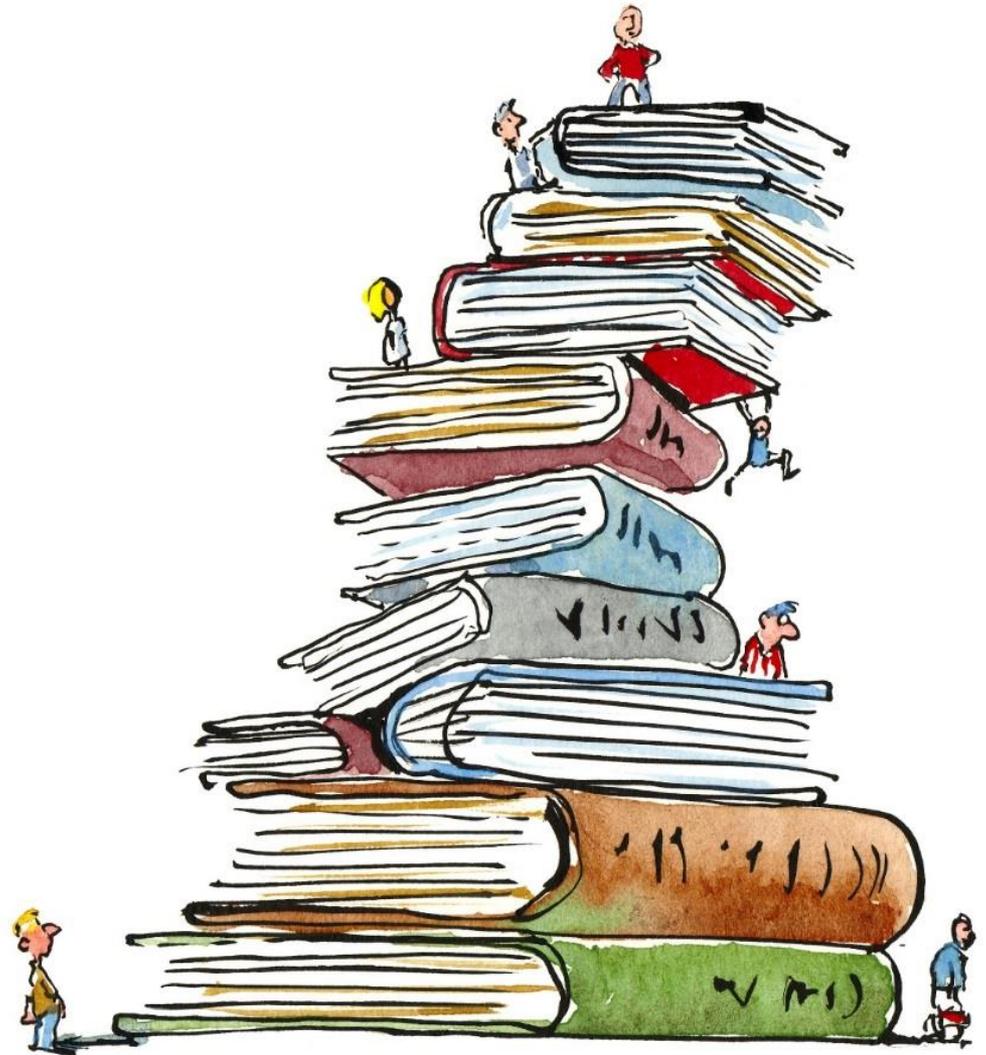
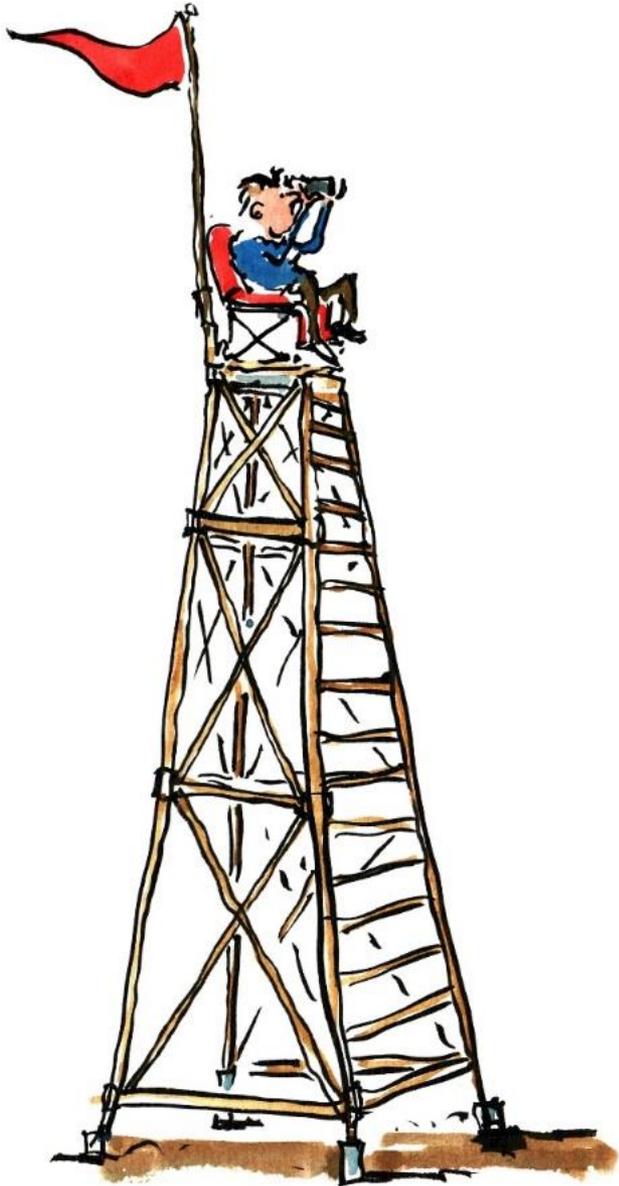


# Открытые альтернативы индексам научного цитирования

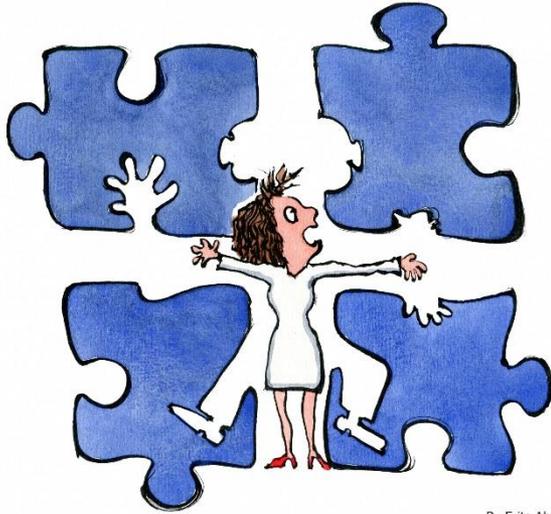
Лутай А.В.

