

ОРИГИНАЛЬНЫЕ СТАТЬИ / ORIGINAL ARTICLES

Открытый доступ сегодня: широкомасштабный анализ распространенности и влияния статей открытого доступа¹

Хизер Пивовар^{1*}, Джейсон Прим^{1*}, Винсент Ларивьер^{2,3}, Хуан Пабло Алперин^{4,5}, Лиза Маттиас⁶,
Брее Норландер^{7,8}, Эшли Фарли^{7,8}, Джейвин Вест⁷, Стефани Хауштайн^{3,9}

¹ *ИмпактСтори (Impactstory), г. Сэнфорд, Северная Каролина, Соединенные Штаты Америки*

² *Школа библиотечных и информационных наук Университета Монреалья, Павийон Лионель-Грульс, ул. Жан-Брийан, 3150, г. Монреаль, Квебек, НЗТ 1N8, Канада*

³ *Научно-техническая обсерватория Межвузовского научно-исследовательского центра науки и техники Университета Квебека в Монреале, Павийон Поль-Жерин-Лажуа (N), ул. Сен-Дени, 1205, г. Монреаль, Квебек, Н2Х 3R9, Канада*

⁴ *Канадский институт исследований в издательском деле Университета Саймона Фрезера, ул. Уэст Хастингс, 515, г. Ванкувер, Британская Колумбия, V6B 5K3, Канада*

⁵ *Проект «Общественные знания» (Public Knowledge Project), г. Ванкувер, Британская Колумбия, Канада*

⁶ *Лаборатория научной коммуникации Университета Саймона Фрезера, ул. Уэст Хастингс, 515, г. Ванкувер, Британская Колумбия, V6B 5K3, Канада*

⁷ *Информационная школа Вашингтонского университета, Мэри Гейтс Холл, апарт. 370, г. Сиэтл, Вашингтон, 98195-2840, Соединенные Штаты Америки*

⁸ *ФлоришОА (FlourishOA), г. Сиэтл, Вашингтон, Соединенные Штаты Америки*

⁹ *Школа информационных исследований Университета Оттавы, Демаре Холл, Восточная Лорье-авеню, 55, г. Оттава, Онтарио, K1N 6N5, Канада*

Аннотация

Хотя интерес к открытому доступу (ОД) к научной литературе выразился в большом количестве публикаций, пока остается неудовлетворенной потребность в широкомасштабном современном исследовании с воспроизводимыми результатами, которое оценило бы распространенность и характеристики публикаций открытого доступа. Мы откликаемся на эту потребность с помощью оаDOI, общедоступного онлайн-сервиса, определившего статус ОД для 67 миллионов статей. Для исследования мы используем три группы статей объемом 100 тысяч каждая: (1) статьи с DOI от CrossRef; (2) статьи последних лет, проиндексированные в Web of Science; и (3) статьи, открытые пользователями Unpaywall, общедоступного браузерного расширения, позволяющего пользователям находить статьи ОД с помощью оаDOI. Наша оценка процента научной литературы открытого доступа — не менее 28% (19 миллионов), причем этот процент растет в основном за счет Золотого и Гибридного ОД. Последний проанализированный период, 2015 г., показал наивысший процент ОД — 45%. Благодаря этому росту и с учетом того, что пользователи чаще всего обращаются к свежим статьям, мы обнаружили, что пользователи Unpaywall достаточно часто просматривают статьи ОД: 47% всех открытых статей. Интересно, что самый распространенный тип ОД — не Золотой, Зеленый или Гибридный, а дискуссионный тип, который мы называем Бронзовым. Это

¹ Перевод статьи: Piwowar H., Priem J., Larivière V., Alperin J.P., Matthias L., Norlander B., Farley A., West J., Haustein S. The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. PeerJ. 2018;6:e4375. <https://doi.org/10.7717/peerj.4375>

статьи в свободном (бесплатном) доступе на сайте издателя, не имеющие открытой лицензии. Мы также исследуем влияние открытого доступа на цитирование статей, подтверждая так называемое преимущество статей ОД в аспекте цитирования: с учетом времени публикации и тематики статей ОД они получают в среднем на 18% больше цитирований. Этот эффект обеспечивается в основном за счет Зеленого и Гибридного ОД. Мы призываем к дальнейшим исследованиям с использованием сервиса oaDOI с целью выработки обоснованной политики и практики открытого доступа.

Ключевые слова: открытый доступ, открытая наука, наукометрия, издательская деятельность, библиотеки, научная коммуникация, библиометрия, научная политика

Для цитирования: Пивовар Х., Прим Дж., Ларивьер В., Алперин Х. П., Маттиас Л., Норландер Б., Фарли Э., Вест Дж., Хауштайн С. Открытый доступ сегодня: широкомасштабный анализ распространенности и влияния статей открытого доступа. *Наука и научная информация*. 2019;2(4):228-247. <https://doi.org/10.24108/2658-3143-2019-2-4-228-247>

The State of OA: a Large-Scale Analysis of the Prevalence and Impact of Open Access Articles

Heather Piwowar^{1*}, Jason Priem^{1*}, Vincent Larivière^{2,3}, Juan Pablo Alperin^{4,5}, Lisa Matthias⁶, Bree Norlander^{7,8}, Ashley Farley^{7,8}, Jevin West⁷, Stefanie Haustein^{3,9}

¹ Impactstory, Sanford, North Carolina, USA

² École de bibliothéconomie et des sciences de l'information, Université de Montréal, Pavillon Lionel-Groulx, rue Jean-Brillant, 3150, Montréal, Quebec, H3T 1N8, Canada

³ Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, Pavillon Paul-Gérin-Lajoie (N), rue Saint-Denis, 1205, Montréal, Québec, H2X 3R9, Canada

⁴ Canadian Institute for Studies in Publishing, Simon Fraser University, West Hastings str., 515, Vancouver, British Columbia, V6B 5K3, Canada

⁵ Public Knowledge Project, Vancouver, British Columbia, Canada

⁶ Scholarly Communications Lab, Simon Fraser University, West Hastings str., 515, Vancouver, British Columbia, V6B 5K3, Canada

⁷ Information School, University of Washington, Mary Gates Hall, Suite 370, Seattle, Washington, 98195-2840, USA

⁸ FlourishOA, Seattle, Washington, USA

⁹ School of Information Studies, University of Ottawa, Desmarais Hall, Laurier Avenue East, 55, Ottawa, Ontario, K1N 6N5, Canada

Abstract

Despite growing interest in Open Access (OA) to scholarly literature, there is an unmet need for large-scale, up-to-date, and reproducible studies assessing the prevalence and characteristics of OA. We address this need using oaDOI, an open online service that determines OA status for 67 million articles. We use three samples, each of 100,000 articles, to investigate OA in three populations: (1) all journal articles assigned a Crossref DOI, (2) recent journal articles indexed in Web of Science, and (3) articles viewed by users of Unpaywall, an open-source browser extension that lets users find OA articles using oaDOI. We estimate that at least 28% of the scholarly literature is OA (19M in total) and that this proportion is growing, driven particularly by growth in Gold and Hybrid. The most recent year analyzed (2015) also has the highest percentage of OA (45%). Because of this growth, and the fact that readers disproportionately access newer articles, we find that Unpaywall users encounter OA quite frequently: 47% of articles they view are OA. Notably, the most common mechanism for OA is not Gold, Green, or Hybrid OA, but rather an under-discussed category we dub Bronze: articles made free-

to-read on the publisher website, without an explicit Open license. We also examine the citation impact of OA articles, corroborating the so-called open-access citation advantage: accounting for age and discipline, OA articles receive 18% more citations than average, an effect driven primarily by Green and Hybrid OA. We encourage further research using the free oaDOI service, as a way to inform OA policy and practice.

Keywords: open access, open science, scientometrics, publishing, libraries, scholarly communication, bibliometrics, science policy

For citation: Piwowar H., Priem J., Larivière V., Alperin J.P., Matthias L., Norlander B., Farley A., West J., Haustein S. The State of OA: a Large-Scale Analysis of the Prevalence and Impact of Open Access Articles. *Scholarly Research and Information*. 2019;2(4):228-247. <https://doi.org/10.24108/2658-3143-2019-2-4-228-247>

Введение

Движение за открытый доступ (ОД) ко всей научной литературе насчитывает более пятнадцати лет. В последние годы ряд событий говорит о том, что многолетняя работа привела к кардинальным изменениям. Во-первых, фонды все больше внедряют поддержку публикаций ОД для грантополучателей. После Национального института здравоохранения США, утвердившего мандатом политику ОД в 2008 г. (<https://publicaccess.nih.gov/index.htm>), сделали обязательной публикации в открытом доступе для своих грантополучателей фонд Билла и Мелинды Гейтс (<http://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy>), Европейская комиссия (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf), Национальный фонд науки США (<https://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf>), the Wellcome Trust (<https://wellcome.ac.uk/press-release/wellcome-trust-strengthens-its-open-access-policy>) и другие. Во-вторых, был разработан ряд сервисов, повышающих удобство пользования ресурсами ОД и увеличивающих их ценность. Среди них — платформы дискавери ScienceOpen и 1Science, а также браузерные расширения Open Access Button, Canary Haz и Unpaywall. В-третьих, то, что Sci-Hub (сайт, предлагающий пиратский свободный доступ к полным текстам статей) собрал огромное количество пользователей, вызвало интенсивное обсуждение этики и эффективности платного доступа [1, 2]. Научные социальные сети ResearchGate и Academia.edu предлагают авторам все более популярные, но сомнительные решения по архивации своих работ [3, 4]. Наконец, возрастающая стоимость подписки, особенно в рамках Big Deals (крупных сделок), вынудила организации массово отказываться от подписок. Среди недавних примеров — Caltech, the University of Maryland, University of Konstanz, Université de Montréal и национальная система Перу [5–8]. Поскольку платный доступ в существующем виде становится слишком дорогим,

организации начинают рассматривать ОД как некий «План Б» для поддержания доступа к необходимой литературе [9].

Таким образом, открытый доступ вызвал новую волну вовлеченности и полемики в больших группах заинтересованных лиц. Возможно, мы приближаемся к очень важному моменту в развитии ОД и более того — в развитии системы научной коммуникации. Однако, несмотря на заметное оживление разработок и обсуждений ОД, мы нуждаемся в больших объемах достоверных данных о количественном росте и структуре литературы открытого доступа. Конкретизируя: нам необходим обзор состояния ОД, основанный на данных, который был бы (а) масштабным, (б) современным и (в) воспроизводимым. Эта статья пытается дать такой обзор, опираясь на новый свободно доступный сервис oaDOI, который находит ссылки на доступные легально научные статьи ОД. На основе данных oaDOI мы отвечаем на следующие вопросы.

