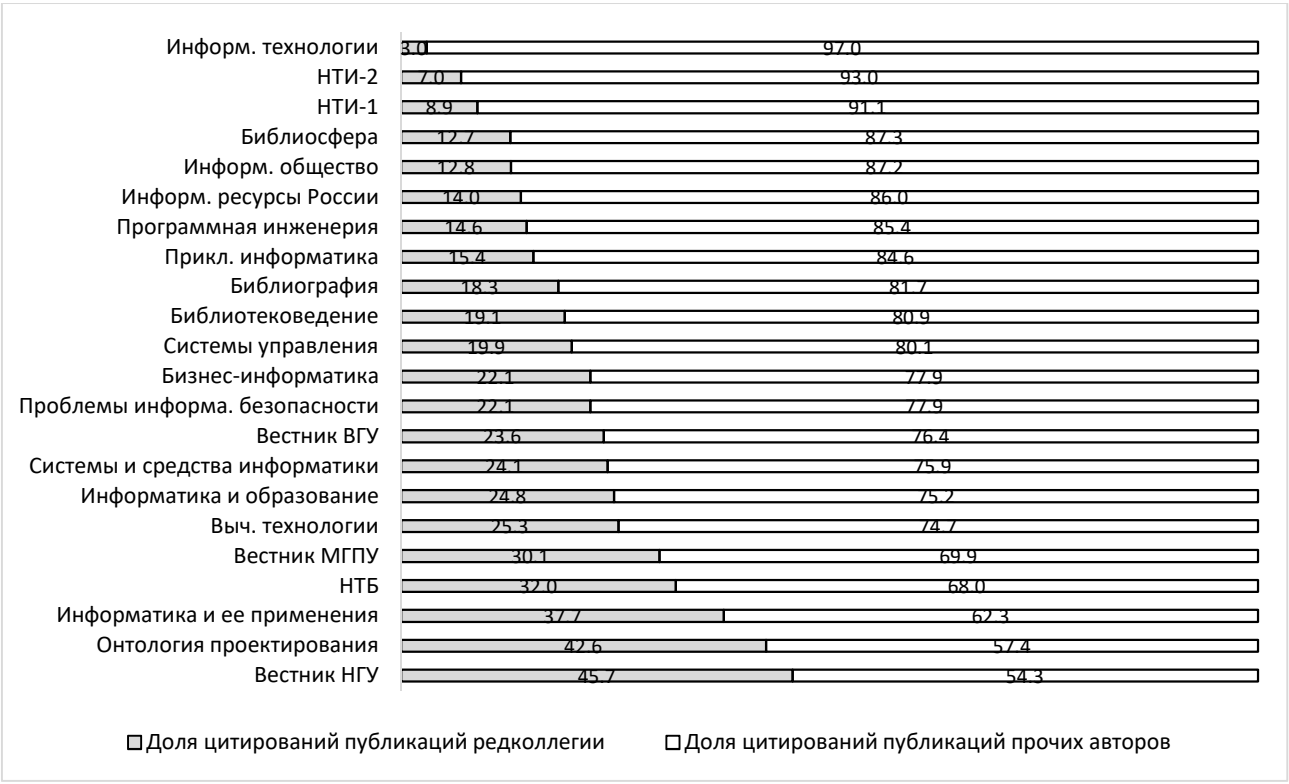


**Рисунок 21 – Соотношение цитируемости статей членов редакционных коллегий, опубликованных в соответствующих журналах по библиотековедению и информатике, и статей внешних авторов в этих журналах.**

Однако редакционная коллегия может влиять на повышение рейтинга своего журнала и иными способами помимо опубликования высококачественных статей, в частности, прибегая к обильному цитированию своего журнала. В основном это так

называемое «скрытое» самоцитирование, под которым мы понимаем цитирование членами редколлегии своего журнала в статьях, публикуемых в иных источниках. При этом такие цитирования формально считаются внешними и не относятся к самоцитированию журнала, фактически же являясь таковыми. Реже это цитирующие свой журнал статьи, публикуемые в нем же. На рис. 22 представлено, в какой мере цитирование своего журнала членами редколлегии может влиять на общую цитируемость издания.