Российский фонд фундаментальных исследований

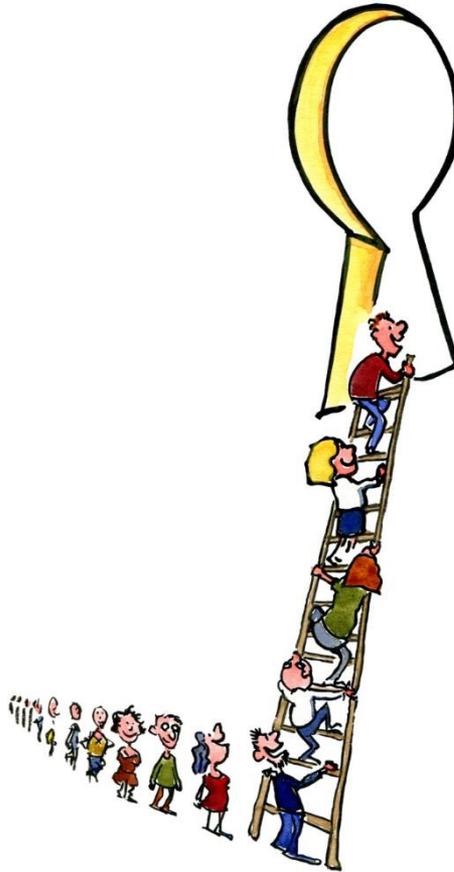
# Инструменты для выбора правильной σκοπός (цель, мишень)



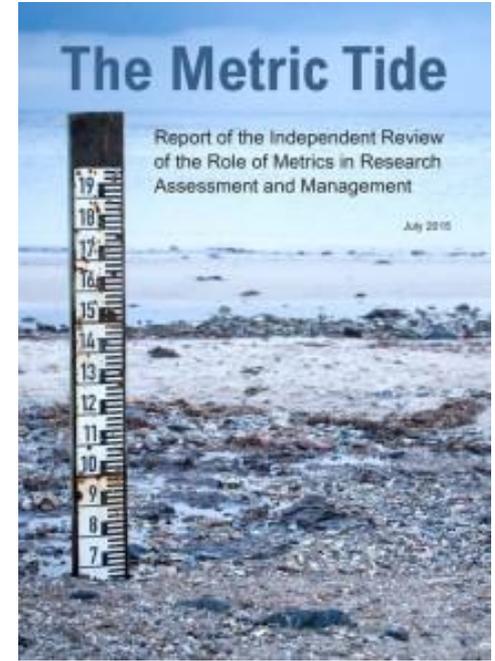
# Инструменты, принесшие в науку тиранию метрик



By Frits Ahlefeldt



HikingArtist

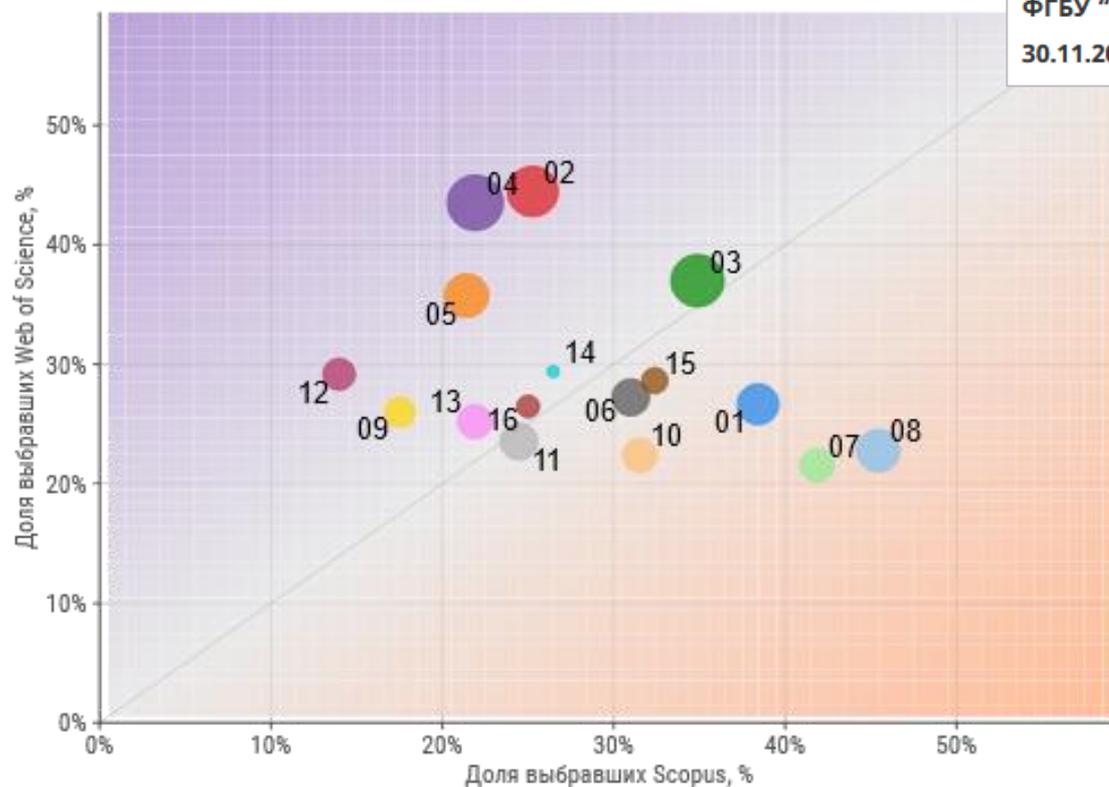


# Насколько востребованы индексы самими исследователями?

## Область научных исследований и выбор индекса

Более существенное влияние на выбор индекса научного цитирования оказывает специализация респондентов. Были взяты из профилей исследователей в КИАС РФФИ. Специалисты в областях {02. физика и астрономия}, {03. химия}, {04. биология и медицина}, {05. геология и география}, {06. инженерные науки}, {07. информатика и информационные технологии}, {08. инженерные науки}, {09. филология и искусствоведение} отдали предпочтение Web of Science, тогда как специалисты в областях {01. информатика и информационные технологии}, {08. инженерные науки} чаще выбирали Scopus. Для того, чтобы увидеть результаты каждого научного направления, наведите курсор на любую точку.

## Область научных исследований и выбор индекса



## Выбор российских исследователей между Scopus и Web of Science: результаты опроса РФФИ

Лутай А.В., Любушко Е.Э.

ФГБУ "Российский фонд фундаментальных исследований"

30.11.2020

- 02
- 08
- 14
- 03
- 09
- 15
- 04
- 10
- 16
- 05
- 11
- 06
- 12

Web of Science      Scopus



Количество ответов

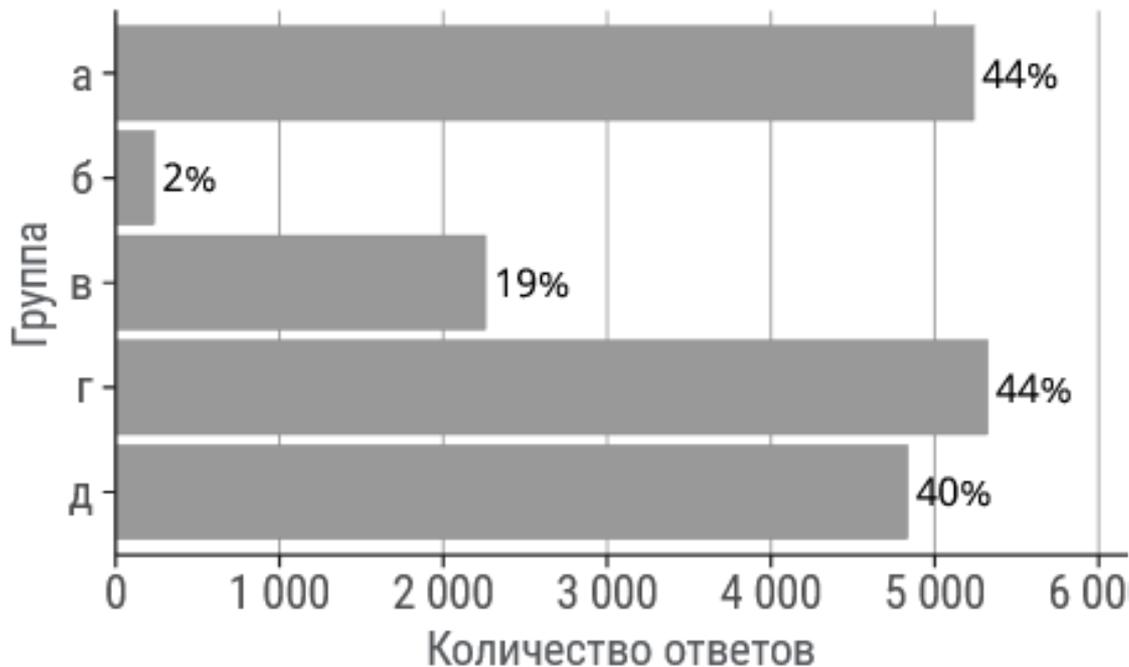
- 200
- 1200
- 400
- 1400
- 600
- 1600
- 800
- 1800
- 1000