Каков процент научной литературы открытого доступа и как он меняется в зависимости от издательства, дисциплины и года публикации?

Действительно ли публикации ОД цитируются лучше, чем платные?

В следующем разделе приводится краткий обзор литературы по теме статьи, описание наборов данных и используемых методов, а также детали определений категорий, используемых oaDOI. Далее приводятся результаты ответов на каждый вопрос исследования, за которыми следует общее обсуждение и выводы.

Обзор литературы

За пятнадцать лет исследований открытого доступа было опубликовано много литературы, исчерпывающий обзор которой не входит в задачи данной статьи. Последние глубокие обзорные исследования представлены в [10, 11]. Здесь мы сосредоточимся на кратком рассмотрении трех основных тем литературы по ОД: определении ОД и его подтипов, оценке распространенности ОД и исследованиях сравнительного влияния публикаций ОД.

Несмотря на обилие литературы по открытому доступу, сам термин остается несколько размытым [9]: слишком сложно предложить его авторитетное общепринятое определение. Наиболее распространенное определение содержится в тексте Будапештской инициативы открытого доступа (Budapest Open Access Initiative — BOAI) 2002 г. В нем ОД определяется как доступ, разрешающий свободное чтение контента и его вторичное использование, а именно возможность индексирования статей, загрузки их программными средствами или использование для любых других законных целей. На практике определение BOAI примерно эквивалентно популярной лицензии Creative Commons CC-BY [12]. Однако ряд других источников предпочитают менее строгое определение ОД, требуя лишь возможности бесплатно читать научную литературу онлайн [13] или чтобы она была цифровой, онлайн-овой и бесплатной [14]. Некоторые авторы считают, что ОД следует понимать как некий диапазон возможностей использования [15].

Исследователи выделяют несколько подтипов ОД; некоторые из них признаются практически всеми, некоторые остаются дискуссионными. Мы не стремимся дать их полный перечень, а приведем те, которые актуальны в контексте данного исследования.

- Libre ОД [16] очень похоже на определение BOAI: оно декларирует права пользователей на чтение и вторичное использование документов, в частности для автоматического индексирования, архивации и иных целей.
- Gratis ОД [16], в противоположность Libre, провозглашает право только на чтение.
- Золотой ОД: статьи публикуются в журнале открытого доступа (в котором все статьи открыты на сайте журнала). На практике журналы открытого доступа определяются как включенные в Directory of Open Access Journals (DOAJ) [17,18].
- Зеленый ОД: статьи публикуются в журнале, распространяемом по платной подписке, но архивируются авторами в открытом архиве. Открытые архивы — это либо тематические репозитории, такие как ArXiv, либо репозитории организаций (institutional repositories — IRs), обычно поддерживаемые университетами; архивированные статьи могут быть как опубликованными версиями, так и электронными препринтами [19]. Большинство статей Зеленого ОД не соответствуют определению BOAI: для них не разрешено вторичное использование; они относятся к типу Gratis ОД.
- Гибридный ОД: статьи публикуются в подписном журнале, но немедленно предоставляются

в свободный доступ под открытой лицензией за счет оплаты автором стоимости публикации (article processing charge — APC) [20, 21].

- Отложенный ОД: статьи публикуются в подписном журнале, но становятся свободно доступными по прошествии периода эмбарго [21, 22]).
- Научные социальные сети (Academic Social Networks — ASN): статьи распространяются их авторами через коммерческие социальные сети, такие как ResearchGate и Academia.edu. Одни исследователи рассматривают это явление как тип ОД [4, 23], другие считают, что это вообще не открытый доступ. В отличие от репозитория Зеленого ОД в сетях ASN не проверяется лицензионная чистота, поэтому примерно половина статей размещена там нелегально [24]. Этот факт вызывает озабоченность стабильностью контента, так как в октябре 2017 г., например, издатели потребовали удаления из ASN большого количества нелегально размещенных статей [25]. Поднимается также вопрос надежности и этики ASN [26]. В силу этих обстоятельств мы исключаем контент ASN из нашего определения ОД.
- Черный ОД: статьи, распространяемые на нелегальных пиратских сайтах, прежде всего, на SciHub и LibGen. Хотя в [27] эти статьи называются подтипом ОД, этот подход не получает поддержки в литературе по ОД. Учитывая это, мы исключаем контент SciHub и LibGen из нашего определения ОД.

Основываясь на консенсусе (а иногда на отсутствии такового) относительно определения ОД и его подтипов, будем использовать в этой статье следующее определение: статьи открытого доступа — это статьи, которые можно свободно читать в онлайн-е на сайте издателя или в репозитории.

Распространенность ОД

Оценка того, какой процент литературы представлен в открытом доступе, проводилась во многих работах, например в [17, 18, 23, 28–31]. Нам неизвестны подобные исследования, опубликованные после 2014 г. Две последние работы оценивают процент публикаций свободного доступа более чем в 50%, причем одна из них учитывает и ОД в строгом понимании, и публикации, размещенные в научных социальных сетях (ASN). Наиболее полное исследование [17] приводит следующие данные для статей 2011–2013 гг.: 12% статей свободно доступны на сайтах журналов, 6% — в репозиториях и 31% — через другие механизмы, включая ASN. В нем также отмечено, что количество документов ОД, опубликованных между 1996 и 2011 гг., возросло

на 4% с апреля 2013 по апрель 2014 г., причем размещение работ прошлых лет происходит в основном в Зеленом ОД. Анализ на уровне отдельных дисциплин совпадает с данными других исследований: наиболее высокий процент публикаций ОД наблюдается в биомедицинских науках и математике, наиболее низкий — в инженерных науках, химии и гуманитарных науках.

Особенно интересно исследование [17], в котором для поиска и идентификации документов ОД применялось автоматизированное «прочесывание» Интернета; более ранние исследования основывались на ручной проверке DOAJ, сайтов издателей, Google и Google Scholar (хотя было одно исключение — работа [32]). Применяв автоматизированные методы формирования выборки, авторы [17] получили сотни тысяч статей, значимо повысив статистическую достоверность и нюансировку результатов. Более того, разработав систему индексирования ОД-контента, они сняли главную озабоченность исследователей ОД, как отмечено в [29]: «Главным вызовом для исследователей... было отсутствие надежного индексирования журналов ОД и опубликованных в них статей». Представленная в [17] автоматизированная система очень точна: она неверно определяет ОД лишь в 1% случаев, находя при этом 75% всех онлайн-документов ОД; однако она не может различить Золотой и Гибридный ОД. Недостатком системы для исследователей является ее закрытость: ее нельзя использовать онлайн для дальнейших изысканий. Напротив, впоследствии данные были использованы для разработки коммерческой базы данных 1science (<http://www.1science.com/oanumbr.html>).

Преимущества ОД в аспекте цитирования

Сравнению показателей цитирования документов ОД и документов по платной подписке посвящены многочисленные исследования. Большинство из них фиксировало превышение цитирования документов ОД, обозначенное специальным термином «преимущества ОД в цитировании» (open access citation advantage — ОАСА); были даже созданы аннотированные библиографические указатели, отражающие эту литературу [33–35]. Однако единодушной поддержки явление ОАСА не получило. Многие работы, доказывающие преимущества ОД в цитировании, критиковались с точки зрения методологии [36], а исследование с применением метода случайной выборки не подтвердило этого преимущества [37]. Однако недавние работы, использовавшие разнообразные методы, вновь обнаружили явление ОАСА. Например, в [38] с использованием сложной статистической моде-

ли для снижения эффекта выборочного размещения авторами в ОД своих лучших работ показано небольшое, но значимое преимущество ОД в 8%. В работе [17] преимущество в 40% продемонстрировано на огромном массиве более чем в миллион статей с учетом нормализованного по дисциплине показателя цитирования. В [39] был применен естественный (не зависящий от выбора авторов) отбор статей, перешедших в ОД по истечении периода эмбарго; превышение цитирования для них составило 19%.

Методы

Определение ОД

■ Классификации

Мы распределяем все публикации на две категории: открытого (ОД) и закрытого доступа (ЗД). Как было сказано выше, мы считаем, что документ находится в открытом доступе, если его можно свободно читать в онлайн-на сайте издателя или в репозитории, все другие документы считаются закрытыми. Далее мы выделяем четыре непересекающиеся подкатегории ОД и в результате получаем пять категорий статей.

- Золотая категория: статьи, опубликованные в журнале открытого доступа, зарегистрированного в DOAJ.
- Зеленая категория: статьи, платные на издательском сайте, для которых имеется свободная копия в открытом репозитории.
- Гибридная категория: статьи, опубликованные в платных журналах на условиях открытой лицензии.
- Бронзовая категория: статьи, свободно доступные для чтения на сайте издателя, у которых отсутствует лицензия, явно описывающая режим использования.
- Закрытая категория: все другие статьи, включая все, доступные только через ASN или Sci-Hub.

Хотя эти категории, по сути, соответствуют пониманию соответствующих терминов в литературе по ОД, необходимы некоторые пояснения. Во-первых, мы, как многие другие авторы, не учитываем контент, распространяемый через научные социальные сети (ASN). Во-вторых, категории не пересекаются — и статьи на издательских сайтах имеют приоритет перед архивированными в репозиториях, то есть если статья опубликована в Золотом журнале и открытом репозитории, она попадает в Золотую категорию. Другими словами, издательский контент «отодвигает в тень» статьи, которые могли бы относиться к Зеленой катего-

рии. Такое понимание Зеленых статей (доступных в репозитории, но недоступных на сайте издателя) часто встречается в литературе по ОД (в том числе у автора терминов «Зеленый ОД» и «Золотой ОД» Стивена Харнада [19]), но признается не всеми. В некоторых работах допускается, что одна и та же статья может принадлежать и Золотой, и Зеленой категории; по сравнению с этим подходом наше исследование недооценивает количество Зеленых статей. Гибридные и Золотые статьи имеют общие характеристики, будучи открыты для чтения и вторичного использования, но различаются местом публикации, так как Гибридные статьи публикуются в журналах, не относящихся к журналам ОД согласно DOAJ, и тем, что они могут иметь период эмбарго. Мы также вводим новую Бронзовую подкатегорию. Бронзовые статьи похожи и на Золотые, и на Гибридные тем, что публикуются на сайтах издателей. Но их нельзя считать Золотыми, так как опубликовавшие их журналы не учтены в DOAJ, и нельзя считать Гибридными, поскольку для них отсутствует явным образом обозначенная лицензия. И хотя отсутствие лицензии может быть ненамеренным, по умолчанию такие статьи доступны только для чтения, но не для вторичного использования. Кроме того, для Бронзовых статей нельзя определить, будут ли они доступны для чтения всегда или только в определенный период.

