

Глубокоуважаемые члены комиссии, сотрудники института!

В связи с тем, что тема цитирования (показатель цитирования ПРНД), несмотря на уже принятое в сентябре 2024 г. решение Учёного совета, продолжает вызывать интерес, ещё раз напоминаем мотивы решения, принятого на УС.

1. Изменение балла за показатель «Цитирование» системы ПРНД

После того, как в мае 2024 г. был закрыт доступ к базу данных (БД) Scopus комиссией было принято решение использовать открытую БД OpenAlex (ОА).

Это решение привело к заметному росту доли цитирования (32%) в балансе суммарного балла ПРНД. Результаты распределения баллов по категориям ПРНД по итогам 2023 г. приведены на Рис.1 (а).

Для того, чтобы вклад показателя цитирования был соразмерен с другими (не основными показателями), было предложено уменьшить значение балла за цитирование в два раза, что и было утверждено на УС 04 сентября 2024 г. С учетом этого изменения диаграмма распределения баллов по категориям ПРНД (Рис. 1 б) стала приближаться к распределению прошлых лет (подробности см. на сайте комиссии [здесь...](#)). Возросла доля текущих публикаций (с 48% до 59%), являющихся основным мерилем продуктивности института. Это в текущей ситуации весьма актуально, имея в виду проблемы с выполнением этого показателя госзадания у некоторых подразделений и сотрудников.

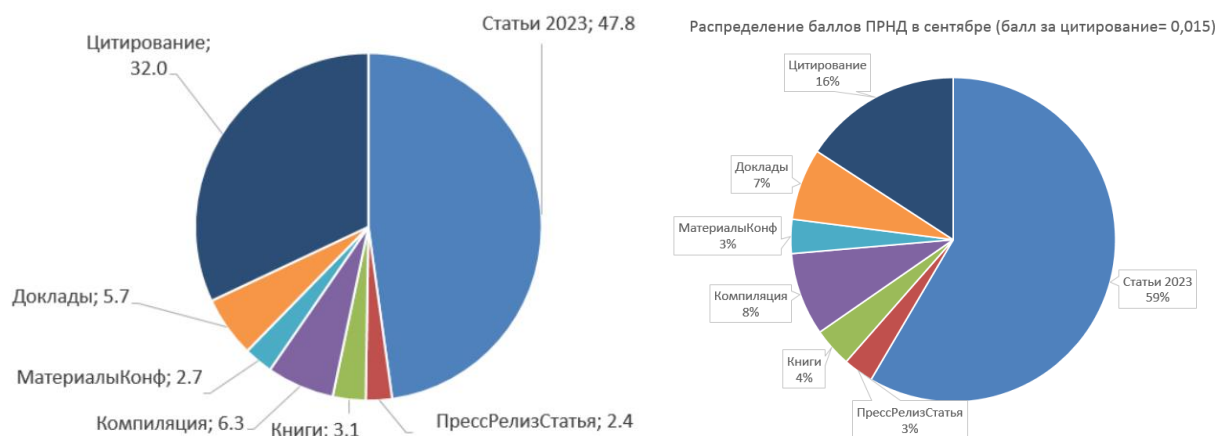


Рис. 1(а) (2023 год)

Рис. 1(б) (с 09-сен-2024 г.)

Рисунок 1. Вклад показателей в общий балл ПРНД

Для того, чтобы закончить обсуждение использования наукометрических БД для расчёта цитирования, ниже приведено их сравнение.

2. Использование альтернативных БД

Ранее комиссией рассматривалось несколько систем учёта цитирования. Приведём здесь описание проблем, с которыми мы столкнулись при их тестовом использовании.

Международные решения.

1) Web of Science (WoS). В настоящее время отсутствует свободный доступ к системе с возможностью составить запрос, который мог бы быть верифицирован секретариатом Института при проверке внесенных данных в Таблицу «Цитирование». Имеется лишь платный API ([Web of Science API Expanded](#)).

2) Scopus. В настоящее время отсутствует свободный доступ к системе с предоставлением данных о цитировании по годам. Информация о количестве цитирований недоступна в API в бесплатном пакете BASIC, только в платном пакете FULL.

3) Google Scholar (Google Академия). Предоставляет данные о цитировании публикаций только зарегистрированных пользователей. Позволяет добавить статьи, не принадлежащие автору, вручную. Учитывает отозванные статьи (см. рис. 1). API отсутствует, хотя существует «рынок платных (и дорогих) “сторонних” API с непонятным юридическим статусом» [1].