© РФФИ, иллюстрация, 2020.  
Дата создания: 30.11.2020.  
Данные: КИАС РФФИ

# Российские исследователи поддерживают профили в разных БД

## Поддержание собственного профиля автора и выбор индекса

- (а) ORCID
- (б) Microsoft Academic AuthorID
- (в) GoogleScholar AuthorID
- (г) ResearcherID
- (д) Scopus AuthorID

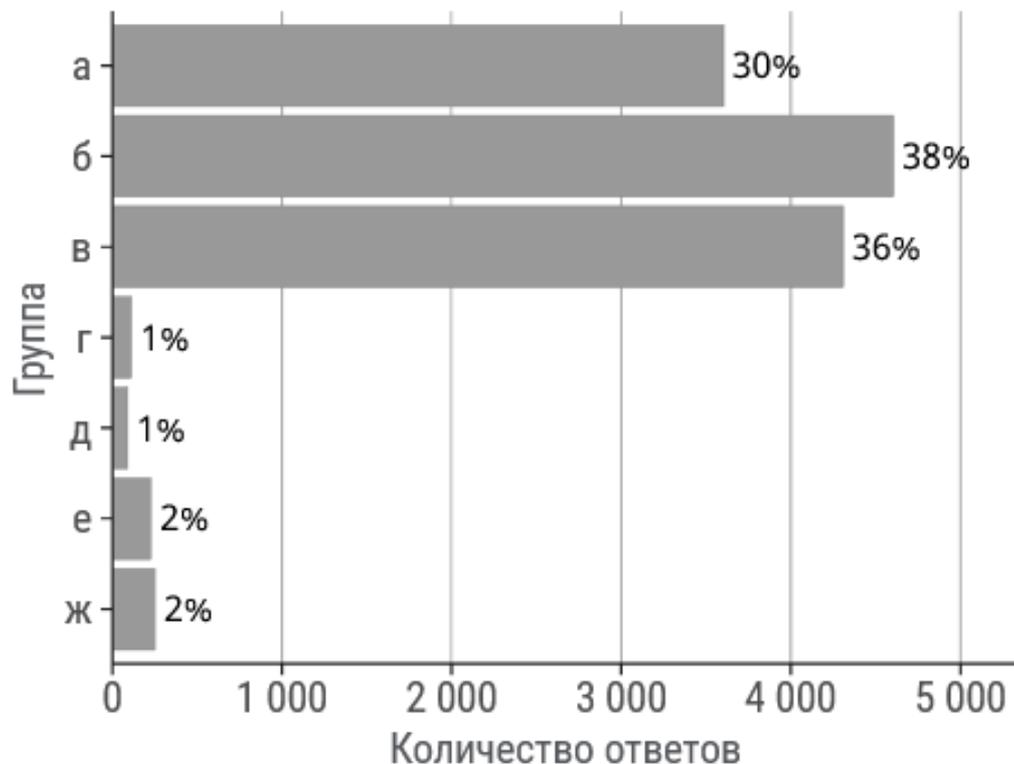


Доля респондентов, которые «*стараяются поддерживать*» собственные профили в ORCID была сопоставима с долей тех, кто поддерживал профили в Scopus AuthorID и ResearcherID

# Только ли индексы цитирования используют для поиска?

## Использование реферативных баз данных для поиска и выбор индекса

(а) Google Scholar, (б) Scopus, (в) Web of Science, (г) Dimensions, (д) Lens, (е) Microsoft Academic, (ж) Semantic Scholar



Доля респондентов, которые не реже раза в неделю использовали для поиска статей Google Scholar составила 30%,

Тогда как доля тех, кто регулярно использовал открытые поисковые системы, не превышала 3%.

# Насколько необходимыми являются индексы по мнению исследователей?

Насколько критичным окажется отсутствие доступа к Web of Science и Scopus для поиска новой информации в Вашей исследовательской работе?

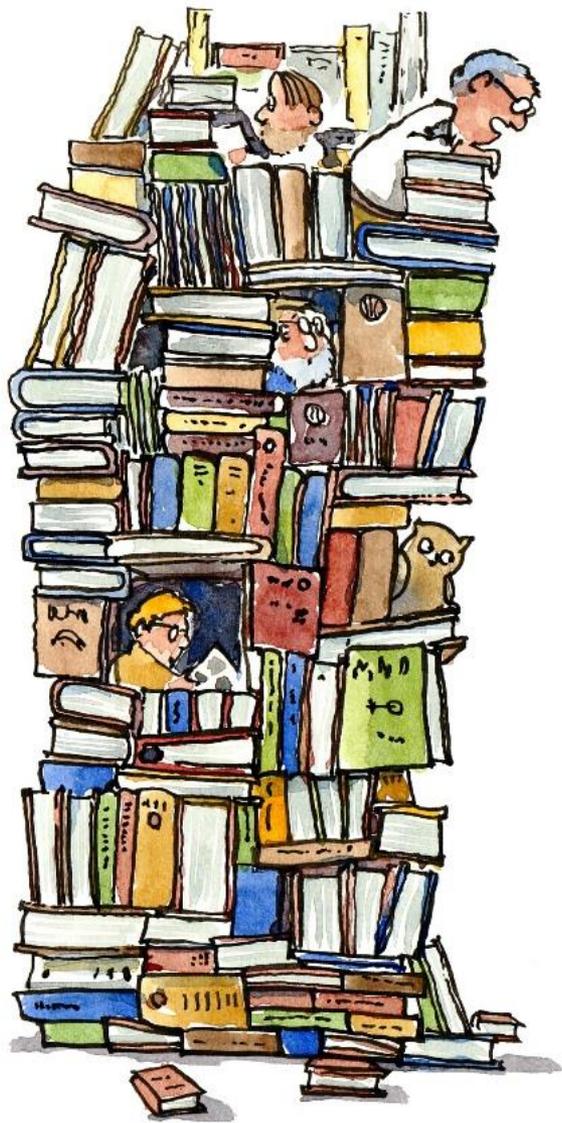
Наличие ученой степени

Доктор наук  Кандидат наук  Аспирант  Нет

Код	Название предметной области в рубрикаторе РФФИ	Ученая степень	Ответов	Не станет проблемой	Вызовет затруднения	Остановит работу
	All	All	⊗	A	All	A
03	ХИМИЯ И НАУКИ О МАТЕРИАЛАХ	Доктор наук	445	22%	39%	38%
03	ХИМИЯ И НАУКИ О МАТЕРИАЛАХ	Кандидат наук	890	22%	40%	37%
04	БИОЛОГИЯ	Доктор наук	539	31%	32%	33%
04	БИОЛОГИЯ	Кандидат наук	1079	32%	37%	28%
08	ФУНДАМЕНТАЛЬНЫЕ ОСНОВЫ ИНЖЕНЕРНЫХ НАУК	Доктор наук	296	36%	36%	25%
05	НАУКИ О ЗЕМЛЕ	Доктор наук	277	41%	32%	24%
08	ФУНДАМЕНТАЛЬНЫЕ ОСНОВЫ ИНЖЕНЕРНЫХ НАУК	Кандидат наук	476	27%	47%	24%
02	ФИЗИКА И АСТРОНОМИЯ	Доктор наук	407	34%	41%	23%
05	НАУКИ О ЗЕМЛЕ	Кандидат наук	540	31%	43%	23%
01	МАТЕМАТИКА И МЕХАНИКА	Кандидат наук	406	31%	43%	22%
02	ФИЗИКА И АСТРОНОМИЯ	Кандидат наук	763	31%	44%	22%

Более трети респондентов с ученой степенью в областях «Биология» и «Химия» заявили о том, что их научная работа невозможна без доступа хотя бы к одному из индексов научного цитирования.