Наконец, необходимо отметить, что хотя выбранные нами категории отражают понимание их в литературе по ОД, они не вполне отражают сложную картину современной системы научных публикаций. К примеру, латиноамериканские организации SciELO и Redalyc являются одновременно и издателями, и репозиториями, поэтому многие статьи на их сайтах невозможно однозначно отнести к одной из выделенных нами категорий [40].

■ Система oaDOI

Присвоение категорий происходило через запросы к сервису oaDOI по DOI статей. Сервис возвращает ссылку на легально доступную ОД-версию статьи при наличии таковой (<https://oadoi.org/>). База oaDOI включает записи для всех 88 миллионов DOI CrossRef. Сервис обходит, агрегирует и нормализует данные из множества источников, в числе которых — PMC (<https://www.ncbi.nlm.nih.gov/pmc/>), BASE (<https://www.base-search.net/about/en/>), DOAJ (<https://doaj.org/>) и множество сайтов репозиториях организаций и издателей. Система oaDOI предоставляет доступ к API без каких-либо ограничений, что позволяет поддерживать другие сервисы и разработки. На момент подготовки статьи oaDOI обрабатывал около 500 тысяч запросов в день, что примерно вдвое больше, чем

в Sci-Hub [1, 41]. Большинство запросов приходит из примерно 700 научных библиотек в ситуациях, когда они помогают читателям найти статьи, недоступные по подписке, решая тем самым проблему их обнаружения [31]. Сервис oaDOI — основа работы браузерного расширения Unpaywall, помогающего пользователям найти легально доступные ОД-версии документов; сейчас Unpaywall активно используют около 80 тысяч человек. oaDOI основан на открытом коде, он бесплатен и открыт через общедоступный API.

■ Точность oaDOI

Для оценки точности автоматического определения ОД мы отобрали случайным образом 500 статей из нашей основной выборки Crossref-DOIs, описанной ниже. Мы вручную провели поиск каждой статьи в Интернете, чтобы определить, доступна ли она на сайте издателя или на другом сайте: репозитории организации, научной социальной сети или на персональной странице. Разрешение DOI проводилось со страницы <https://doi.org>. Если полный текст был доступен через эту ссылку, статья маркировалась как доступная на сайте издательства. Если требовался оплаченный доступ, проводился поиск по заглавию в Google Scholar (GS) и Google, чтобы найти альтернативные версии (препринты или архивированные копии). Если полный текст был найден на любом издательском сайте или в открытом репозитории, он маркировался как свободно доступный в архиве. Если единственная открытая копия была доступна в научной социальной сети (Academia.edu или ResearchGate), делалась пометка об этом, но в рамках этого исследования она не относилась ни к одной из категорий ОД, а маркировалась как Закрытый доступ.

Ниже показаны итоговые показатели работы oaDOI в сравнении с ручной проверкой доступности статей. В онлайн-приложениях (см. раздел «Дополнительная информация») приводится полный набор сравниваемых данных. На их основе вычислялись полнота и точность работы системы. Показатель полноты отвечает на вопрос «Если статья находится в открытом доступе, насколько часто oaDOI верно определяет ее открытость?». Полученное значение полноты составило 77%, то есть для 77% свободно доступных статей oaDOI правильно определил это. Показатель точности отвечает на вопрос «Когда oaDOI определяет, что статья свободно доступна, в каком проценте случаев это верно?». Полученное значение точности составило 96,6%, то есть в 96,6% случаев ответ oaDOI был правильным.

Эти результаты примерно соответствуют полученным в исследовании [17] в результа-

те работы их автоматизированной системы: 86,4 и 99,1% соответственно. Показатели полноты и точности также рассчитывались на выборке в 500 документов; статистическая погрешность в обоих исследованиях составила $\pm 4,5\%$. В работе [17] выборка формировалась на временном интервале с 1996 г., в то время как в данной работе ограничения по времени не применялись. Период с 1996 г. был более однородным в плане распространенности ОД, что отчасти объясняет лучшие показатели исследования коллег.

Сервис oaDOI оптимизирован под показатель точности в большей степени, чем под полноту. Очень высокий показатель точности говорит о том, что любые оценки на основе базы oaDOI можно считать консервативными по отношению к реальным показателям ОД в потоке литературы. То есть мы можем с уверенностью утверждать, что если oaDOI показывает определенный процент ОД, реальный процент не ниже этого, а скорее выше, учитывая то, что показатель полноты далек от идеала. Другими словами, oaDOI порождает очень мало ложноположительных результатов (случаев, когда он ошибочно считает статью открытой), но относительно много ложноотрицательных (когда он ошибочно считает статью закрытой), как показано в таблице 1. Мы планируем усовершенствовать систему, чтобы повысить полноту, не снижая показатель точности.

Выборки, сформированные для данного исследования

Показатели трех выборок научных ресурсов, имеющих DOI, представлены в таблице 2 и описаны ниже.

■ Выборка из Crossref

Первая выборка «Crossref-DOIs» — случайная выборка в 100 тысяч журнальных статей,

имеющих DOI, любых лет публикации. На май 2017 г. в базе Crossref было 88 млн записей. Чтобы исключить книги, наборы данных и другой контент, мы делали отбор по значению элемента метаданных type равному journal-article; всего таких записей в базе оказалось 66 млн. Для проверки точности Crossref мы проверили вручную 150 единиц контента и обнаружили, что 93% действительно были журнальными статьями, а оставшиеся 7% — в основном первыми разделами журналов типа страницы содержания или инструкции для авторов.

Назначение этой выборки — грубая оценка научной литературы в целом. Это определяет ее сильные и слабые стороны. Одна из слабых сторон — то, что хотя в Crossref есть информация о цитировании и дисциплине, эти данные далеко не полные, следовательно, их нельзя было использовать. Другая заключается в том, что исследователи в области наукометрии и ОД в основном полагаются на иные источники, особенно на Scopus и Web of Science (WoS) для оценки потока литературы в целом, поэтому наши результаты трудно сравнивать с результатами прежних исследований. Наконец, известно, что DOI реже присваиваются публикациям в некоторых дисциплинах, в частности в гуманитарных науках [42], в некоторых регионах, в особенности в развивающихся странах, и для более старых работ [43]; следовательно, эти сегменты хуже представлены в нашей выборке. Однако известно, что Scopus и Web of Science также недостаточно учитывают важные сегменты литературы [44], так что этот недостаток свойствен не только Crossref. Кроме того, выборка из Crossref имеет важное преимущество по сравнению с индексами. Хотя никакая выборка научной литературы не может быть полной во всех аспектах, Crossref содержит больше записей: в июле 2017 года

Таблица 1. Показатели работы прототипа сервиса oaDOI, использованного в данной работе

Table 1. Accuracy of the prototype version of the oaDOI service used in this study

	oaDOI определяет как ОД / oaDOI reports Open	oaDOI определяет как ЗД / oaDOI reports Closed	Подсчет вручную всего / Manual count Total (ground truth)
Открытый доступ / Open	144	43	187
Закрытый доступ / Closed	5	308	313
Всего/Total	149	351	500

DOI: 10.7717/peerj.4375/table-1

Таблица 2. Обобщенные данные выборок, использованных в данной работе

Table 2. Summary of samples used in this study

Название выборки / Sample name	Размер выборки / Sample size	Наполнение выборки / Population sampled	Целевое назначение / Purpose	Объем выборки / Population size
Crossref-DOIs	100,000	All journal articles with Crossref DOIs, all years	Estimate percentage of the literature that is OA	66,560,153
WoS-DOIs	100,000	All citable WoS articles with DOIs, 2009–2015	Estimate citation impact of recent OA papers, and also OA prevalence by discipline	8,083,613
Unpaywall-DOIs	100,000	All articles accessed by Unpaywall users over a 1-week period in 2017	Estimate percentage of OA experienced by users of the Unpaywall extension	213,323

DOI: 10.7717/peerj.4375/table-2

в нем было 67 млн журнальных статей, а в Scopus — 30 млн (<https://www.elsevier.com/solutions/scopus/content>). Crossref — система, открытая для использования, а Scopus и Web of Science доступны по подписке; благодаря этому преимуществу данные исследования по Crossref можно размещать в открытом доступе, что позволяет воспроизводить и повторять результаты нашего исследования. Однако нам пришлось обратиться к Web of Science, чтобы оценить цитирование и распределение по дисциплинам статей ОД, поскольку данные Crossref для этого недостаточно.

■ Выборка из Web of Science

Вторая выборка, WoS-DOIs, — случайная выборка в 100 тысяч журнальных статей с DOI, индексируемых в Web of Science. Она была сформирована из локальной базы WoS, используемой в Observatoire des sciences et des technologies (OST) в Университете Квебека в Монреале. В выборку включены только статьи, определяемые WoS как потенциально цитируемые (citable items), благодаря чему в нее не попадает нерецензируемый контент типа материалов «от редактора» и новостей. Выборка ограничена периодом с 2009 по 2015 г. — периодом, когда присвоение DOI стало распространенным явлением. Выборка формировалась случайным образом из массива в 8 млн статей и обзоров, имеющих DOI, по состоянию на май 2017 г.

Поскольку выборка из WoS ограничивалась определенным периодом и наличием DOI, она не подходила для оценки процента документов ОД в общем массиве литературы. Однако в других аспектах она лучше, чем выборка из Crossref: для каждой статьи есть информация о количестве

цитирований и информация о дисциплине (о чем подробнее ниже). Поэтому мы используем эту выборку для оценки распространенности ОД в различных дисциплинах и влияния на цитирование ОД в последние годы. Мы не рекомендуем сравнивать показатели ОД в двух первых выборках, поскольку они по-разному формировались.