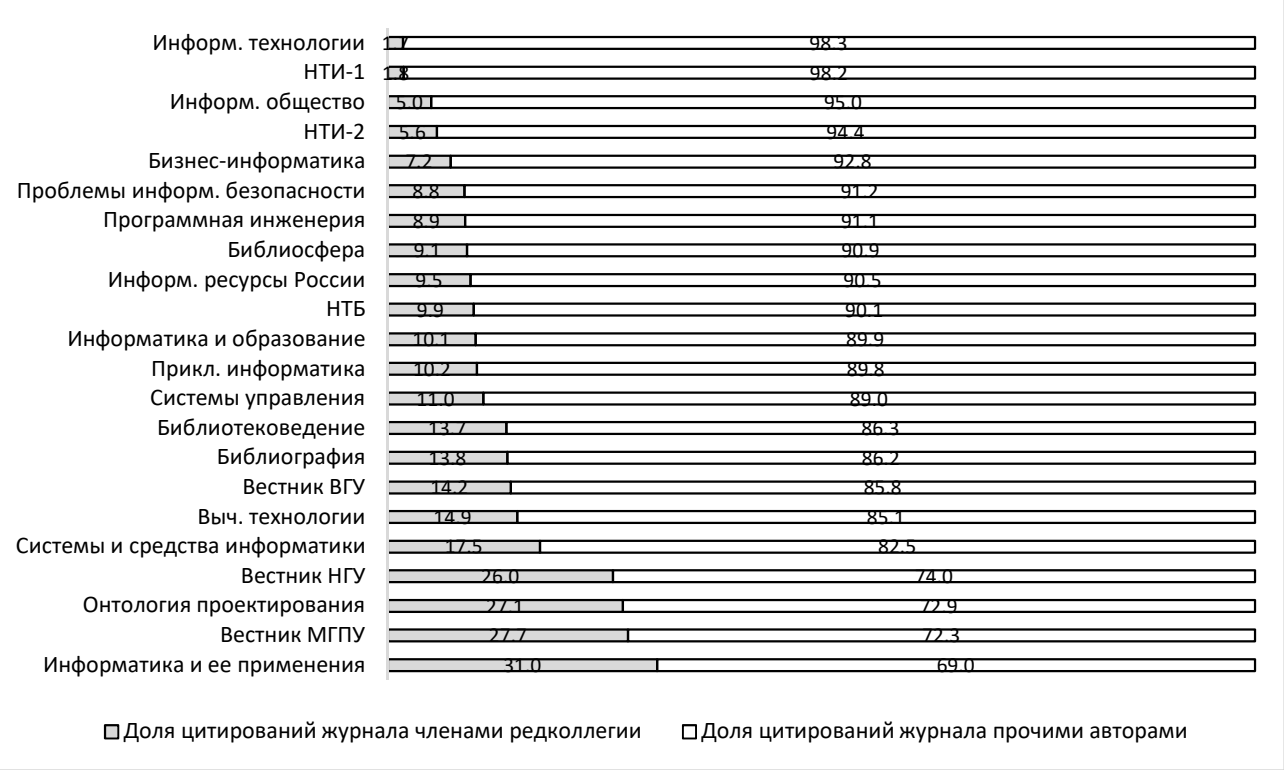


**Рисунок 22 – Цитирование журналов по библиотековедению и информатике членами их редколлегий. Показаны доля цитирований журналов членами их редакционных коллегий и доля цитирований другими авторами.**

### 3.1.3 Географическое распределение членов редколлегий журналов по наукам о Земле и выявление экспертных центров

В редакционных коллегиях 19 наиболее авторитетных российских журналов по геонаукам работают 434 человека. Наибольшее представительство (почти 87 %) ожидаемо за российскими гражданами. На долю других стран приходится всего около 13 % (табл. 18).

География российской части редколлегий анализируемых журналов представлена почти тремя десятками городов, однако определяющую роль играют только семь из них: без малого две трети членов редколлегий (63 %) в журналах по наукам о Земле представляют столичные организации, второе место занимает Новосибирск (10 %), вдвое

меньше представителей Санкт-Петербурга (5 %), затем следует Петропавловск-Камчатский (3 %) и, наконец, по 2 % членов редколлегий представляют организации Иркутска, Хабаровска и Владивостока. Такое распределение вполне отчетливо отражает географию 105 крупных научных центров и организаций выбранного нами профиля, из которых в представленной выборке половина находится в Москве, девять – в Санкт-Петербурге, шесть – в Новосибирске.

**Таблица 18 – Географическое распределение участников редакционных коллегий в журналах по наукам о Земле.**

| **Название журнала** | **Число** | | **Число и доля (%) зарубежных представителей** |
|---|---|---|---|
| | **членов редколлегии** | **стран членов редколлегии** | |
| Криосфера Земли | 24 | 7 | 11 (45,8) |
| Геология и геофизика | 41 | 7 | 12 (29,3) |
| Петрология | 18 | 5 | 5 (27,8) |
| Геоморфология | 24 | 7 | 6 (25,0) |
| Геотектоника | 27 | 6 | 5 (18,5) |
| Почвоведение | 20 | 4 | 3 (15,0) |
| Геохимия | 30 | 5 | 4 (13,3) |
| Горный журнал | 40 | 6 | 5 (12,5) |
| Палеонтологический журнал | 16 | 3 | 2 (12,5) |
| Нефтехимия | 30 | 4 | 3 (10,0) |
| Вулканология и сейсмология | 33 | 3 | 2 (6,1) |
| Литология и полезные ископаемые | 20 | 2 | 1 (5,0) |
| Геология рудных месторождений | 19 | 1 | 0 |
| Геомагнетизм и аэрономия | 28 | 1 | 0 |
| Основания, фундаменты и механика грунтов | 21 | 1 | 0 |
| Стратиграфия. Геологическая корреляция | 17 | 1 | 0 |
| Тихоокеанская геология | 21 | 1 | 0 |
| Физика Земли | 21 | 1 | 0 |
| Физико-технические проблемы разработки полезных ископаемых | 28 | 1 | 0 |

Почти 90 % членов редколлегий научных журналов в области наук о Земле, т. е. наиболее квалифицированных специалистов в этой области, связаны с РАН или с крупнейшими исследовательскими университетами (табл. 19 и 20).

**Таблица 19 – Организации, представляющие наибольшее число членов редколлегий в области наук о Земле.**

| Организация | Число членов редколлегий | Город |
|---|---|---|
| МГУ им. М.В. Ломоносова | 34 | Москва |
| Геологический институт РАН | 30 | Москва |
| Институт физики Земли им. О.Ю. Шмидта | 19 | Москва |
| Институт геохимии и аналитической химии им. В.И. Вернадского | 17 | Москва |
| Институт геологии рудных месторождений, петрографии, минералогии и геохимии | 16 | Москва |
| Палеонтологический институт имени А.А. Борисяка | 12 | Москва |
| Институт нефтегазовой геологии и геофизики им. А.А. Трофимука | 12 | Новосибирск |
| Институт горного дела им. Н.А. Чинакала | 11 | Новосибирск |
| Институт вулканологии и сейсмологии ДВО РАН | 11 | Петропавловск-Камчатский |
| Институт географии РАН | 10 | Москва |
| Институт нефтехимического синтеза им. А.В. Топчиева | 10 | Москва |
| Институт геологии и минералогии им. В.С. Соболева | 12 | Новосибирск |
| Национальный исследовательский технологический университет «МИСиС» | 8 | Москва |
| Санкт-Петербургский горный университет | 8 | Санкт-Петербург |