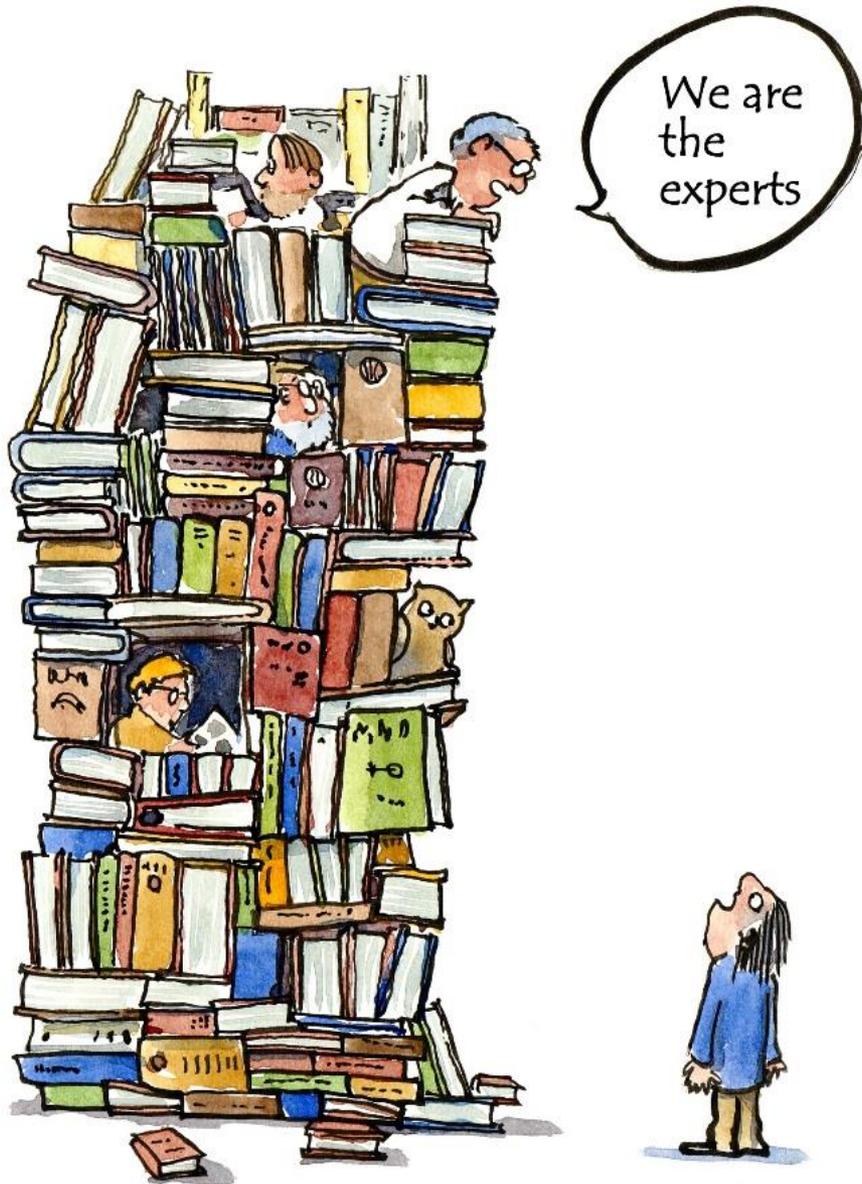
# И наступит день, и придёт человек, и задаст главный вопрос



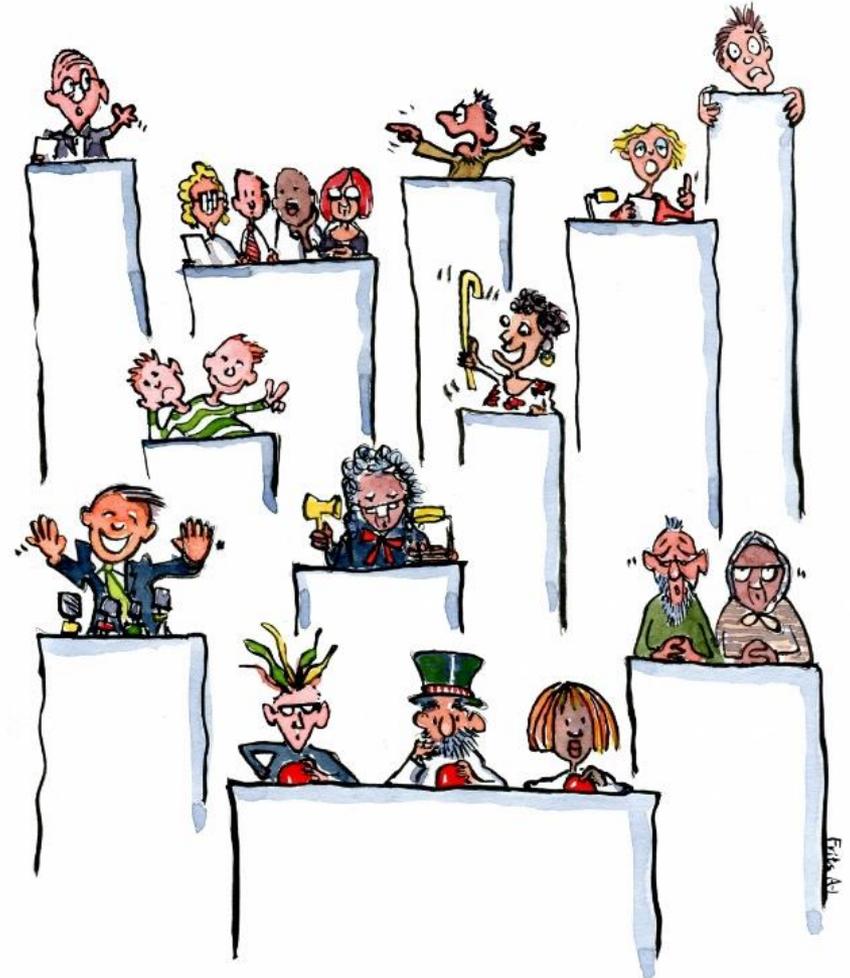
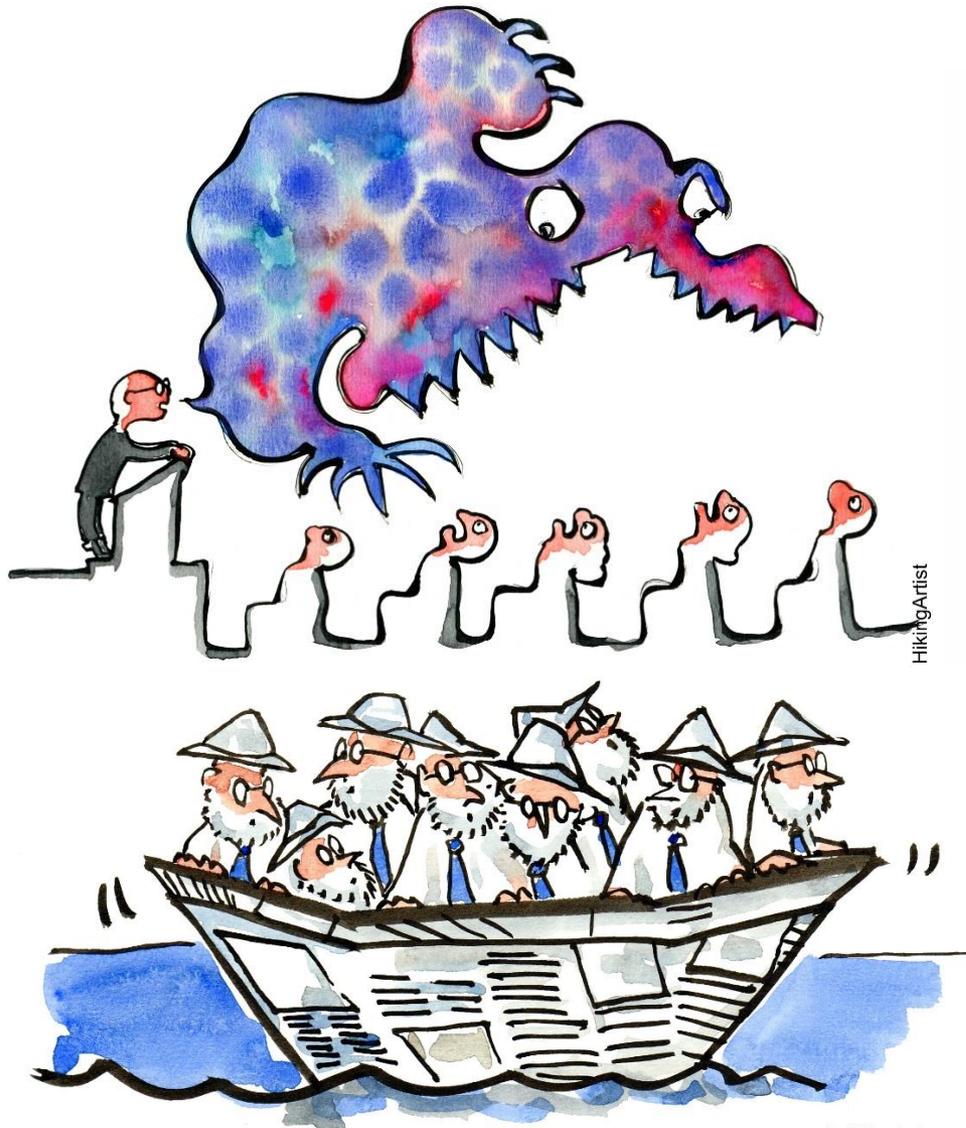
Доступ к ...  
не работает.  
Как мне теперь  
делать науку!?