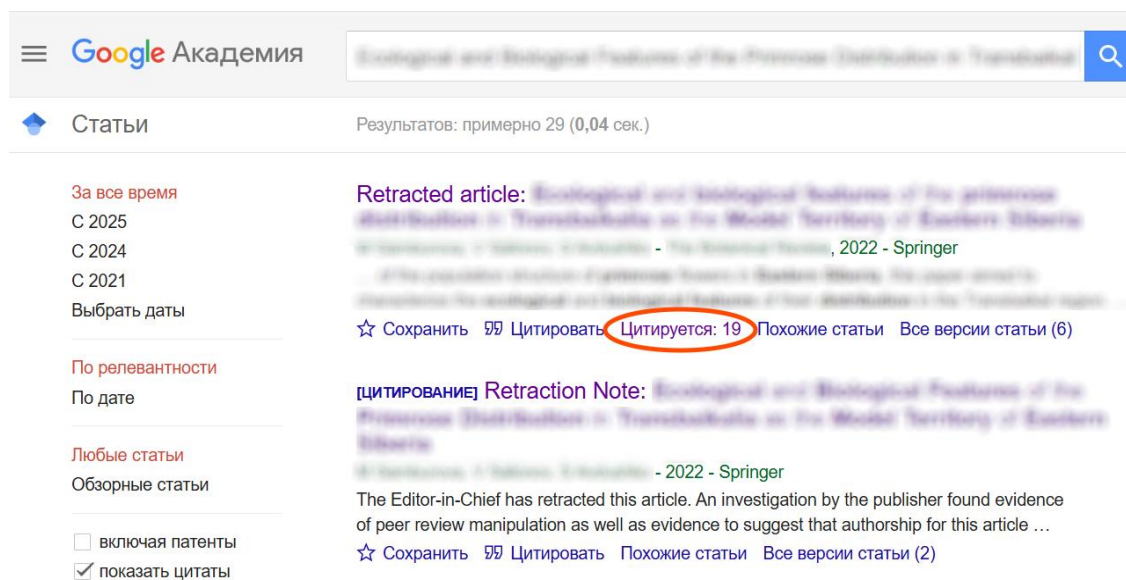


Рисунок 2. Учёт цитирований отозванной статьи в Google Академия

4) The Lens. Тяжелый интерфейс, сложность составления запросов, невозможность бесплатной выгрузки данных по окончании пробного (14 дней) периода. При этом, информация по цитированию в БД The Lens, по данным [2], обладает высоким консенсусом с ОА.

Российские решения.

5) Индекс цитирования РЦНИ. В настоящее время, предоставляет информацию только по запросу DOI-DOI, т.е. требует предварительного сбора информации по всем публикациям автора.

6) РИНЦ – платный API (от 36 000 рублей, в зависимости от числа сотрудников), вопросы к индексируемым изданиям.

Переход на альтернативные базы данных обсуждался Комиссией по ПРНД, и этот вопрос выносился на заседания Учёного совета (18 апреля и 04 сентября 2024 г.).

К сожалению, ввиду недостатка времени (а это трудоемкая процедура), не удалось в полной мере детально проанализировать возможность учёта годового цитирования в БД Crossref, OpenCitations, COCI, I4OC, Semantic Scholar, Dimensions, Wizdom.ai, Scilit, FATCAT, BASE, [VOSviewer](#), а также в ряде БД с использованием ПО Publish or Perish. Сотрудники РЦНИ [3] сопоставили ряд этих БД, хотя, вероятно, для некоторых сотрудников нашего Института, наполнение этих БД в пределах погрешности может отличаться от БД OpenAlex.

Важно, что все перечисленные выше базы данных не имеют **бесплатного публичного интерфейса программирования приложения** (application programming interface, API) (БД The Lens – не предлагает **бесплатного** API). Поэтому работа с показателями цитирования в таких БД проходит в режиме «черного ящика», а вопросы верификации и анализа данных о цитировании остаются открытыми (о критериях открытости – см. [4]). Наукометрический центр Высшей Школы Экономики даёт такое определение открытости метаданных (в т.ч. данных по цитированию): «полноценные открытые метаданные - это не просто данные, доступные всем бесплатно, это данные, свободное использование которых - в т.ч. коммерческое - разрешено официально <...>. Поэтому, например, [Google Scholar](#) к Open Access-системам не относится» [3]. Единственной БД с открытой лицензией ([Creative Commons 0](#)) является OpenAlex. А так как детальный и верифицируемый анализ возможен только по данным БД с открытой лицензией, то по этому критерию особого выбора кроме, OpenAlex нет.

3. Сбор расширенной информации о цитировании

Возможности открытого API OpenAlex позволяют получить не только общую количественную информацию о значении персонального цитирования, но и получить возможность проанализировать этот показатель постатейно.