**Таблица 20 – Распределение членов редколлегий по ведомствам.**

| № | Ведомство | Число членов редколлегий | Доля членов редколлегий (%) |
|---|---|---|---|
| 1 | Министерство науки и высшего образования (научно-исследовательские организации) | 253 | 68,9 |
| 2 | Министерство науки и высшего образования (вузы) | 74 | 20,2 |
| 3 | Ведомственные институты | 21 | 5,7 |
| 4 | Коммерческие предприятия | 19 | 5,2 |

### 3.1.4 Библиометрический анализ публикационной активности членов редколлегий журналов по наукам о Земле

Для определения взаимосвязи рейтинга журнала и публикационной активности членов его редколлегии мы провели сравнительный анализ 3-летних индексов Хирша всех членов редколлегии и статей рассматриваемых журналов за 3 года (рис. 23), который показал, что большинство журналов демонстрирует существенную корреляцию сводного индекса Хирша своих статей и публикаций членов редколлегии.

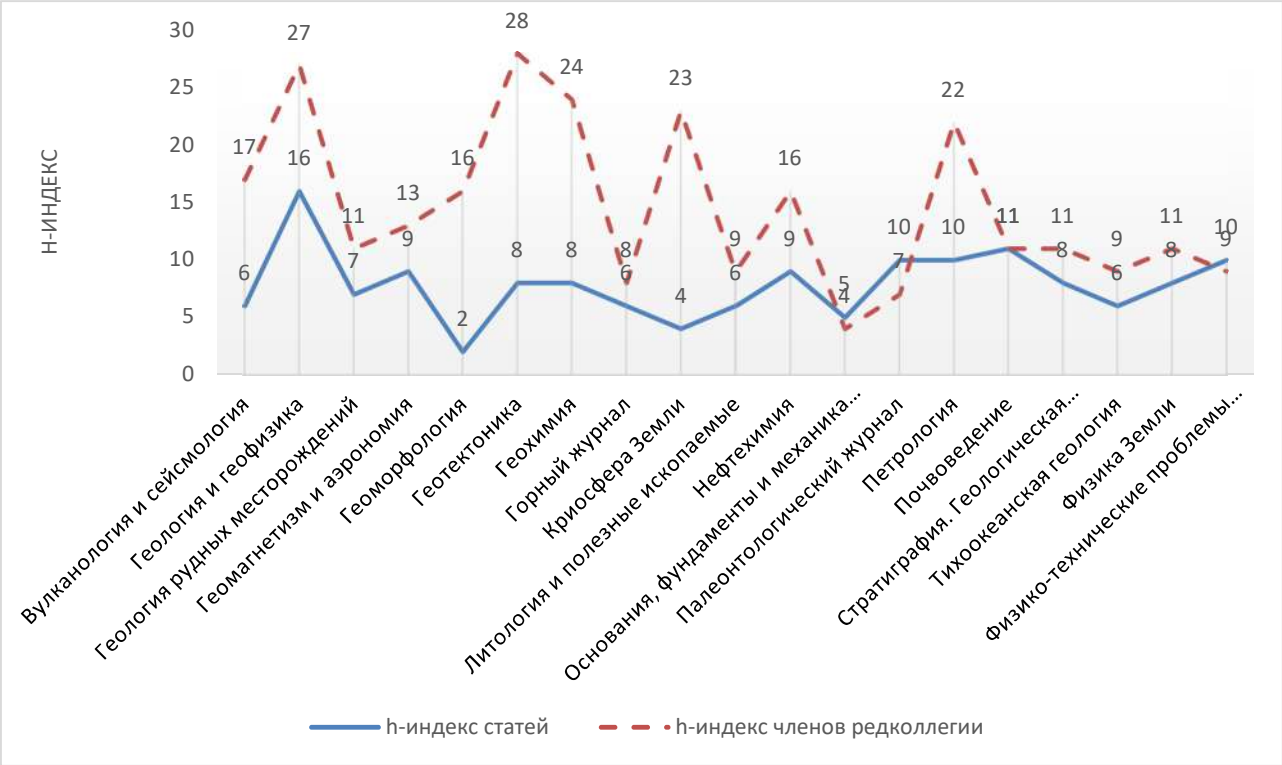


**Рисунок 23 – Сравнение показателей публикационной активности членов редколлегии с рейтинговыми показателями журнала (Scopus 2013–2015).**

Значительная корреляция данных наблюдается при сопоставили средней цитируемости одной статьи члена редколлегии со средней цитируемостью статьи журнала (рис. 24).

**Рисунок 24 – Сравнение средней цитируемости одной статьи членов редколлегии со средней цитируемостью одной статьи журнала (Scopus 2013–2015).**

Заметное исключение в обоих случаях составляет журнал «Криосфера Земли», в составе редколлегии которого больше всех зарубежных и авторитетных участников, но невысокие библиометрические показатели самих статей журнала. Схожая ситуация наблюдается у журнала «Геоморфология», в редколлегии которого также высокая доля зарубежных представителей. Объяснением может быть либо формальное привлечение зарубежных членов редколлегии, либо небольшой срок их работы в этих журналах.

Запрос по публикациям членов редколлегий показал, что в 12 из 19 анализируемых нами изданий на первом месте стоит их собственный журнал (табл. 21).