# А мы ему такие: «Ок, сейчас всё объясним»



# Дебаты о важности индексов научного цитирования для науки



# А что же экспертное мнение?



# Что такое хорошо и что такое плохо для индексов научного цитирования?

Чтобы правильно задать  
вопрос, надо знать  
большую часть ответа



# Варианты использования индексов научного цитирования

## ЗАДАЧИ

1. Поиск научных работ, исследователей, проектов, организаций
2. Анализ цитирований (метрики и сеть)
3. Подсчет публикаций в определенных научных изданиях
4. Наполнение собственных ИС метаданными и метриками

## ПРЕИМУЩЕСТВА ИНДЕКСОВ

1. Качество индексации полей метаданных + алгоритмы поиска
2. Качество распознавания списков литературы
3. Обновляемые списки изданий, метрики и готовые инструменты
4. Современные схемы, удобство интеграции, качество метаданных

## Сравнение качества метаданных в БД CrossRef, Lens, OpenAlex, Scopus, Semantic Scholar, Web of Science Core Collection

Лутай А.В., Любушко Е.Э.

ФГБУ "Российский фонд фундаментальных исследований"

21.02.2022

**Массив** – 60 тыс. статей (РФ, 2016-21 г.)

**Цель анализа** – оценка качества полей метаданных и сведений о цитировании (полнота, однородность, консенсус между источниками):

авторы | ORCID | аффилиации | наличие аннотации | финансирование  
списки литературы | количество цитирований

# Что за ресурсы включены в анализ и почему только они?

НАЗВАНИЕ РЕСУРСА	ОХВАТ ИЗДАНИЙ	ДОСТУП (CSV/API)
Google Scholar, Dimensions, Baidu Scholar, J-Global, SciLit, Naver Academic, CNKI Scholar, etc.	+	-
MEDLINE, DBLP, AMiner, РИНЦ, etc.	-	+
CrossRef	+	+
The Lens	+	+
Semantic Scholar	+	+
Open Alex	+	+

## ОСНОВАНИЯ ДЛЯ ОПТИМИЗМА

## РЕАЛЬНОСТЬ

Мощности для сбора данных (максимальный охват)

Безразличие к уровню и качеству издания

Наличие наработок в области ИИ для анализа и распознавания метаданных

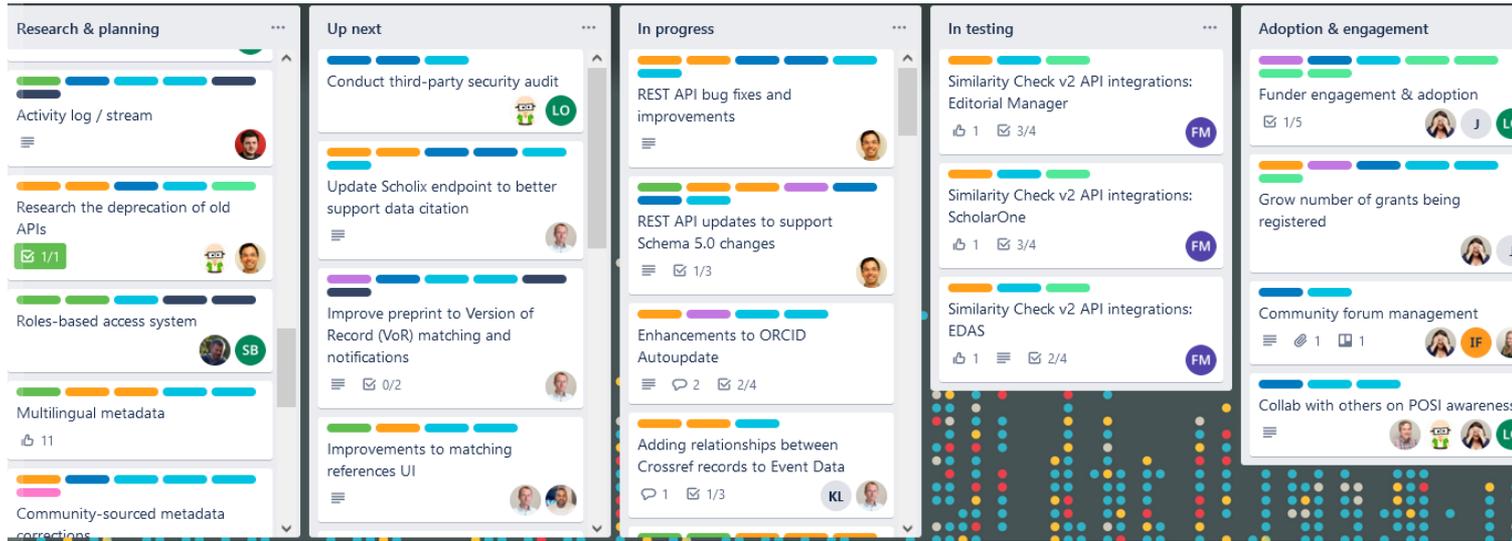
Отсутствие опыта и людей для проверки

Привязка профилей к учетным записям глобальных сервисов

Отсутствие видимого интереса к персонализации сервисов

Отсутствие стратегии развития, связывающей направления роста с потребностями пользователей сервиса!