В 2025 году, на R 4.4.2 [5], в RStudio 2024.09.0 [6], с использованием библиотеки openalexR 1.4.0 [7], был подготовлен скрипт¹ сбора данных о цитировании публикации за год. Скрипт с подробными комментариями на русском языке выложен в Интернете и с его текстом можно ознакомиться по ссылке [здесь...](#)

Скрипт позволяет **автоматически за 15-20 минут** (в зависимости от скорости интернет-соединения и производительности ПК) собрать данные по цитированию каждой публикации **всех сотрудников Института**, информация о которых имеется в БД ОА, за каждый год. Собранный файл с данными доступен на сайте комиссии по ПРНД, [здесь...](#)

Структура данных файла (описание по [8]):

Поле	Описание
name	ФИО сотрудника
id	порядковый номер публикации сотрудника, учитываемый в расчёте
year	год учёта цитирования
cited_by_count.x	количество цитирований публикации в данный год, ед.
link	ссылка на карточку публикации в ОА
title	заголовок публикации
display_name	заголовок публикации
author	(данные отсутствуют в файле, в выгрузке - отдельная таблица с указанием авторов, их идентификаторов, аффилиаций, стран, является ли автор корреспондирующим и т.д. – всего 12 полей для каждого из авторов публикации)
ab	аннотация (при наличии в ОА)
publication_date	дата публикации
so	издание
so_id	ссылка на карточку издания в ОА
host_organization	издательство
issn_1	ISSN журнала

¹ автор скрипта-В.С.Б, является членом комиссии по ПРНД

url	ссылка на публикацию
pdf_url	ссылка на pdf-файл публикации (при наличии в ОА)
license	лицензия
version	версия публикации
first_page	номер первой страницы публикации
last_page	номер последней страницы публикации
volume	том
issue	выпуск
is_oa	наличие публикации в открытом доступе на сайте журнала
is_oa_anywhere	наличие публикации в открытом доступе
oa_status	тип открытого доступа: (green – легальные постпринты в репозиториях, gold – финальные версии на сайтах журналов и в репозиториях, diamond – Gold без APC, bronze – без внятных лицензий, подробнее об открытом доступе см. напр. https://clck.ru/3Fms5u)
oa_url	ссылка на открытый доступ
any_repository_has_fulltext	доступность полного текста в репозитории
language	язык (en - английский, ru - русский и т.д.)
grants	(данные отсутствуют в файле, в выгрузке ссылки на источники финансирования, при наличии в ОА)
cited_by_count.y	общее число цитирований данной статьи
counts_by_year	(данные отсутствуют в файле, в выгрузке - отдельная таблица с указанием цитирования статьи, аналогичная представленной в выгрузке)
publication_year	год публикации
cited_by_api_url	ссылка на данные по цитированию статьи в формате JSON
ids	(данные отсутствуют в файле, в выгрузке - отдельная таблица с указанием идентификаторов статьи в других БД)
doi	doi публикации
type	тип публикации (article - статья, book - монография и т.д.)
referenced_works	(данные отсутствуют в файле, в выгрузке - отдельная таблица с указанием списка литературы статьи)
related_works	(данные отсутствуют в файле, в выгрузке - отдельная таблица с указанием близких по теме работ)
is_paratext	является ли статья паратекстом
is_retracted	отозвана ли статья
concepts	(данные отсутствуют в файле, в выгрузке - отдельная таблица с указанием направлений статьи)
topics	(данные отсутствуют в файле, в выгрузке - отдельная таблица с указанием тем статьи)
info	технический столбец
oaaid	перечень идентификаторов авторов, использованный для анализа, из формы 10

Полученные данные (которые можно загрузить на ваш компьютер) позволяют провести детальный анализ цитирований за любой требуемый период как для отдельного сотрудника, так и для групп сотрудников (лаборатории, отдела, должности и т.д.).

Появилась возможность **автоматизированного** анализа как цитирования **каждой статьи**, так и **каждого цитирования** в отдельности, что невозможно сделать ни в одной другой базе данных!

Как у комиссии, так и у сотрудников, появляется **возможность верификации сведений**, которые, прежде, представляли собой только цифровое значение цитирования без детализации, направляемое ответственным за ввод данных. Это позволяет отслеживать тенденции с цитированием работ сотрудников/подразделений и, вероятно, поможет принимать организационные меры.

4. Предварительный анализ

В качестве примера приведём несколько результатов **предварительного** анализа информации для всех сотрудников Института за период 2020-2024 гг. (общий пул ~44 тыс. цитирований).