**Таблица 21 – Число членов редколлегий в собственных журналах. Жирным шрифтом отмечены журналы, которые занимают первое место по числу публикаций своих членов редакционной коллегии.**

| Название журнала | Число статей в своем журнале (2013-2015) | Доля статей в своем журнале (2013-2015) | Название журнала | Число статей в своем журнале (2013-2015) | Доля статей в своем журнале (2013-2015) |
|---|---|---|---|---|---|
| Литология и полезные ископаемые | 27 | 29,3 | Петрология | 17 | 18,3 |
| Геотектоника | 27 | 28,4 | **Стратиграфия. Геологическая корреляция** | 24 | 18,3 |
| **Геоморфология** | 31 | 26,7 | **Геомагнетизм и аэрономия** | 70 | 16,7 |
| Тихоокеанская геология | 27 | 24,3 | **Почвоведение** | 64 | 15,7 |
| **Нефтехимия** | 78 | 23,9 | **Основания, фундаменты и механика грунтов** | 19 | 12,9 |
| Геология и геофизика | 82 | 21,6 | **Криосфера Земли** | 13 | 12,6 |
| Вулканология и сейсмология | 18 | 21,2 | Геология рудных месторождений | 18 | 12,2 |
| **Физика Земли** | 48 | 20,9 | **Горный журнал** | 83 | 10,6 |
| **Физико-технические проблемы разработки полезных ископаемых** | 82 | 20,2 | **Палеонтологический журнал** | 40 | 10,1 |
| **Геохимия** | 55 | 19,6 | | | |

### 3.1.5 Библиометрический анализ публикационной активности членов редколлегий журналов по истории и археологии

В 2019 году проводилось исследование по теме «Наукометрический анализ редакционных коллегий ведущих археологических и этнографических журналов России».

Производился анализ списков участников редакционных коллегий и их научной деятельности за последние 5 лет и географическое распределение членов редколлегий журналов. В составе редколлегий был представлен цвет российской археологической науки: 10 академиков РАН, 10 член-корр. РАН, 1 академик академии наук республики Киргизстан. Высокой квалификацией обладали иностранные ученые.

**Рисунок 25 – Численность членов редколлегий по журналам**

Всего в 12 журналах археологического и этнографического профиля числится 330 членов редакционного совета и редколлегий. При этом 10 человек являются членами редколлегий двух журналов, 11 человек являются членами редколлегий 3 журналов. По одному члены редколлегий представлены в 5, 6 и 8 журналах. Это указывает на достаточно узкое профессиональнее сообщество, а с другой на интенсивную работу членов редколлегии. Наибольшее число членов редколлегии представлено журналами: «Археология этнография и антропология Евразии» и «Вестником НГУ: Серия археология и история» (рис.25).

Как видно из таблицы 22, наибольшую вклад в развитие журнала вносят члены редколлегии, работающие в организации-учредителе журнала. Около 15% публикаций в среднем дают члены редколлегии из других российских организаций, тогда как публикации иностранных членов являются большой редкостью, из чего можно сделать предварительный вывод об их весьма номинальном участии в этой работе.

**Таблица 22 – Число публикаций членов редколлегий ведущих журналов в области истории и археологии**

| Журналы | Местные | Приглашенные | Иностранные | Всего |
|---|---|---|---|---|
| Вестник НГУ: серия история | 291 | 18 | 6 | 315 |
| Археология, этнография и антропология Евразии | 294 | 66 | 5 | 365 |
| Краткие сообщения института археологии РАН | 77 | 11 | 0 | 88 |
| Вестник археологии, этнографии и антропологии | 163 | 14 | 0 | 177 |
| Теория и практика археологических исследований | 150 | 51 | 0 | 201 |
| Уральский исторический вестник | 178 | 67 | 0 | 245 |
| Известия Иркутского государственного университета: геоархеология, этнография, антропология | 47 | 36 | 0 | 83 |
| Вестник Томского государственного университета. История | 68 | 37 | 0 | 105 |
| Российская археология | 329 | 41 | 0 | 370 |
| Этнографическое обозрение | 291 | 60 | 0 | 351 |
| Археологические вести | 43 | 4 | 0 | 47 |
| Вестник Московского университета. Серия 8: История | 15 | 1 | 0 | 16 |
| Всего | 1946 | 406 | 11 | 2363 |

Распределение цитируемых публикаций между авторами и членами редакционных коллегий, как показано в табл. 23, оказалось неоднозначным, с существенными перекосами как в одну, так и в другую сторону. Так, более 90% цитируемых авторских публикаций наблюдаются в журналах «Краткие сообщения института археологии РАН», «Вестник Томского государственного университета. История», «Археологические вести», «Вестник Московского университета. Серия 8: История». Более 75% цитируемых публикаций членов редакционной коллегии зафиксированы в журналах «Археология, этнография и антропология Евразии», «Уральский исторический вестник», «Этнографическое обозрение». Причины таких неоднородностей и аномалий будут изучены в дальнейшем.

Корреляционный анализ показал относительно высокий коэффициент корреляции Пирсона между величинами «численность редколлегий» и «показатель журнала в рейтинге Science Index» R=73.8%". Однако рис. 26 показывает, что высокая корреляция обусловлена наличием одной выпадающей точки (АЭАЕ: 59 членов редколлегии, 3.814 – показатель журнала в рейтинге Science Index). В остальном множестве журналов зависимость между количеством членов редколлегии и его рейтингом если и наблюдается, то очень слабая.