DOI ORCID  
 API ROR  
 FundRef  
 I4OC, COCI  
 I4OA



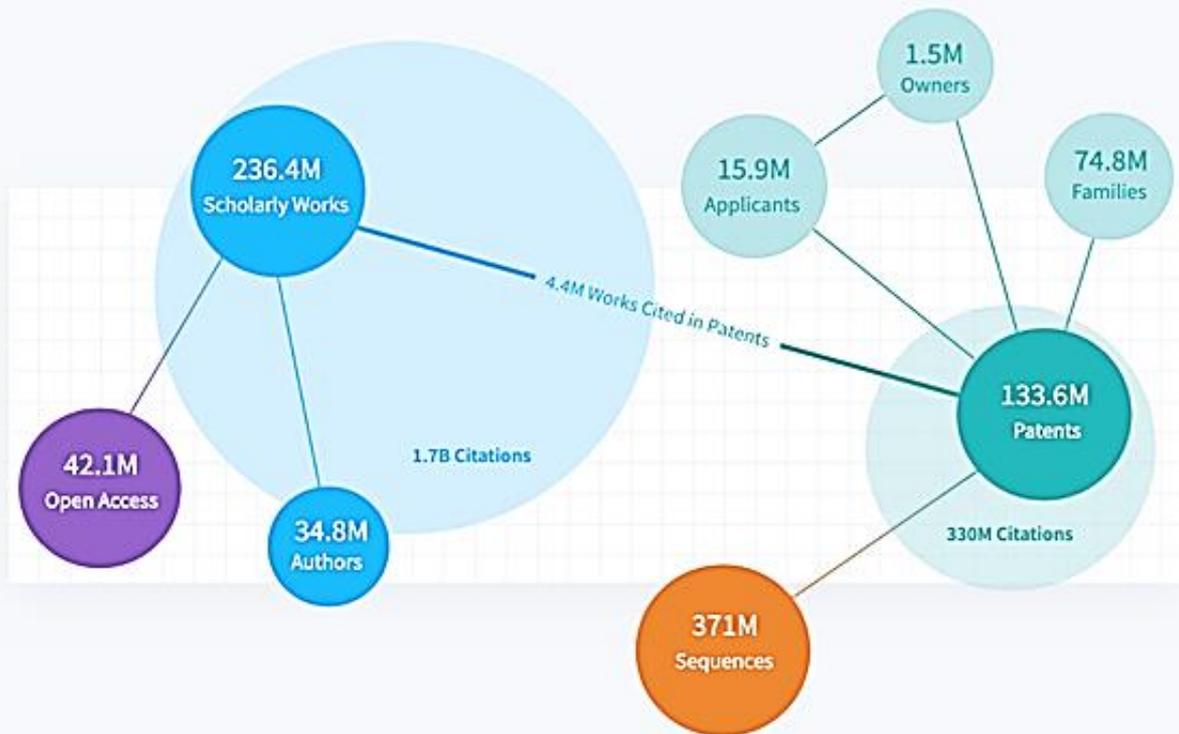
## ОСНОВАНИЯ ДЛЯ ОПТИМИЗМА

- Crossref DOI = стандарт
- POSI +
- Максимум возможностей для загрузки корректных и полных метаданных

## ОСНОВАНИЯ ДЛЯ ПЕССИМИЗМА

- Качество в руках издателя
- Необходимость взносов (исключает ряд издателей)

Статьи  
 Патенты  
 Профили  
 Отчеты  
 API  
 PatCite



### ОСНОВАНИЯ ДЛЯ ОПТИМИЗМА

- Дорожная карта
- Акцент на доступность
- Регулярные обновления

### ОСНОВАНИЯ ДЛЯ ПЕССИМИЗМА

- Унификация записей
- Ограниченность источников
- Развитие платных сервисов

Статьи

TLDR\*

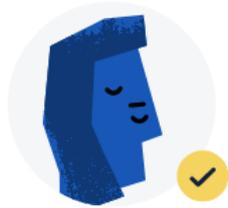
Impact\*

Топики\*

API-S2AG

Reader\*

Профили



**Author Homepages**  
 We use AI to showcase authors' impact on science and highlight their most influential papers.



**Recommendations and Alerts**  
 Easily stay up-to-date with customized recommendations based on saved papers.



**TLDRs**  
 Automatically generated single-sentence paper summaries to help prioritize which papers to read in-depth.



**Semantic Reader**  
 An augmented reading application makes the reading experience more accessible and richly contextual by providing citation information directly within the context of a paper.

## ОСНОВАНИЯ ДЛЯ ОПТИМИЗМА

- Проект AI2 -> технологии
- Бесплатный API и датасеты (S2AG, S2ORC)

## ОСНОВАНИЯ ДЛЯ ПЕССИМИЗМА

- Отсутствие публичной стратегии развития
- Фокус на открытых данных

## Data

The OpenAlex dataset describes scholarly *entities* and how those entities are connected to each other. There are five types of entities:

- 📄 **Works** are papers, books, datasets, etc; they *cite* other works
- 👤 **Authors** are people who *create* works
- 🌐 **Venues** are journals and repositories that *host* works
- 🏠 **Institutions** are universities and other orgs that are *affiliated with* works (via authors)
- 💡 **Concepts** tag Works with a topic

### ОСНОВАНИЯ ДЛЯ ОПТИМИЗМА

- Опытная команда\*
- Лицензия CCO, POSI +
- API/дампы
- Максимум данных
- Веб-интерфейс (2022-09)

### ОСНОВАНИЯ ДЛЯ ПЕССИМИЗМА

- Качество индексации (?)
- Качество источников

## Цитируемость статей

Базы данных: SC = Scopus, WS = Web of Science Core Collection, CR = CrossRef, LN = Lens, S2 = Semantic Scholar, OA = OpenAlex  
 Массив статей: 58634 публикаций 2016-2021 гг. с участием исследователей из РФ.

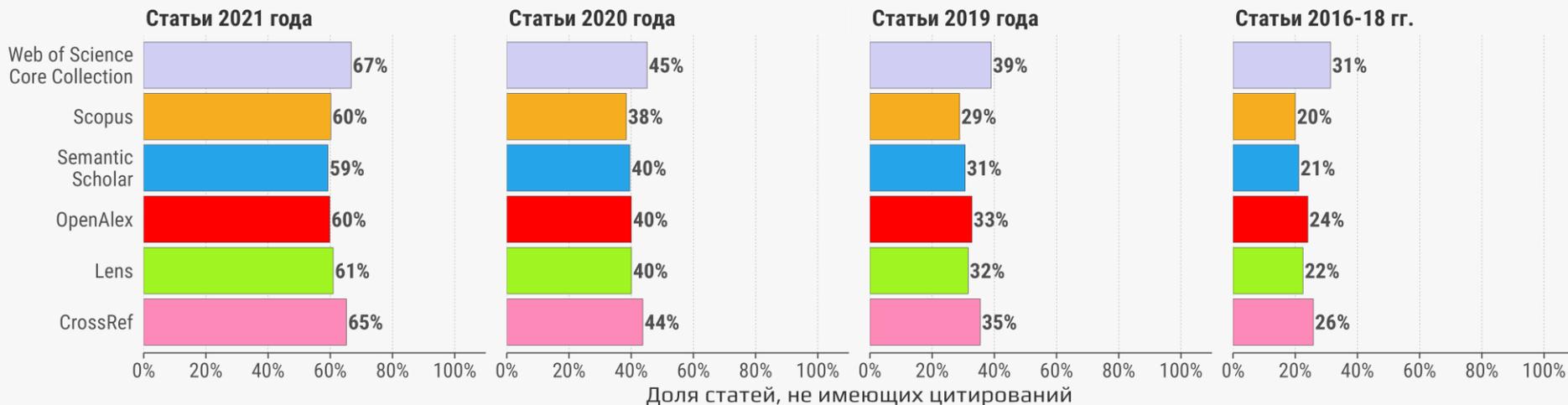


© РФФИ, иллюстрация, 2022  
Дата создания: январь-февраль 2022

Во всех подгруппах суммарное количество цитирований в WoSCC меньше, чем в других БД. CrossRef опережает WoSCC, но уступает Scopus. Количество цитирований в Lens, S2 и OpenAlex приблизительно одинаково и больше наблюдаемого для Scopus.

## Статьи без цитирований

Базы данных: SC = Scopus, WS = Web of Science Core Collection, CR = CrossRef, LN = Lens, S2 = Semantic Scholar, OA = OpenAlex  
Массив статей: 58634 публикаций 2016-2021 гг. с участием исследователей из РФ.