Цитирований по журналам (более 500 ед. за период)

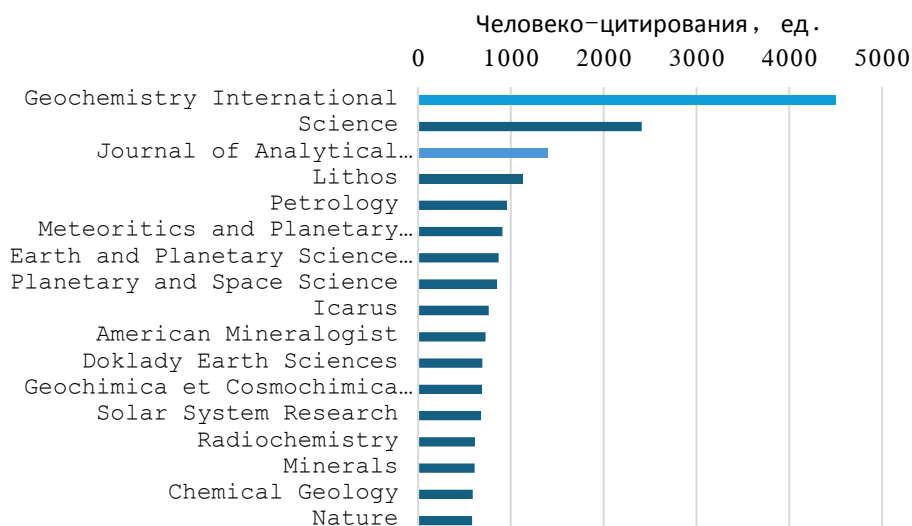


Рисунок 3. Количество человеко-цитирований за 2020-2024 гг. по журналам, в которых опубликована статья.

Цитирований по году публикации

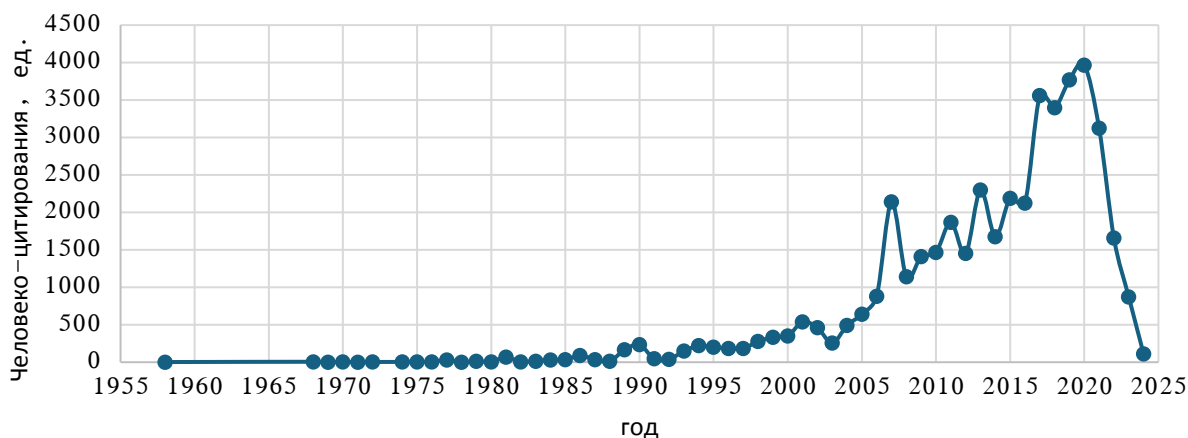


Рисунок 4. Количество человеко-цитирований за 2020-2024 гг. по году публикации

5. Заключение

Сервисами БД ОА активно пользуется РЦНИ.

Комиссия по ПРНД готова к конструктивному диалогу со всеми, кто заинтересован в выработке мер, стимулирующих повышение эффективности научной деятельности института.

Список литературы

1. Стерлигов И.А. Онлайн-руководство по наукометрии. — URL: <https://sciguide.hse.ru/sources/opendata/> (дата обращения: 07.12.2023).
2. Лутай А.В. Сравнение качества метаданных в БД CrossRef, Lens, OpenAlex, Scopus, Semantic Scholar, Web of Science Core Collection / А.В. Лутай, Е.Э. Любушко; Российский фонд фундаментальных исследований. – Текст : электронный // РФФИ : [сайт]. – URL: https://podpiska.rfbr.ru/storage/reports2021/2022_meta_quality.html. – Дата публикации: 21.02.2022.
3. Лутай А.В. Открытые сервисы для поиска научных публикаций/ А.В. Лутай, О.В. Черченко; Российский центр научной информации. – Текст : электронный // РЦНИ : [сайт]. – URL: https://podpiska.rfbr.ru/materials/2022_open_search_solutions/. – Дата публикации: 28.12.2022.
4. Peroni, S., & Shotton, D. (2018). Open Citation: Definition. <https://doi.org/10.6084/M9.FIGSHARE.6683855.V1>
5. R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
6. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
7. Aria, M., Le T., Cuccurullo, C., Belfiore, A. & Choe, J. (2024), openalexR: An R-Tool for Collecting Bibliometric Data from OpenAlex, The R Journal, 15(4), 167-180, DOI: <https://doi.org/10.32614/RJ-2023-089>.
8. Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. ArXiv. <https://arxiv.org/abs/2205.01833>