**Таблица 23 – Число цитируемых публикаций**

| Журналы | Редакторы | Авторы | Всего |
|---|---|---|---|
| Вестник НГУ: серия история | 597 | 2091 | 2688 |
| Археология, этнография и антропология Евразии | 2943 | 174 | 3117 |
| Краткие сообщения института археологии РАН | 287 | 4493 | 4780 |
| Вестник археологии, антропологии и этнографии | 1043 | 486 | 1529 |
| Теория и практика археологических исследований | 891 | 670 | 1561 |
| Уральский исторический вестник | 1790 | 331 | 2121 |
| Известия Иркутского государственного университета: геоархеология, этнография, антропология | 218 | 418 | 636 |
| Вестник Томского государственного университета. История | 152 | 2299 | 2451 |
| Российская археология | 1623 | 952 | 2575 |
| Этнографическое обозрение | 2545 | 800 | 3345 |
| Археологические вести | 28 | 361 | 389 |
| Вестник Московского университета. Серия 8: История | 51 | 499 | 550 |
| Всего | 12158 | 13574 | 25732 |

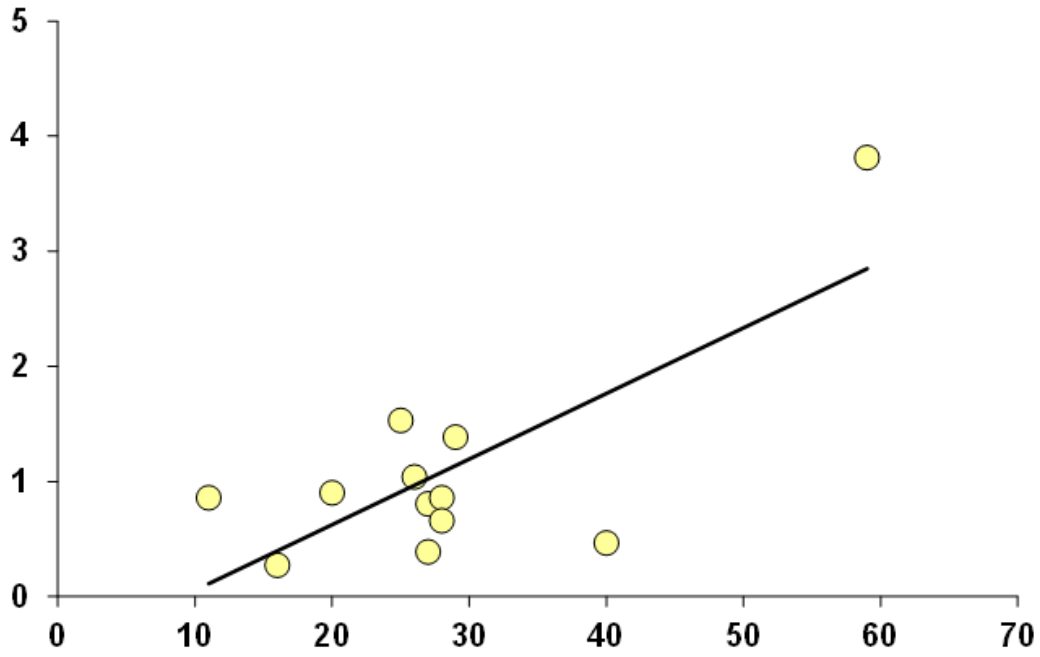


**Рисунок 26 – Корреляция численности редколлегий и рейтинга журналов.**

## ЗАКЛЮЧЕНИЕ

В 2019 г. сотрудниками лаборатории наукометрии были решены все поставленные по проекту промежуточные задачи.

По **первому блоку работ** было проанализировано качество библиографических данных на примере БД Scopus и возможности для его улучшения. Показано, что ошибки идентификации приводят к наличию дублирующих профилей у 77% организаций и 24% авторов. Выявлено 12 случаев возникновения дублей организаций и 4 случая возникновения дублей авторов. Наиболее часто причинами являются использование сокращенных названий, ошибки написания и различная транслитерация. Выполнена оценка влияния этих ошибок на количественные показатели результативности организаций, в среднем теряющих около 17% публикаций; отдельных авторов, для них потери составляют 11%.

Исследована возможность классификации научных текстов на основе метода компрессии их аннотаций. Идея метода состоит в том, что алгоритмы компрессии типа PPM (Prediction by partial matching) сжимают терминологически близкие тексты существенно лучше, чем далекие. Для изучения метода были сформированы 30 тематических категорий, для каждой из которых было подготовлено ядро (обучающая выборка) из 100 аннотаций научных статей и 20 аннотаций для тестирования. В качестве инструмента компрессии использовался архиватор WinRAR, в котором реализован алгоритм PPMd. Выявлено, что построение обучающего ядра на основе высокоцитируемых публикаций уменьшает количество ошибок классификации до 12% против 32% при случайной выборке. К ограничениям метода следует отнести то, что на качество классификации влияет изначальное количество категорий: чем меньше категорий и чем больше терминологическое различие между ними, тем выше ее качество.

Также была определена роль библиотек в информационном обеспечении научных исследований, включая задачи улучшения качества библиографических данных, а также необходимые для этого ресурсы и компетенции.

Была сформирована база данных расширенных библиографических описаний публикаций российских исследователей за 2000-2018 гг. на основе БД Scopus, сформированы справочники организаций и авторов, выполнена частичная дедупликация данных, на основании дополнительной информации (характеристики журналов, географическая и секторальная привязка аффилиаций) расширены и дополнены библиографические данные.