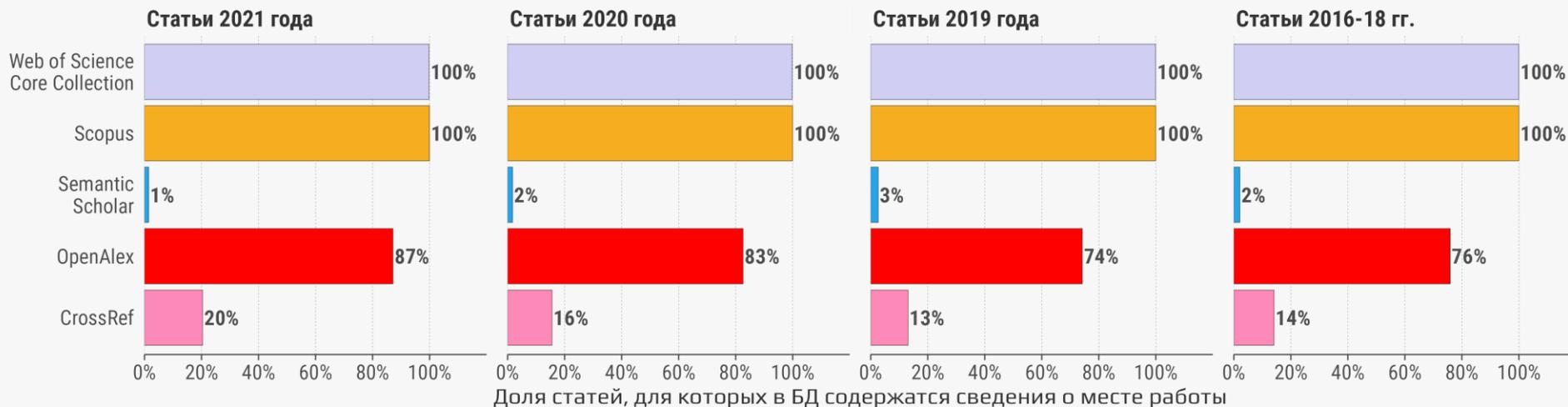


© РФФИ, иллюстрация, 2022  
Дата создания: январь-февраль 2022

В подгруппе статей 2021 года доля публикаций без цитирований для Scopus, S2, Open Alex, Lens была сопоставимой (около 60%), с увеличением возраста статей разница в пользу Scopus увеличивалась до нескольких процентов. Разница между Scopus и WoSCC от 7 до 10%.

## Наличие информации о месте работы авторов

Базы данных: SC = Scopus, WS = Web of Science Core Collection, CR = CrossRef, S2 = Semantic Scholar, OA = OpenAlex  
Массив статей: 57570 публикаций 2016-2021 гг. с участием исследователей из РФ. Исключены статьи с 100 и более авторов.



© РФФИ, иллюстрация, 2022  
Дата создания: январь-февраль 2022

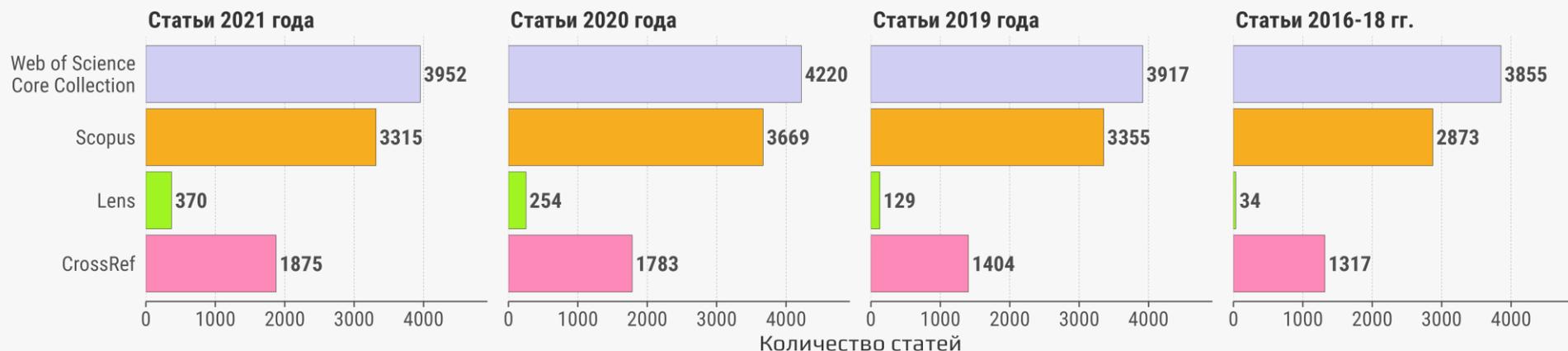
WoSCC и Scopus намного всех опережают! В CrossRef новые статьи заметно больше имеют аффилиаций, но лишь 20% для статей 2021 года. OpenAlex показывает ROR, к которому пока остается немало вопросов.

## Количество статей, содержащих сведения о гранте по маске поиска

Базы данных: SC = Scopus, WS = Web of Science Core Collection, CR = CrossRef, LN = Lens

Массив статей: 58634 публикации 2016-2021 гг. с участием исследователей из РФ.

Маска поиска: (№1) RFBR|RSF|Russian Foundation for Basic Research|Russian Science Foundation, (№2) [0-9]{2}-[0-9]{2}-[0-9]{5}

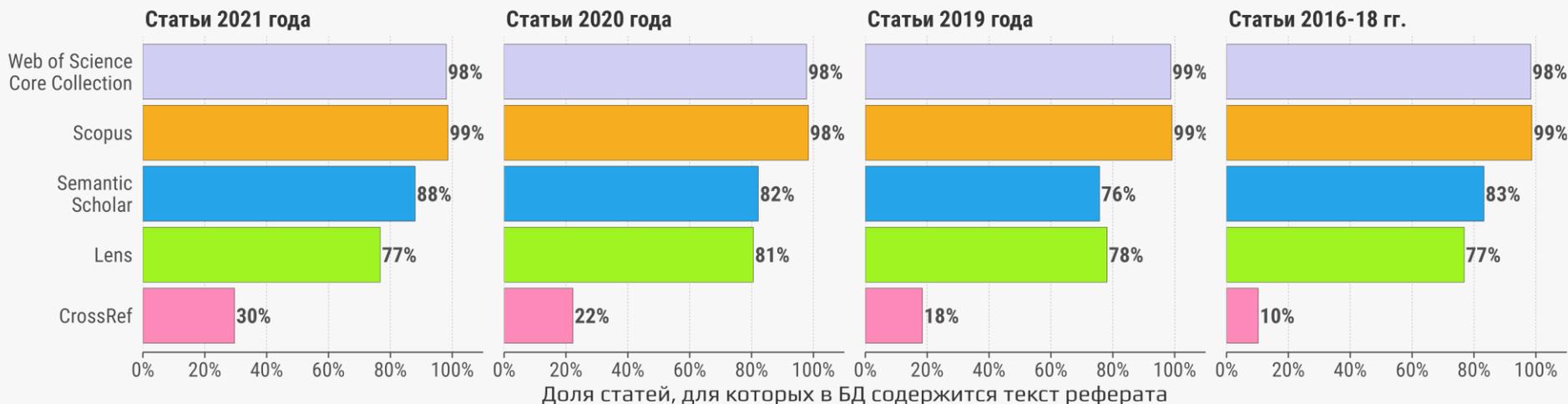


© РФФИ, иллюстрация, 2022  
Дата создания: январь-февраль 2022

В 2019-2021 гг. в WoSCC обнаруживается приблизительно на 15% больше удовлетворяющих условиям поиска статей, чем в Scopus, тогда как на выборке 2016-2018 гг. отличие в 2 раза больше. В CrossRef по условиям поиска обнаруживается в 2 раза меньшее количество статей, чем в WoSCC. Lens пока рано рассматривать в качестве источника сведений о финансировании.

## Наличие реферата в метаданных

Базы данных: SC = Scopus, WS = Web of Science Core Collection, CR = CrossRef, LN = Lens, S2 = Semantic Scholar  
Массив статей: 58634 публикаций 2016-2021 гг. с участием исследователей из РФ.