Документы в выборке WoS-DOIs классифицировались по журнальной классификации Национального фонда науки (National Science Foundation — NSF). Каждому журналу в ней присваивается одно название дисциплины (категория верхнего уровня) и одно название специальности (категория детального уровня). Поскольку это классификация журналов, все статьи журнала получают названия дисциплины и специальности, присвоенные журналу. Недостаток этой классификации — в том, что мультидисциплинарные журналы (например, Nature, PNAS, PLOS ONE) относятся к одной категории биомедицинских исследований, хотя публикуют множество статей других направлений исследований. Для таких журналов мы выполнили постатейное индексирование. Каждая статья из списка мультидисциплинарных журналов была приписана к той специальности NSF, которая чаще всего фигурировала в ее списке цитируемой литературы. Другими словами, каждая статья из этих журналов классифицировалась на уровне статьи, а не журнала, и относилась к той специальности, статьи по которой она чаще всего цитировала.

Мы оцениваем относительное влияние закрытых и открытых статей с помощью ссылок на них как показателя их научного влияния. Но есть ряд свойств статей, которые могут отрицательно вли-

ять на такую оценку. Главные из них — дисциплина (некоторые области исследования цитируются лучше, чем другие) и возраст статьи (у статей прошлых лет больше времени для накопления ссылок). Чтобы учесть это обстоятельство, мы вычислили нормализованное ожидаемое число ссылок (цитат) для каждой статьи на базе ее возраста и специальности, сравнив его со средним количеством цитат для подобных статей.

На основе этого подхода каждая статья получила показатель среднего относительного цитирования (average relative citation — ARC). Значение ARC 1.0 показывает, что документ был процитирован в соответствии с ожидаемым значением для документов, опубликованных в тот же год по той же специальности NSF, а значение ARC больше или меньше единицы означало, что документ был процитирован лучше или хуже среднего значения соответственно. Используя эти показатели, нормализованные по области исследования, мы можем сравнить цитирование между дисциплинами и годами публикации. Мы можем также вычислить средние значения ARC для групп статей «все открытые статьи» и «все закрытые статьи» и сравнить их нормализованные показатели цитирования. При анализе результатов на уровне дисциплин NSF мы не показываем данные по гуманитарным наукам и искусству, поскольку они слабо представлены и в Web of Science вообще и в плане наличия DOI в частности.

■ Выборка Unpaywall

Третья выборка, Unpaywall-DOIs, — случайная выборка в 100 тысяч статей, которые открыли пользователи браузерного расширения Unpaywall в течение одной недели. Мы собрали данные об IP-адресах и DOI для запросов к oаDOI через Unpaywall, поступивших с 5 по 11 июня 2017 г. За этот период всего поступило 374 703 запроса с 213 323 уникальными DOI с 42 894 уникальными IP-адресами; из этих данных случайным образом было выбрано 100 тысяч DOI.

На основе этой выборки мы оценивали распространенность документов ОД в запросах пользователей Unpaywall. Эта выборка удобна для оценки того, что хотят прочитать пользователи; она позволяет нам приблизительно оценить процент литературы ОД в их предпочтениях. Однако эта выборка серьезно ограничена демографией: мы не знаем, откуда приходят пользователи, а по IP-адресам видим, что преобладают пользователи США. Кроме того, мы не можем обобщить результаты с учетом уровня образования пользователей, их специальности и целей изучения научной литературы.

Результаты

Вопрос 1. Каков процент литературы открытого доступа?

■ Сколько литературы находится в открытом доступе?

На выборке Crossref-DOIs мы обнаружили, что 27,9% (95% CI [27,6–28,2]) всех статей, имеющих DOI, находятся в открытом доступе. Это означает, что имеется 18,6 млн статей ОД с DOI Crossref (95% CI [18,4–18,8]). Таково общее количество статей, которые может идентифицировать oаDOI. С учетом того что, как было описано в разделе «Методы», сервис oаDOI находит только 77% документов ОД по сравнению с поиском вручную, мы можем приблизительно оценить количество документов ОД, которые не обнаруживает oаDOI, еще в 3,5 млн.

Те, кто читает литературу, пользуясь расширением Unpaywall, обнаруживают гораздо больше документов ОД: 47% документов из выборки Unpaywall (95% CI [46,7–47,3]) — открытого доступа. В основном это обусловлено годом публикации статьи: поскольку эта выборка базируется на поведении реальных читателей, в ней преобладают статьи последних лет; половина из них опубликованы за последние два года. Как показано ниже, в разделе «Как меняется количество документов ОД в зависимости от года публикации», статьи последних лет чаще являются общедоступными, чем более старые.

■ Какие типы открытого доступа наиболее распространены?

Как показано на рисунке 1 и в таблице 3, соотношение подтипов ОД довольно ровное для всех трех выборок. Отчасти это объясняется тем, что самоархивированные статьи считаются относящимися к Зеленому ОД, только если для них нет соответствия на сайте издательства; таким образом, статьи Зеленого ОД находятся в тени Золотого, Гибридного и Бронзового ОД. Во всех выборках преобладает процент Бронзового ОД. Это особенно интересно, поскольку очень мало исследователей обращали внимание на его роль. Мы просмотрели небольшое количество статей Бронзового ОД, чтобы лучше понять особенности этой подкатегории; мы обнаружили, что хотя многие «бронзовые» статьи имели период эмбарго на сайтах платных издательств, почти половина была опубликована в журналах полностью открытого доступа (100% ОД), но они не были зарегистрированы в DOAJ и не обозначали явным образом условия лицензии (CC-BY или другие). Такие журналы лучше было бы называть «темно-золотыми» или «тайными золотыми», чем бронзовыми. Углубленное изучение Бронзово-

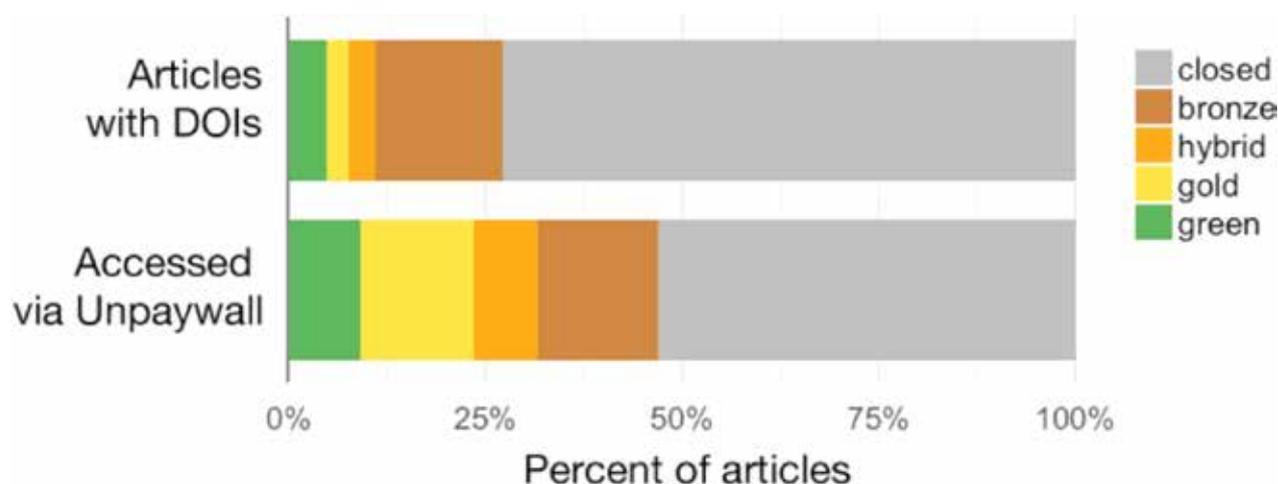


Рис. 1. Процент статей со статусом ОД: выборка Crossref-DOIs против выборки Unpaywall-DOIs

Fig. 1. Percent of articles by OA status, Crossref-DOIs sample vs Unpaywall-DOIs sample

DOI: 10.7717/peerj.4375/fig-1

Таблица 3. Процент литературы ОД по типам в трех выборках в 100 тысяч статей с доверительными интервалами в 95%

Table 3. Percent of the literature that is OA, by type, in three samples of 100,000 journal articles, with 95% confidence intervals

Тип доступа / Access type	Все журнальные статьи с DOI за все годы / Crossref-DOI All journal articles with Crossref DOIs, all years. ("Articles with DOIs" in Fig. 1)		Все потенциально цитируемые статьи WoS с DOI с 2009 по 2015 г. / WoS-DOIs All citable WoS articles with DOIs, 2009–2015		Все статьи, к которым обратились пользователи Unpaywall за одну неделю 2017 г. / Unpaywall-DOIs All articles accessed by Unpaywall users over a 1-week period in 2017	
	estimate, %	95% CI	estimate, %	95% CI	estimate, %	95% CI
ОД (все типы) / OA (all types)	27,9	27,6–28,2	36,1	36,0–36,2	47,0	46,7–47,3
Бронзовый ОД / Bronze OA	16,2	16,0–16,5	12,9	12,6–13,2	15,3	15,0–15,6
Гибридный ОД / Hybrid OA	3,6	3,3–3,9	4,3	4,0–4,6	8,3	8,0–8,6
Золотой ОД / Gold OA	3,2	2,9–3,5	7,4	7,1–7,7	14,3	14,0–14,6
Зеленый ОД / Green OA	4,8	4,5–5,1	11,5	11,2–11,8	9,1	8,8–9,4
Закрытый доступ / Closed	72,0	71,8–72,4	63,9	63,8–64,0	53,0	52,7–53,3

DOI: 10.7717/peerj.4375/table-3

го ОД не входит в задачи настоящего исследования, оно будет выполнено в последующих работах.

■ Как меняется количество документов ОД в зависимости от года публикации?

На рисунке 2 представлены данные по количеству (рис. 2А) и процентам (рис. 2В) документов в зависимости от категории и года публикации. В последние 20 лет количество статей ОД непрерывно возрастает, причем эта тенденция не замедляется. Чем «моложе» статья, тем более вероятно, что она окажется открытой; среди статей последнего обследованного года, 2015-го, обнаружено наибольшее число открытых: 44,7% (95% CI [43,3–46,2%]), из них 17,6% бронзовых (95% CI [16,2–19,1]), 9,4% гибридных (95% CI [8,0–10,9]), 11,3% золотых (95% CI [9,9–12,8]) и 6,3% зеленых (95% CI [4,9–7,8]). В 2015 году было опубликовано более 1 млн статей. Тенденция роста числа статей ОД в основном обусловлена значительным увеличением Золотого и Гибридного ОД с 2000 г. Тем не менее более 20% статей, опубликованных до наступления «цифрового века», находятся в свободном доступе. Большинство таких статей — Бронзовые; можно предположить, что это связано с окончанием периода эмбарго, но это требует дальнейшего изучения. Процент статей Бронзового ОД остается поразительно стабильным для всех лет издания.