По **второму блоку работ** был выполнен анализ динамики позиционирования российской науки в мировом контексте, определены и описаны реперные точки изменений национальной научной политики, прослежено и проанализировано их влияние на изменение публикационной активности как российских исследователей в целом, так и в региональном и секторальном разрезе. Показано, что опережающий рост публикационной активности обеспечен сегментом высшего образования и локализован в регионах Сибири, Урала и Поволжья. Для проведения многомерного анализа был разработан ряд алгоритмов обработки данных. Анализ публикационного потока с детализацией по авторам и организациям показал, что основным драйвером роста публикационной активности является расширение количества активных исследователей и организаций, публикующих результаты работ в индексируемых в международных базах данных источниках, в связи с чем возрастает актуальность задачи изучения кадровой базы российской науки и перспектив ее развития. Было показано, что данные государственной статистики не вполне адекватно отражают текущее состояние и историю развития кадров науки, социологические исследования также не позволяют обрисовать полную картину, в связи с этим обоснована необходимость и возможность использования в таких исследованиях библиометрических данных.

На основе анализа данных и литературных источников выполнена классификация факторов, искажающих показатели национальной публикационной активности, сделаны оценки их влияния. Выполнен обзор литературных источников по изучению академической мобильности как одного из важных факторов результативности ученых. Подробно проанализировано влияние феномена международной синхронной мобильности на показатели результативности российских организаций и российской науки в целом. Обнаружены свидетельства предположительно недобросовестных практик ряда российских организаций, финансирующих в рамках программ по привлечению ведущих зарубежных исследователей «покупку» публикаций в форме вознаграждения за указание этими исследователями российской аффилиации в качестве дополнительной в своих статьях.

Выполнен анализ проведенной в 2017 г. ведомственной оценки результативности научных исследований организаций, подведомственных ФАНО России, определены недостатки методики количественной оценки, обоснована необходимость перехода на использование фракционного метода счета публикаций, разработаны методы исследования последствий перехода с целого на фракционный счет. Сделанный анализ показал, что наибольшие потери при переходе на фракционный счет понесут организации, находящиеся в последние годы под сильным «наукометрическим» давлением и, в частности в связи с

этим, активно сотрудничающие в научных коллаборациях с исследовательскими группами как из российских так и зарубежных организаций.

В **третьем блоке работ** было определено ядро из наиболее авторитетных журналов и разработана методика для его выделения, подробно изложенная в опубликованных по теме статьях. Сформированы базы данных членов редакционных коллегий отечественных журналов по информатике, библиотековедению и наукам о Земле, которые могут представлять интерес для главных редакторов при выборе рецензентов.

Анализ географического распределения членов редколлегий анализируемых дисциплин указывает на достаточно большое присутствие в редколлегиях зарубежных представителей, работа которых в отечественных журналах, тем не менее, зачастую носит формальный характер. Это подтверждается фактическим отсутствием как публикаций зарубежных членов редколлегий в российских журналах, так и отсутствием с их стороны цитирований статей в этих изданиях. В науках о Земле это связано с их региональным характером, поэтому анализируемые журналы имеют свой специфический пул авторов и читателей, в значительной мере ограниченный русскоязычной аудиторией ряда стран, граничащих с Россией. В свою очередь, это оказывает непосредственное влияние и на формирование состава редколлегий. В информатике и библиотековедении мы не видим объективных причин для такого ограничения и связываем невысокие на мировом уровне показатели журналов со слабой международной коллаборацией, которая нуждается в усилении.

Центры экспертных групп, представленных членами редколлегий анализируемых журналов, находятся преимущественно в столице, Санкт-Петербурге и Новосибирске. Для информатики и библиотековедения это также областные центры Европейской части России, а для наук о Земле – крупные города Сибири. При этом основными организациями – местами работы специалистов – для информатики и библиотековедения являются университеты, а для наук о Земле – научно-исследовательские организации Министерства науки и высшего образования. Для некоторых журналов показано влияние редколлегии на рейтинговые показатели журнала, включая случаи публикации высококачественных статей, впоследствии привлекающих большое число цитирований, а также «скрытое» цитирование своего журнала в других источниках. Для таких журналов редколлегии фактически являются самой основой их существования, а также «поставщиком» основных библиометрических показателей, таких как публикуемость и цитируемость, что противоречит рекомендациям международных комитетов по публикационной этике. В то же время для ряда других изданий указанная зависимость отсутствовала.

По тематике проекта было опубликовано 10 статей в рецензируемых научных журналах, в том числе 4 – в журналах, индексируемых в Web of Science Core Collection и Scopus, 5 статей в сборниках материалов конференций, в том числе 4 – в сборниках, индексируемых в Scopus.

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ


1. Valderrama-Zurián J.C. и др. A systematic analysis of duplicate records in Scopus // Journal of Informetrics. Elsevier Ltd, 2015. Т. 9, № 3. С. 570–576.
2. Kawashima H., Kawashima H., Tomizawa H. Accuracy evaluation of scopus author ID based on the largest funding database in japan // Scientometrics. Springer Netherlands, 2015. Т. 103, № 3. С. 1061–1071.
3. Aksnes D.W. When different persons have an identical author name. How frequent are homonyms? // Journal of the American Society for Information Science and Technology. 2008. Т. 59, № 5. С. 838–841.
4. Franceschini F., Maisano D., Mastrogiacomo L. Empirical analysis and classification of database errors in Scopus and Web of Science // Journal of Informetrics. Elsevier Ltd, 2016. Т. 10, № 4. С. 933–953.
5. Franceschini F., Maisano D., Mastrogiacomo L. The museum of errors/horrors in Scopus // Journal of Informetrics. 2016. Т. 10, № 1. С. 174–182.
6. Demetrescu C., Ribichini A., Schaerf M. Accuracy of author names in bibliographic data sources: an Italian case study // Scientometrics. Springer Netherlands, 2018. Т. 117, № 3. С. 1777–1791.
7. Schulz J. Using Monte Carlo simulations to assess the impact of author name disambiguation quality on different bibliometric analyses // Scientometrics. 2016. Т. 107, № 3. С. 1283–1298.
8. Perianes-Rodriguez A., Ruiz-Castillo J. A comparison of the Web of Science and publication-level classification systems of science // Journal of Informetrics. Elsevier Ltd, 2017. Т. 11, № 1. С. 32–45.
9. Shu F. и др. Comparing journal and paper level classifications of science // Journal of Informetrics. Elsevier Ltd, 2019. Т. 13, № 1. С. 202–225.
10. Waltman L., van Eck N.J. A new methodology for constructing a publication-level classification system of science // Journal of the American Society for Information Science and Technology. 2012. Т. 63, № 12. С. 2378–2392.
11. Block M., Khvatova T. University transformation: explaining policy making and trends in higher education in Russia // Journal of Management Development. 2017. Т. 36, № 6. С. 761–779.
12. Moed H.F., Markusova V., Akoev M. Trends in Russian research output indexed in Scopus and Web of Science // Scientometrics. Springer Netherlands, 2018. Т. 116, № 2. С. 1153–1180.