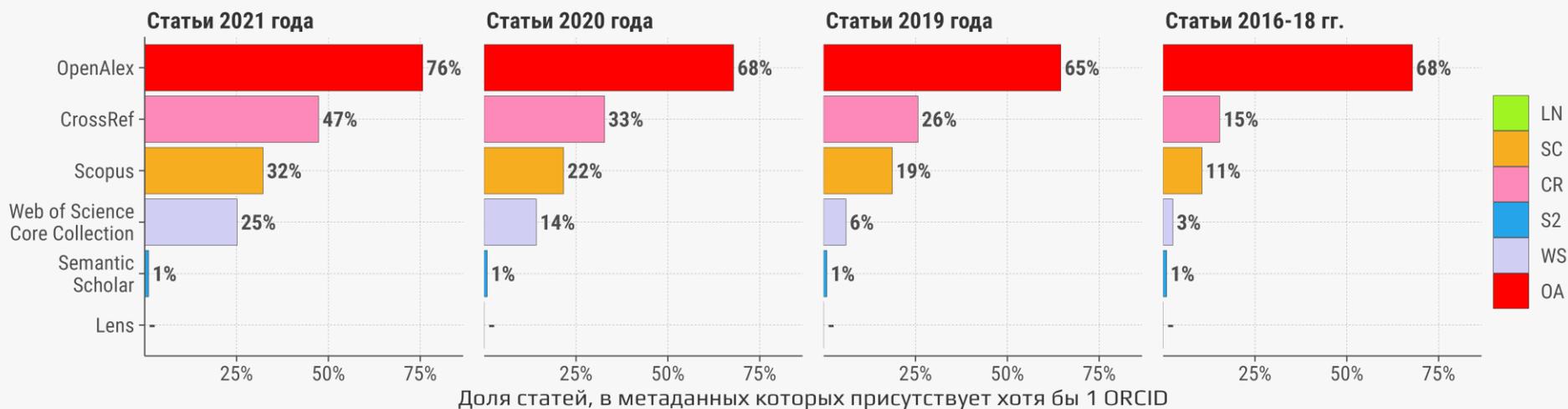


© РФФИ, иллюстрация, 2022  
Дата создания: январь-февраль 2022

WoSCC ~Scopus > S2 > Lens >> CrossRef (в OpenAlex на момент исследования API не возвращал реферат). Если OpenAlex не наладит индексацию, то в S2 и Lens доля статей с рефератом СНИЗИТСЯ.

## Присутствие ORCID в доступных метаданных публикаций

Базы данных: SC = Scopus, WS = Web of Science Core Collection, CR = CrossRef, LN = Lens, S2 = Semantic Scholar, OA = OpenAlex  
 Массив статей: 57569 публикаций 2016-2021 гг. с участием исследователей из РФ. Исключены статьи с 500 и более авторов.



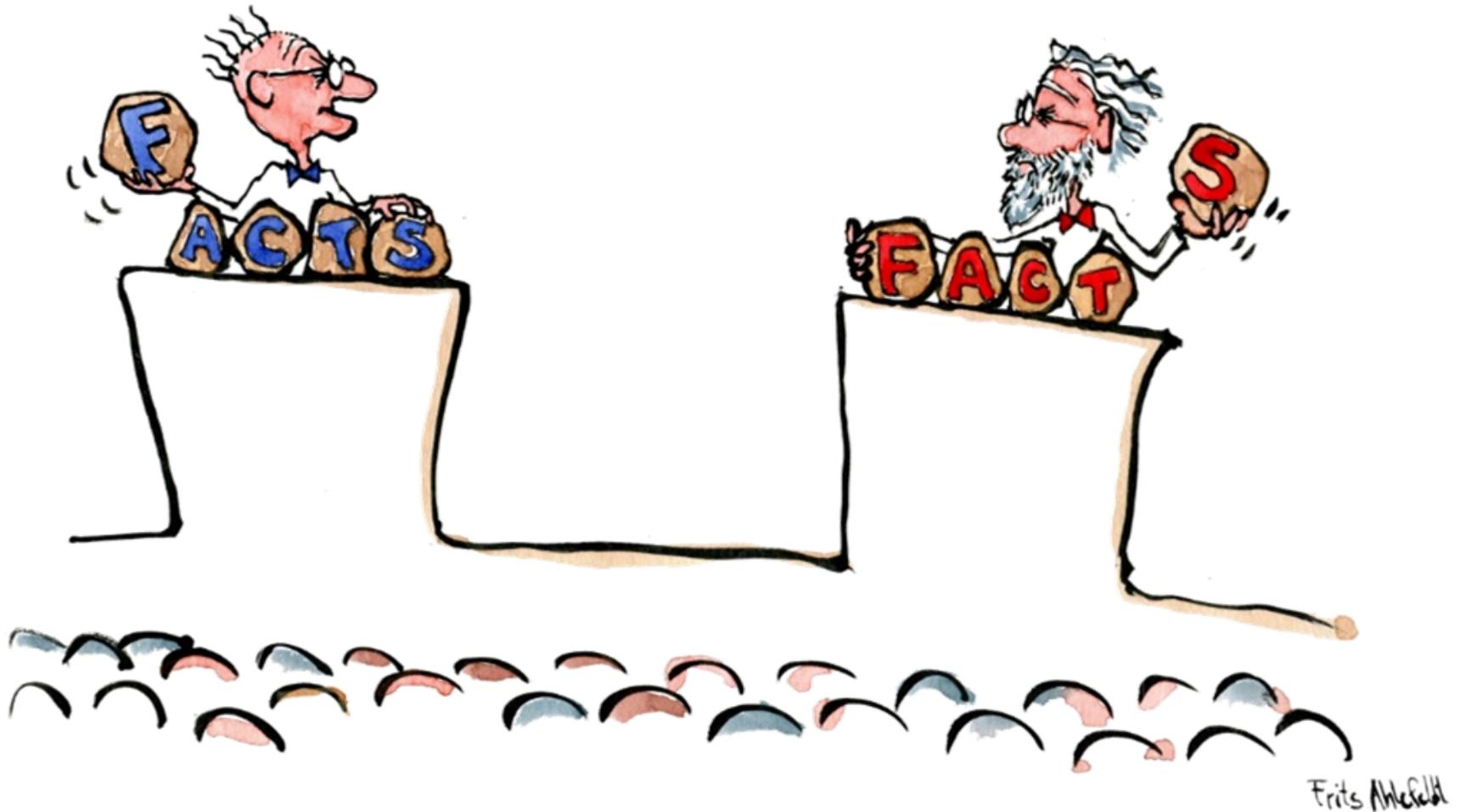
© РФФИ, иллюстрация, 2022  
Дата создания: январь-февраль 2022

Доля статей, в которых встречается хотя бы один ORCID, гораздо выше в статьях последних лет. В OpenAlex частота встречаемости ORCID в несколько раз выше, чем в Scopus и в WoSCC – три четверти статей 2021 г. из исследуемого массива имели хотя бы 1 ORCID в OpenAlex, тогда как в CrossRef лишь 47% (\*).

- Эволюция открытых БД Crossref, OpenAlex, Lens и Semantic Scholar не вызывает сомнения – качество их содержимого и уровень сервисов быстро повышается.
- На сегодняшний день открытые БД\* имеют убедительный охват изданий и предлагают широкий выбор инструментов (в т.ч. рекомендательных) для исследователей, желающих выполнить поиск по теме.
- Качество метаданных вполне достаточно для построения и анализа сетей цитирования и сотрудничества авторов.
- Некоторые поля метаданных (аффилиации, финансирования, реферат) пока недостаточно полны, чтобы заменить данные из проприетарных ИНЦ для решения задач учета и мониторинга.

- ORCID уверенно растет в популярности. OpenAlex и Lens дополняют свои данные из ежегодного открытого массива данных ORCID. Проверка корректности данных в ORCID остается под вопросом.
- Проект ROR делает попытку занять место ORCID среди держателей реестров организаций (уже есть в Crossref, Lens, OpenAlex). Проект сейчас занят разработкой новой схемы данных, которая будет учитывать многообразие отношений у организаций. Возможно, в будущем будет хорошо, но пока больше вопросов.
- Связки ROR – ОГРН для российских научных организаций в проекте (Russian Index of Research Organization / [RIRO](#)).
- [O Wikidata и профилях организации на русском языке.](#)

# Теперь, можно и пообсуждать! ВОПРОСЫ?



# СПАСИБО ЗА ВНИМАНИЕ!



**Frits Ahlefeldt-Laurvig**

[hikingartist.com](http://hikingartist.com)

[fritsahlefeldt.com](http://fritsahlefeldt.com)

The illustrations protected by copyright have been purchased @ [fritsahlefeldt.com](http://fritsahlefeldt.com). Some illustrations are licensed under CC-BY-ND 3.0 ([see](#))