Следует с особой осторожностью подходить к оценке количества и процентов Зеленых статей. Это обусловлено несколькими факторами. Во-первых, в отличие от статей ОД на издательских сайтах (Золотых, Бронзовых и Гибридных) дата пре-

доставления статьи в открытом доступе обычно отличается от даты первой официальной публикации. Нередко авторы самоархивируют статьи гораздо позже (а иногда и раньше, если речь идет о препринтах) даты их официальной публикации. В статье [17] это явление Зеленого ОД называется *backfilling* (ретроспективное наполнение). Таким образом, графики не могут отразить рост Зеленого ОД во времени: это требовало бы продолжительных наблюдений за несколько лет, что не входит в задачи настоящей работы. Фактически графики отражают только число и проценты Зеленых статей в зависимости от года их официальной публикации. Во-вторых, многие статьи нельзя легально самоархивировать до истечения периода эмбарго, что предположительно влияет на появление видимого плато на графике рисунка 2 для Зеленого ОД. Наконец, как было отмечено выше, многие самоархивированные статьи могли бы считаться Зелеными, если бы не были «затенены» Золотой, Бронзовой или Гибридной версией этой статьи. Более подробно рост «затененных» Зеленых статей показан на рисунках SA2 и SA3².

■ Как меняются показатели ОД в зависимости от издательства?

Мы проанализировали подмножество выборки Crossref-DOIs по издательствам (указанным в метаданных Crossref), чтобы понять, как распределяется процент ОД и типы ОД для разных издательств за период с 2009 по 2015 г. Как видно на рисунке 3А, крупнейшие издательства, возглавляемые «Эльзевиром», публикуют наибольшее количество статей

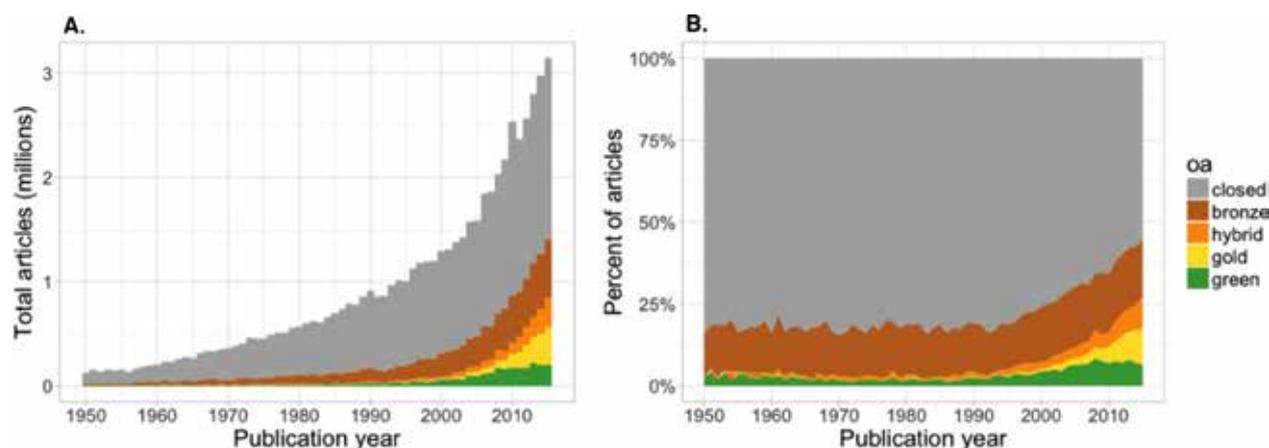


Рис. 2. Число статей (А) и процент статей (В), открытого доступа по случайной выборке из 100 000 статей с Crossref DOI

Fig. 2. Number of articles (A) and proportion of articles (B) with OA copies, estimated based on a random sample of 100,000 articles with Crossref DOIs

DOI: 10.7717/peerj.4375/fig-2

2 Номера рисунков, начинающиеся с "S", отсылают к дополнительным материалам к статье <http://dx.doi.org/10.7717/peerj.4375#supplemental-information>. Прим. переводчика.

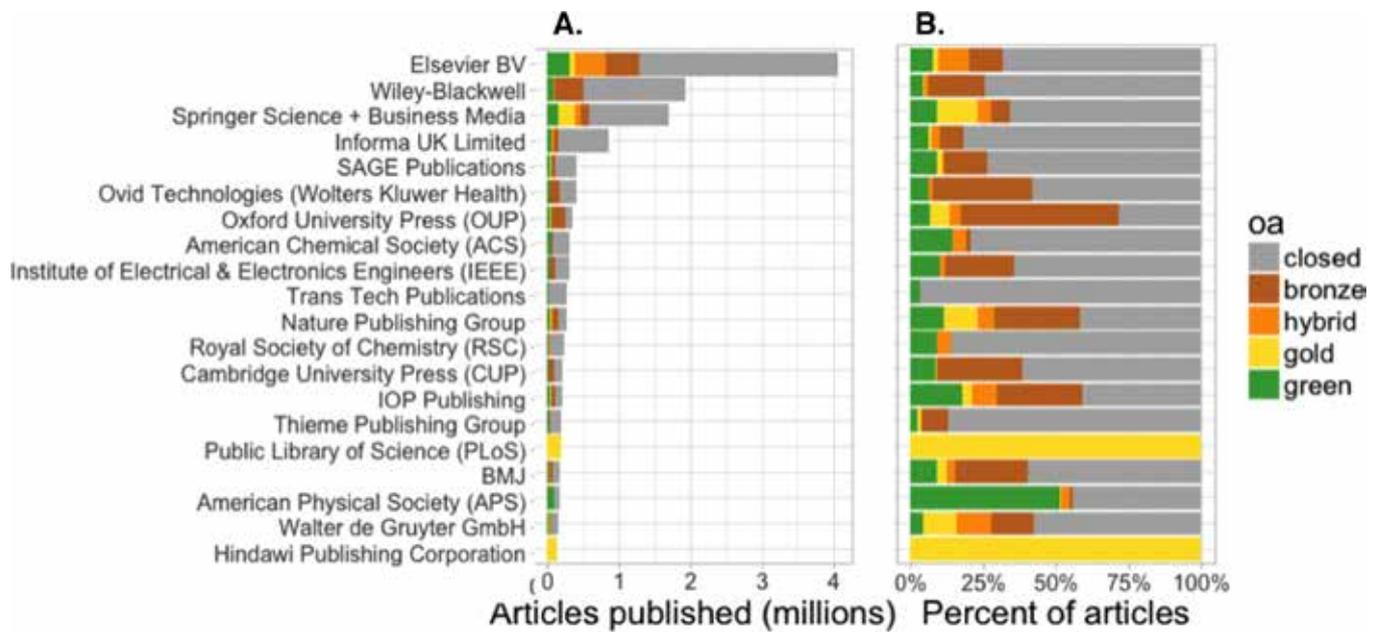


Рис. 3. Количество (А) и процент (В) статей ОД по издательствам для 20 крупнейших издательств. На основе выборки в 27 894 статей с Crossref DOI, опубликованных в период 2009–2015 гг.

Fig. 3. Number (A) and proportion (B) of articles with OA copies, by publisher, for the 20 most prolific publishers. Based on sample of 27,894 Crossref DOI-assigned articles published between 2009–2015

DOI: 10.7717/peerj.4375/fig-3

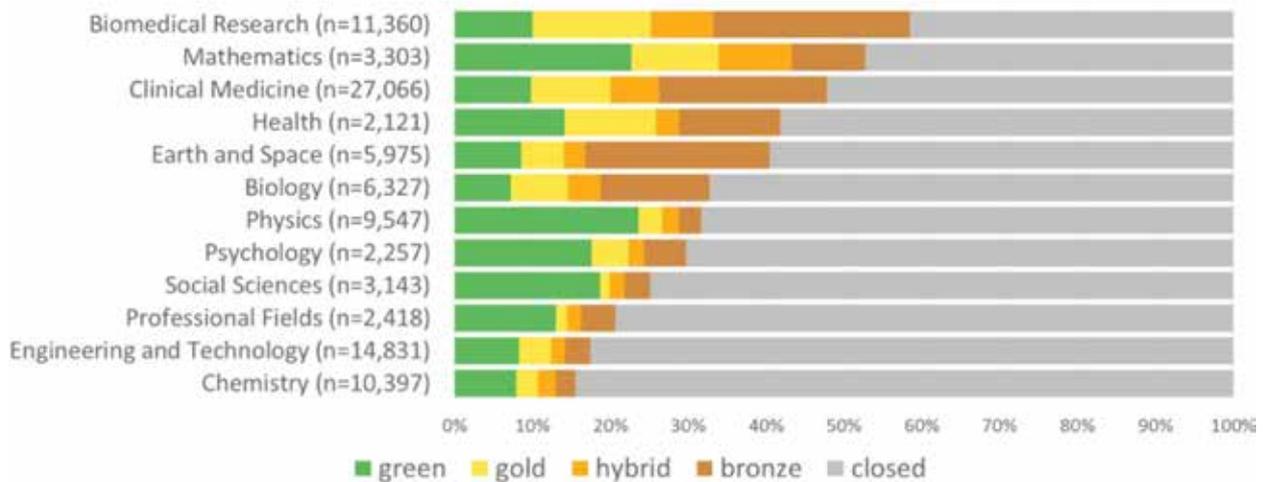


Рис. 4. Процент различных типов доступа в случайной выборке статей и обзоров из WoS, имеющих DOI, опубликованных в период с 2009 по 2015 г. по дисциплинам NSF (кроме Arts and Humanities)

Fig. 4. Percentage of different access types of a random sample of WoS articles and reviews with a DOI published between 2009 and 2015 per NSF discipline (excluding Arts and Humanities)

DOI: 10.7717/peerj.4375/fig-4

открытого доступа. Что касается процента статей ОД, только издательства PLOS и Hindawi в первой двадцатке показывают 100% открытый доступ. Издательства Oxford University Press, Nature Publishing Group, IOP Publishing и the American Physical Society (APS) предоставляют в открытый доступ более половины статей. В случае APS ОД обеспечивается в основном размещением статей в репозиториях, таких как arXiv. Более подробно данные по репозиториям приведены на рисунке SA1.