13. Мкртчян Н., Флоринская Ю. Квалификационная миграция в России: баланс потерь и приобретений // Мониторинг экономической ситуации в России. Тенденции и вызовы социально-экономического развития. 2018. № 1 (32). С. 15–18.
14. Минобрнауки России провело мониторинг и опубликовало данные о средней заработной плате штатных работников профессорско-преподавательского состава федеральных вузов за октябрь 2012 г. [Электронный ресурс] // Новости образования. URL: http://news-edu.ru/minobrnauki-rossii-provelo-monitori/.
15. «Копейка» в карман вузовского преподавателя [Электронный ресурс] // Информационный портал о высшем и дополнительном образовании в России ACADEMICA.RU. URL: https://academica.ru/stati/stati-o-pervom-vysshem-obrazovanii-i-magistrature/764140-kopejka-v-karman-vuzovskogo-prepodavatelja.
16. Мимо кассы. Ректоры двух вузов уволены из-за зарплат преподавателей // Российская газета - Столичный выпуск №204. 2015.
17. Зарплаты российских ученых впервые достигли почти 100 тыс. руб. [Электронный ресурс] // РБК. URL: https://www.rbc.ru/economics/10/04/2018/5ac77e589a794729bdf0c004.
18. Городникова Н.В. и др. Индикаторы науки: 2018 : статистический сборник. Москва: Нац. исслед. ун-т «Высшая школа экономики»., 2018. 320 с.
19. Kosyakov D., Guskov A. Impact of national science policy on academic migration and research productivity in Russia // Procedia Computer Science. 2019. Т. 146. С. 60–71.
20. Гуськов А.Е., Косяков Д.В., Селиванова И.В. МЕТОДИКА ОЦЕНКИ РЕЗУЛЬТАТИВНОСТИ НАУЧНЫХ ОРГАНИЗАЦИЙ, «Вестник Российской академии наук» // Вестник Российской академии наук. 2018. № 5. С. 430–443.
21. Macháček V., Srholec M. Predatory publications in Scopus: Evidence on cross-country differences. Prague: Charles University in Prague, Institute of Economic Studies (IES), 2019. № 20/2019.
22. Russian site says it has brokered authorships for more than 10,000 researchers [Электронный ресурс] // Retraction Watch. 2019. URL: https://retractionwatch.com/2019/07/18/exclusive-russian-site-says-it-has-brokered-authorships-for-more-than-10000-researchers/.
23. Chawla D. Russian journals retract more than 800 papers after ‘bombshell’ investigation // Science. 2020.
24. Jöns H. Transnational mobility and the spaces of knowledge production: a comparison of global patterns, motivations and collaborations in different academic fields // Social Geography. 2007. Т. 2, № 2. С. 97–114.

25. De Filippo D., Casado E.S., Gómez I. Quantitative and qualitative approaches to the study of mobility and scientific performance: a case study of a Spanish university // Research Evaluation. 2009. Т. 18, № 3. С. 191–200.
26. Trippl M. Scientific Mobility and Knowledge Transfer at the Interregional and Intraregional Level // Regional Studies. 2013. Т. 47, № 10. С. 1653–1667.
27. Mahroum S. Scientists and global spaces // Technology in Society. 2000. Т. 22, № 4. С. 513–523.
28. Шматко Н.А., Волкова Г.Л. Мобильность и карьерные перспективы исследователей на рынке труда // Высшее образование в России. 2017. Т. 208, № 1. С. 35–46.
29. Franzoni C., Scellato G., Stephan P. Patterns of international mobility of researchers: evidence from the GlobSci survey. Brisbane, Australia, 2012. 32 с.
30. Yudkevich M., Sivak E. University Inbreeding: An Impact on Values, Strategies and Individual Productivity of Faculty Members // SSRN Electronic Journal. 2012.
31. Horta H. Deepening our understanding of academic inbreeding effects on research information exchange and scientific output: new insights for academic based research // Higher Education. 2013. Т. 65, № 4. С. 487–510.
32. Scellato G., Franzoni C., Stephan P. Mobile Scientists and International Networks. Cambridge, MA, 2012.
33. Mokhnacheva Y. V., Beskaravainaya E. V., Kharybina T.N. The representation and professional success of scientific diasporas of the Pushchino Scientific Center abroad // Scientific and Technical Information Processing. 2016. Т. 43, № 4. С. 275–279.
34. Kharybina T.N., Beskaravainaya E.V., Mokhnacheva Y.V. Analysis of the model of an interaction with the Russianspeaking scientific diaspora abroad by the example of Institute of Biochemistry and Physiology of Microorganisms of RAS // Information and Innovations. 2018. Т. 13, № 4. С. 89–98.
35. Adams J. The fourth age of research // Nature. 2013. Т. 497, № 7451. С. 557–560.
36. Adams J.D. и др. Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981-1999 // Research Policy. 2005. Т. 34, № 3. С. 259–285.
37. Shashnov S., Kotsemir M. Research landscape of the BRICS countries: current trends in research output, thematic structures of publications, and the relative influence of partners // Scientometrics. 2018.
38. Guskov A.E., Kosyakov D. V., Selivanova I. V. Boosting research productivity in top Russian universities: the circumstances of breakthrough // Scientometrics. 2018. Т. 117, № 2. С. 1053–1080.