■ Как распределяются статьи открытого доступа по дисциплинам?

Для определения распространенности статей ОД в отдельных дисциплинах мы использовали выборку WoS-DOIs, поскольку из метаданных индекса WoS легко извлечь информацию о дисциплине. Результаты представлены на рисунке 4. В биомедицинских науках и математике больше половины публикаций свободно доступны, в то время как в химии и инженерных науках этот показатель меньше 20%.

На рисунке 4 явно обозначена популярность Зеленого ОД в физике и математике, где более 20% работ доступны только в репозиториях (в основном в arXiv). Гибридные статьи особенно распространены в математике (9,4%), биомедицинских исследованиях (8,1%) и клинической медицине (6,3%); в журналах Золотого ОД чаще всего публикуются авторы исследований по биомедицине (15,3%), здравоохранению (11,7%), математике (11,2%) и клинической медицине (10,3%).

Большое разнообразие в пропорциях ОД наблюдается на более детальном уровне специальностей NSF (рис. SA5). Более 80% статей ОД мы наблюдаем в астрономии и астрофизике (87%), фертильности (86%), тропической медицине (84%) и эмбриологии (83%). На противоположной стороне спектра находятся фармакология (7%), неорганическая и ядерная химия (7%), химическая технология (9%), где за платёжным барьером (paywall) находится более 90% статей.

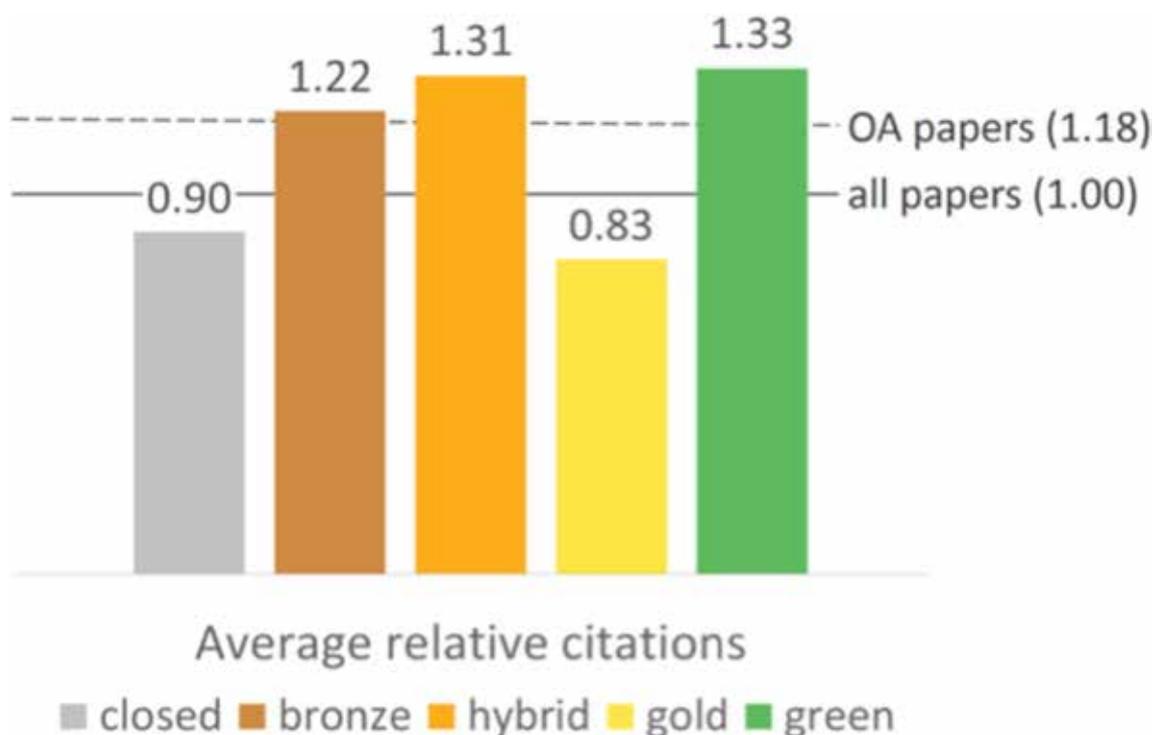


Рис. 5. Средние относительные значения цитирования для различных типов доступа по случайной выборке статей и обзоров из WoS, имеющих DOI, опубликованных с 2009 по 2015 г.

Fig. 5. Average relative citations of different access types of a random sample of WoS articles and reviews with a DOI published between 2009 and 2015

DOI: 10.7717/peerj.4375/fig-5

Вопрос 2. Каково цитатное влияние статей открытого доступа?

Преимущества ОД в цитировании (ОАСА) подтверждается сравнением средних относительных показателей цитирования разных категорий доступа. Платные статьи цитируются на 10% ниже среднемировых значений (ARC = 0,90), а свободно доступные статьи получают в среднем на 18% больше цитирований, чем среднее ожидаемое значение (ARC = 1,18). Однако цитатное влияние различно для разных типов ОД. Наиболее ярко преимущество проявляется для статей, доступных только в Зеленом ОД (ARC = 1,33) и Гибридном ОД (ARC = 1,31), которые наиболее часто цитируются (более чем на 30% выше ожидаемых значений). Бронзовые статьи цитируются на 22% лучше среднемировых

значений; Золотые статьи получают значение ARC в 0,83, что на 17% ниже среднемировых значений и на 9% ниже значений для платных статей. Эти данные показаны на рисунке 5.

Как показано на рисунке 6, эти тенденции меняются с годами. Хотя показатель ARC платных статей остается ниже среднемировых значений весь изученный период, он меняется от 0,86 в 2009 г. до 0,93 в 2014 и 2015 гг. Для всех типов ОД показатель ARC всегда выше среднемировых значений, меняясь в интервале от 1,15 до 1,22. Эти флуктуации связаны с типом доступа, причем цитатное влияние Гибридных документов с годами возрастает. Среднее значение цитирования Зеленых статей остается относительно стабильным, а наивысшее значение достигается в 2015 г. Бронзовыми и Гибридными

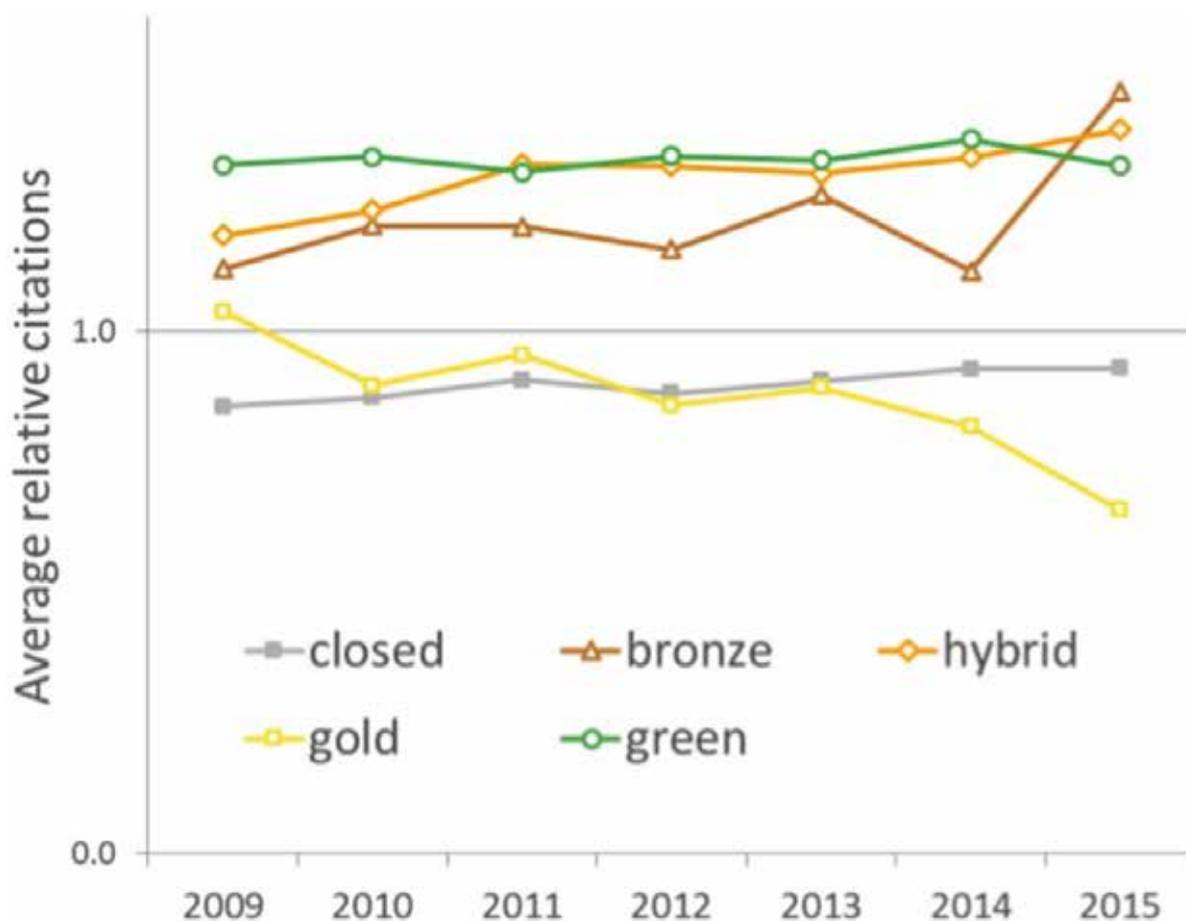


Рис. 6. Процент и влияние статей разных типов доступа по случайной выборке из статей и обзоров WoS, имеющих DOI, по годам публикации

Fig. 6. Percentage and impact of different access types of a random sample of WoS articles and reviews with a DOI, by year of publication

DOI: 10.7717/peerj.4375/fig-6

статьями. Из всех типов ОД только Золотые статьи показывают снижение от года к году. Однако результаты последних лет основаны на узком окне цитирования и могут меняться в дальнейшем с накоплением цитат.

Обсуждение и выводы

В наши дни доступ к научной литературе — это центральная тема обсуждений в научном сообществе. Все больше научных фондов требуют от своих грантополучателей предоставлять результаты исследований в открытом доступе. В то же время рост стоимости подписки вынуждает отказываться от нее все больше университетских библиотек. В этих условиях возник ряд способов получить доступ к литературе, легальных или нелегальных. Используя данные, полученные одним из таких способов, oaDOI, настоящая работа пытается ответить на два важных для исследователей вопроса: какой процент литературы свободно доступен в разных типах ОД и каково в среднем научное влияние этой литературы. Для ответа на них мы использовали три больших выборки (1) 100 тысяч статей, имеющих DOI Crossref, что позволило нам оценить процент статей ОД в общем объеме статей; (2) 100 тысяч журнальных статей из WoS, имеющих DOI, что позволило оценить научное влияние статей ОД и не-ОД; (3) 100 тысяч статей, к которым обращались пользователи браузерного расширения Unpaywall, что позволило оценить процент статей ОД в этом массиве.

Мы обнаружили, что 28% журнальных статей в онлайн свободно доступны (в выборке Crossref-DOIs). К радости сторонников ОД, этот процент стабильно растет на протяжении последних 20 лет, особенно для Золотых и Гибридных статей. Последний обследованный год, 2015-й, показал наибольший процент ОД, 45%, а также рекордное количество статей ОД, опубликованных за год. Высокий процент статей ОД в последние годы наряду с предпочтениями читателей к свежим статьям приводит к благоприятной ситуации: фактически процент статей ОД, к которым они обращаются, выше среднего по большой выборке. Пользователи Unpaywall через oaDOI получают свободный доступ к почти половине статей — к 47%. Этот эффект определенно имеет место не только для пользователей Unpaywall: скорее всего, все читатели обычно предпочитают новые статьи, а значит, и они выигрывают от роста количества Золотых, Бронзовых и Гибридных статей в текущем потоке. Для того чтобы оценить это количественно, нужны дополнительные исследования.

Интересно отметить, что большинство статей ОД — Бронзовые, то есть размещенные на сайтах

издателей без лицензии вообще или без открытой лицензии. Это особенно удивительно, поскольку об этом типе статей редко пишут, а они явно заслуживают большего внимания со стороны сторонников ОД. В частности, Бронзовому ОД нужно уделить больше внимания при разработке политики ОД, так как в отличие от других статей на издательских сайтах для Бронзовых разрешено только чтение, но не вторичное использование (Gratis OA). Дальнейшего изучения требуют характеристики Бронзового ОД. Какое количество статей этого типа, в принципе, доступны под открытыми лицензиями, но это не обозначено явным образом? Возможно, Бронзовые статьи преимущественно не рецензируются? Каков процент статей отложенного доступа в этой категории? Сколько Бронзовых статей свободно доступны в целях продвижения и рекламы и как долго они остаются в свободном доступе? Сколько из них публикуются в неявных Золотых журналах, не включенных в DOAJ? Почему такие журналы не размещают лицензии для своего контента и можно ли стимулировать их идти в этом направлении? Ответы на эти и другие вопросы нельзя дать в настоящей статье, но они могут стать стимулом для дальнейших исследований ОД.

Лишь 7% всей литературы (и 17% литературы ОД) — в Зеленом открытом доступе. На первый взгляд это может показаться разочарывающим результатом, учитывая годы пропаганды Зеленого ОД и растущее количество мандатов Зеленого ОД [45]. Однако для ОД в целом есть основания для оптимизма. Во-первых, много статей, архивированных в репозиториях, в данном исследовании не считались Зелеными, поскольку на платформах издателей они доступны по модели Золотого, Гибридного или Бронзового ОД. Такие «затененные» Зеленые статьи образуют надежную страховочную сеть, которая обеспечит свободный доступ к контенту, если издатели его прекратят. Для определения распространенности «затененных» статей в различных дисциплинах нужны дополнительные исследования. Во-вторых, многие авторы архивируют свои статьи не только текущего, но и прошлых лет; это означает, что хотя график Зеленого ОД не показывает стабильного ежегодного роста, как Золотой и Гибридный, это может измениться по мере того, как авторы будут все больше наполнять архивы ранее опубликованными работами. Это предположение подтверждается результатами работы [46]. Наконец, сравнительно низкое количество Зеленых статей оставляет надежду на грядущий устойчивый рост. Хотя большинство журналов крупнейших издательств (Elsevier, Wiley, Springer и других) разрешают самоархивирование, некоторые иссле-

дования показали, что лишь небольшой процент статей этих издателей заархивирован в репозиториях. Например, в [47] на выборке статей по исследованию здоровья в мире продемонстрировано, что только для 39% из них авторы воспользовались правом на самоархивирование.

Полученные нами результаты подтверждают преимущество статей ОД в цитировании, ранее выявленное в других исследованиях: открытые статьи получают на 18% больше ожидаемого количества цитирований. Отчасти это преимущество можно объяснить тем, что при открытом доступе со статьями могут ознакомиться, а значит и процитировать, больше читателей, но такое объяснение трудно подтвердить объективными данными и против него есть много возражений. Наиболее распространенное — так называемый «постулат смещения выбора» (*selection bias postulate*) [48]. Он подразумевает, что авторы продвигают в открытый доступ свои наиболее сильные работы. В настоящей работе не изучались причины и направленность преимущества статей ОД в цитировании, в нем только подтверждено его существование на очень большой выборке, репрезентативной для всего потока литературы. В преимуществе статей ОД в цитировании также могут иметь значение требования фондов: наиболее влиятельные фонды чаще выдвигают требование открытого доступа; кроме того, хорошо профинансированные работы имеют преимущества в цитировании над работами, выполненными на скромные средства [49]. Интересный факт: меньше всего цитируются Золотые работы, вероятно, из-за роста количества новых небольших журналов этой категории. Часть этих журналов издается в регионах, которые исторически не охватываются WoS, публикуются не на английском языке и не считаются престижными, поскольку у них не было достаточного времени на завоевание позиций и накопление цитирования [23]. С другой стороны, на проигрыш Золотых статей в цитировании, вероятно, влияет рост числа мегажурналов, таких как PLOS ONE [50]. Каковы бы ни были причины, относительно низкое влияние Золотого ОД означает, что общее преимущество в цитировании ОД имеет место благодаря Зеленому, Гибричному и Бронзовому контенту. Суммируя полученные результаты, скажем, что хотя на наблюдаемую разницу в цитировании могут влиять некоторые факторы, которые трудно оценить, факт остается фактом: ученые скорее прочтут и процитируют доступные работы, чем недоступные. Мы надеемся, что наличие свободно доступного индекса контента ОД поможет дальнейшим исследованиям явления преимущества ОД в цитировании.

Сравнительно высокий процент статей ОД, особенно среди пользователей Unpaywall, может сильно повлиять на деятельность научных библиотек. Им все труднее оплачивать подписные пакеты в рамках «крупных сделок», цены на которые растут, а отказ от них, о котором раньше нельзя было и помыслить, становится реальностью. Знание того, что около половины востребованной литературы доступно без подписки, может склонить некоторые библиотеки к отказу от нее, особенно с учетом того, что этот процент стабильно растет. Интересно, что информация об отказе Монреальского университета от подписки на пакет журналов издательства Taylor & Francis [5] сопровождалась рекомендацией пользоваться Unpaywall и другими вариантами выхода на открытые статьи. Такое решение может показаться радикальным, однако отказ от платных подписок давно входит в дорожную карту движения за открытый доступ [51]. Даже если доля ОД недостаточна для того, чтобы отказаться от подписки, ее может вполне хватить для переговоров о снижении цены на основе модели «платы за доступ с учетом ОД» (*OA-adjusted Cost Per Access*) [9]). Тем не менее необходимы дальнейшие исследования распространенности контента ОД в отдельных журналах и пакетах крупных сделок; нужно также разработать средства анализа ОД для библиотекарей-практиков в помощь принятию решений об отказе от подписок.

Это исследование имеет несколько существенных ограничений. Анализируемые данные включали только журнальные статьи с DOI. Это означает, что в нем плохо представлены дисциплины и регионы, где преобладают материалы конференций и статьи, не имеющие DOI. Наша выборка из Crossref содержит около 7% материалов, помещенных на первых страницах журналов, которые имеют DOI и тип «журнальная статья», однако чаще всего это информация о редколлегии и подбные. Бронзовая категория в нашей интерпретации включает статьи, фактически опубликованные в журналах ОД, не зарегистрированных в DOAJ; в дальнейшем необходимо идентифицировать их и отнести соответствующие статьи к Золотому ОД. Как оговорено в нашем определении ОД, мы не учитывали открытые документы из научных социальных сетей. Сервис oaDOI отчасти учитывает статьи, опубликованные на персональных веб-страницах, но значительно уступает в этом глобальным сервисам типа Google. oaDOI индексирует тысячи репозиторий, но отнюдь не все. Проведенная нами оценка точности показывает, что около 23% контента ОД, который обнаруживается поисковиками, не охвачено oaDOI,

следовательно, наши оценки преуменьшают ОД приблизительно на 30%. Наконец, наше исследование не учитывает дат размещения статей в репозиториях. Поскольку репозитории часто наполняются ретроспективно, в данной работе мы не измеряем увеличения или уменьшения Зеленого ОД от года к году; оценивается только процент Зеленых статей на момент сбора данных.

Помимо полученных эмпирических результатов, эта работа демонстрирует потенциал использования оаDOI в дальнейших исследованиях. Этот открытый сервис обеспечивает ученым возможность оценки и отслеживания динамики моделей доступа к научной литературе на больших массивах данных, а также факторов, влияющих на них. Так, наши результаты демонстрируют рост доли ОД в потоке литературы и повышенный уровень цитирования работ ОД по сравнению с закрытыми статьями. Предположительно, на это влияют несколько факторов, но они пока недостаточно изучены. Привлечение других наборов данных (из WoS, Scopus или Crossref-оаDOI) позволит оценить на больших массивах влияние различных мандатов на депонирование или отследить динамику доступности документов, чтобы, например, выяснить, когда авторы самоархивируют статьи, или определить устойчивость доступа к материалам, открытым в целях продвижения контента. Обобщенные на уровне журналов и издательских платформ, эти данные обеспечат библиотекарей индикаторами для принятия обоснованных решений об отказе от подписки и смягчения последствий таких решений. Широкомасштабное применение оаDOI позволит также более тщательно изучить эффект преимущества ОД в цитировании по отдельным дисциплинам и периодам. Как показано в [52], влияние осложняющих факторов можно ослабить за счет привлечения метаданных статей, опубликованных в тех же журнальных выпусках, на ту же тему или опубликованных автором в тот же период. Мы надеемся, что другие исследователи глубже изучат такие данные, чтобы лучше понимать распространение ОД и влияющие

на него факторы. Это чрезвычайно важно для научной коммуникации будущего.