## ПРИЛОЖЕНИЕ.
## СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ

### Статьи, проиндексированные в Web of Science или Scopus

1. Selivanova I.V. The Impact of Errors in the Scopus Database on the Research Assessment / Selivanova I.V., Kosyakov D.V., Guskov A.E. // Scientific and Technical Information Processing. Volume 46, Issue 3, 1 July 2019. Pp 204-212. http://doi.org/10.3103/S0147688219030109
2. Гуреев, В. Н. Феномен научной мобильности в информетрических исследованиях / В. Н. Гуреев, Н. А. Мазов, А. Е. Гуськов // Научные и технические библиотеки. - 2019. - № 10. - С. 40-55. - http://doi.org/10.33186/1027-3689-2019-10-40-55
3. Selivanova I.V. Classification of Scientific Texts Based on the Compression of Annotations to Publications / Selivanova I.V., Kosyakov D.V., Guskov A.E. // Automatic Documentation and Mathematical Linguistics. – 2019. – Vol. 53, No. 6. – pp. 329–342 – http://doi.org/10.3103/S0005105519060062
4. Guskov, Andrey. Tuning national performance-based science policy: Introducing fractional count / A. Guskov, D. Kosyakov // Proceedings of the 17th International Conference on Scientometrics & Informetrics (ISSI 2019) with a Special STI Indicators Conference Track (Rome, Italy, 02-05 September 2019 г.). - Rome, 2019. - Vol. 2. - P. 2512-2513
5. Kosyakov D. Synchronous scientific mobility and international collaboration: case of Russia / Kosyakov Denis, Guskov Andrey // Proceedings of the 17th International Conference on Scientometrics & Informetrics (ISSI-2019) (2-5 September 2019, Rome, Italy), Volume 1. – 2019. – С. 1319-1328
6. Muhamedyev, R. Assessment of the dynamics of publication activity in the field of natural language processing and deep learning / Ravil Muhamedyev, Yan Kuchin, Kirill Yakunin, Denis Kosyakov, Sanzhar Murzakhmetov, Adilkhan Symagulov // Digital Transformation and Global Society, 4th International Conference, DTGS 2019, St. Petersburg, Russia, June 19–21, 2019, Revised Selected Papers, Communications in Computer and Information Science. – Т. 1038. –2019. – С.744-753. – http://doi.org/10.1007/978-3-030-37858-5_63
7. Kosyakov, D. V. Impact of national science policy on academic migration and research productivity in Russia / D. V. Kosyakov, A. E. Guskov // Procedia Computer Science. - 2019. - Vol. 146. - P. 60-71. - http://doi.org/10.1016/j.procs.2019.01.080

8. Kosyakov, D. V. Research assessment and evaluation in Russian fundamental science / D. V. Kosyakov, A. E. Guskov // Procedia Computer Science. - 2019. - Vol. 146. - P. 11-19. - http://doi.org/10.1016/j.procs.2019.01.072

**Статьи в рецензируемых журналах из списка ВАК**

1. Селиванова, И. В. Влияние ошибок в базе данных SCOPUS на оценку результативности научных исследований / И. В. Селиванова, Д. В. Косяков, А. Е. Гуськов // Научно-техническая информация. Серия 1. Организация и методика информационной работы. - 2019. - № 9. - С. 25-32. - http://doi.org/10.36535/0548-0019-2019-09-3
2. Селиванова, И. В. Классификация научных текстов на основе компрессии аннотаций публикаций/ И. В. Селиванова, Д. В. Косяков, А. Е. Гуськов // Научно-техническая информация. Серия 2. Информационные процессы и системы. - 2019. - № 12. - С. 25-38.

**Статьи в рецензируемых журналах, проиндексированные в РИНЦ**

1. Гуськов, А. Е. Матрица задач, ресурсов и компетенций для научных библиотек / А. Е. Гуськов, Д. В. Косяков, О. В. Макеева // Библиосфера. - 2019. - № 3. - С. 35-46. - http://doi.org/10.20913/1815-3186-2019-3-35-46
2. Мазов, Н. А. Состояние российского библиотековедения и информатики через исследование отраслевых журналов / Н. А. Мазов, В. Н. Гуреев // Библиосфера. - 2019. - № 3. - С. 56-70. - http://doi.org/10.20913/1815-3186-2019-3-56-70
3. Гуськов, А. Е. Проблемы мониторинга научных кадров / А. Е. Гуськов, Д. В. Косяков // Труды ГПНТБ СО РАН. - 2019. - № 1. - С. 55-61. - http://doi.org/10.20913/2618-7515-2019-1-55-61

**Публикации в сборниках докладов и тезисов конференций**

1. Костюк Д.М., Широков Н.К. Технология связывания авторских профилей РИНЦ с открытыми данными ЕГИСУ НИОКТР // Тезисы XX Всероссийской конференции молодых ученых по математическому моделированию и информационным технологиям. г. Новосибирск, Россия, 28 октября – 1 ноября 2019 г. — Новосибирск: ИВТ СО РАН, 2019. С.66.
2. Костюк Д.М., Широков Н.К. Технология дедупликации авторских профилей в библиографических базах данных // В книге: Информационные технологии материалы 57-й Международной научной студенческой конференции. 2019. С. 163.