Дополнительная информация

Вклад авторов

Хизер Пивовар*, **Джейсон Прим*** и **Стефани Хауштайн** — концепция и планирование экспериментов, выполнение экспериментов, анализ данных, обеспечение реагентов / материалов / аналитических сервисов, подготовка текста статьи, подготовка рисунков и таблиц, работа над черновиками статьи.

Винсент Ларивьер — концепция и планирование экспериментов, выполнение экспериментов, анализ данных, обеспечение реагентов / материалов / аналитических сервисов, подготовка текста статьи, работа над черновиками статьи.

Хуан Пабло Алперин — концепция и планирование экспериментов, выполнение экспериментов, анализ данных, подготовка текста статьи, работа над черновиками статьи.

Лиза Матиас — выполнение экспериментов, анализ данных, работа над черновиками статьи.

Брее Норландер — анализ данных, подготовка текста статьи, работа над черновиками статьи.

Эшли Фарли — подготовка текста статьи, работа над черновиками статьи.

Джевин Вест — работа над черновиками статьи.

* Эти авторы внесли одинаковый вклад в подготовку статьи.

Доступность данных

Zenodo: <http://doi.org/10.5281/zenodo.837902>.

Данные, анализируемые в настоящей статье, свободно доступны по следующим адресам:

- <http://dx.doi.org/10.5281/zenodo.837902> and the R statistics code can be found,
- <https://github.com/Impactstory/oadoi-paper1> — код R статистики,
- <https://github.com/impactstory/oadoi> — код оаDOI,
- <https://oadoi.org/api> — информация по оаDOI API.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Bohannon J. Who's downloading pirated papers? Everyone. *Science*. 2016;352(6285):508–512.
2. Greshake B. Looking into Pandora's Box: the content of Sci-Hub and its usage. *F1000Research*. 2017;6:Article 541.
3. Björk B. Hybrid open access—a longitudinal study. *Journal of Informetrics*. 2016;10(4):919–932.
4. Björk B-C. The open access movement at a crossroad: Are the big publishers and academic social media taking over? *Learned Publishing*. 2019;29(2):131–134. <https://doi.org/10.1002/leap.1021>

5. Université de Montréal. *UdeM Libraries cancel Big Deal subscription to 2231 periodical titles published by Taylor & Francis Group*. 2017.
6. Schiermeier Q., Mega E.R. Scientists in Germany, Peru and Taiwan to lose access to Elsevier journals. *Nature News*. 2017;541(7635):13.
7. Anderson R. *When the wolf finally arrives: big deal cancellations in North American Libraries*. The Scholarly Kitchen. URL: <https://scholarlykitchen.sspnet.org/2017/05/01/wolf-finally-arrives-big-deal-cancellations-north-american-libraries/> (accessed 9 January 2018).
8. Universität Konstanz. *Teurer als die Wissenschaft erlaubt*. 2014.
9. Antelman K. Leveraging the growth of open access in library collection decision making. *Proceeding from ACRL 2017: at the helm: leading transformation*. 2017.
10. Tennant J.P., Waldner F., Jacques D.C., Masuzzo P., Collister L.B., Hartgerink C.H. The academic, economic and societal impacts of Open Access: an evidence-based review. *F1000 Research*. 2016;5:Article 632.
11. McKiernan E., Bourne P., Brown C., Buck S., Kenall A., Lin J., McDougall D., Nosek B.A., Ram K., Soderberg C.K. How open science helps researchers succeed. *eLife*. 2016;5:e16800.
12. Creative Commons. *Attribution 4.0 International (CC BY 4.0)*. 2018.
13. Willinsky J. The nine flavours of open access scholarly publishing. *Journal of Postgraduate Medicine*. 2003;49:263–267.
14. Matsubayashi M., Kurata K., Sakai Y., Morioka T., Kato S., Morioka T., Kato S., Mine S., Ueda S. Status of open access in the biomedical field in 2005. *Journal of the Medical Library Association*. 2009;97(1):4–11.
15. Chen X., Olijhoek T. Measuring the degrees of openness of scholarly journals with the open access spectrum (OAS) evaluation tool. *Serials Review*. 2016;42(2):108–115.
16. Suber P. Gratis and libre open access. *SPARC Open Access Newsletter*. 2008;124.
17. Archambault É., Amyot D., Deschamps P., Nicol A., Provencher F., Rebout L., Roberge G. *Proportion of open access peer-reviewed papers at the European and world levels—1996–2013*. Brussels: European Commission, 2014.
18. Gargouri Y., Larivière V., Gingras Y., Carr L., Harnad S. *Green and gold open access percentages and growth, by discipline*. [Preprint]. 2012.
19. Harnad S., Brody T., Vallières F., Carr L., Hitchcock S., Gingras Y., Oppenheim C., Hajjem C., Hilf E.R. The access/impact problem and the green and gold roads to open access: an update. *Serials Review*. 2008;34(1):36–40.
20. Walker T.J., Soichi T. Free internet access to traditional journal. *Journal of Information Processing and Management*. 1998;41(9):678–694.
21. Laakso M., Björk B.C. Delayed open access: an overlooked high-impact category of openly available scientific literature. *Journal of the American Society for Information Science and Technology*. 2013;64(7):1323–1329.
22. Willinsky J. *The access principle: the case for open access to research and scholarship*. 1st edition ed. Cambridge: MIT Press, 2009.
23. Archambault É., Amyot D., Deschamps P., Nicol A., Provencher F., Rebout L., Roberge G. *Proportion of open access peer-reviewed papers at the European and world levels—2004–2011*. Brussels: European Commission, 2013.
24. Jamali H.R. Copyright compliance and infringement in ResearchGate full-text journal articles. *Scientometrics*. 2017;112(1):241–254.
25. Chawla D. Publishers take ResearchGate to court, alleging massive copyright infringement. *Science News*. 2017.
26. Fortney K., Gonder J. *A social networking site is not an open access repository*. 2015. URL: <http://osc.universityofcalifornia.edu/2015/12/a-social-networking-site-is-not-an-open-access-repository>
27. Björk B.-C. Gold, green, and black open access. *Learned Publishing*. 2017;30(2):173–175. <https://doi.org/10.1002/leap.1096>
28. Björk B.C., Welling P., Laakso M., Majlender P., Hedlund T., Guðnason G. Open access to the scientific journal literature: situation 2009. *PLoS One*. 2010;5(6):e11273.
29. Laakso M., Welling P., Bukvova H., Nyman L., Björk B.C., Hedlund T. The development of open access journal publishing from 1993 to 2009. *PLoS One*. 2011;6(6):e20961.
30. Laakso M., Björk B.C. Anatomy of open access publishing: a study of longitudinal development and internal structure. *BMC Medicine*. 2012;10(1):124.
31. Chen X. Journal article retrieval in an age of Open Access: how journal indexes indicate Open Access articles. *Journal of Web Librarianship*. 2013;7(3):243–254.
32. Hajjem C., Harnad S., Gingras Y. *Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact*. [Preprint]. 2006.
33. SPARC Europe. *The open access citation advantage: list of studies until 2015*. 2015.

34. Wagner A.B. Open access citation advantage: an annotated bibliography. *Issues in Science and Technology Librarianship*. 2010;60:2.
35. Tennant J. *The open access citation advantage*. 2017. URL: <https://www.scienceopen.com/collec-tion/996823e0-8104-4490-b26a-f2f733f810fb>
36. Davis P.M., Walters W.H. The impact of free access to the scientific literature: a review of recent research. *Journal of the Medical Library Association*. 2011;99(3):208–217.
37. Davis P.M. Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *FASEB Journal*. 2011;25(7):2129–2134.
38. McCabe M., Snyder C. Identifying the effect of open access on citations using a panel of science journals. *Economic Inquiry*. 2014;52(4):1284–1300.
39. Ottaviani J. The post-embargo open access citation advantage: it exists (probably), it's modest (usually), and the rich get richer (of course). *PLoS One*. 2016;11(8):e0159614.
40. Packer A.L. The SciELO open access: a gold way from the south. *Canadian Journal of Higher Education*. 2010;39(3):111–126.
41. Himmelstein D.S., Romero A.R., McLaughlin S.R., Tzovaras B.G., Greene C.S. Sci-Hub provides access to nearly all scholarly literature. *PeerJ Preprints*. 2017 (No. e3100v1).
42. Gorraiz J., Melero-Fuentes D., Gumpenbergera C., Valderrama-Zuriánc J.-C. Availability of digital object identifiers (DOIs) in Web of Science and Scopus. *Journal of Informetrics*. 2016;10(1):98–109.
43. Boudry C., Chartron G. Availability of digital object identifiers in publications archived by PubMed. *Scientometrics*. 2017;110(3):1453–1469.
44. Mongeon P., Paul-Hus A. The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*. 2016;106(1):213–228.
45. Björk B.C., Laakso M., Welling P., Paetau P. Anatomy of green open access. *Journal of the Association for Information Science and Technology*. 2014;65(2):237–250.
46. Archambault É., Côté G., Struck B., Voorons M. *Research impact of paywalled versus open access papers*. 2016.
47. Smith E., Haustein S., Mongeon P., Fei S., Ridde V., Larivière V. Knowledge sharing in global health research; the impact, uptake and cost of open access to scholarly literature. *BMC Health Research Policy and System*. [In Press].
48. Craig I.D., Plume A.M., McVeigh M.E., Pringle J., Amin M. Do open access articles have greater citation impact? *Journal of Informetrics*. 2007;1(3):239–248.
49. Berg J. Measuring the scientific output and impact of NIGMS grants [Blog post]. *NIGMS Feedback Loop Blog*. URL: <https://loop.nigms.nih.gov/2010/09/measuring-the-scientific-output-and-impact-of-nigms-grants/>
50. PLoS. *Reviewer guidelines: criteria for publication*. 2018.
51. Anderson R. *The forbidden forecast: thinking about open access and library subscriptions*. The Scholarly Kitchen. URL: <https://scholarlykitchen.sspnet.org/2017/02/21/forbidden-forecast-thinking-open-access-library-subscriptions/> (accessed 15 July 2017).
52. Gargouri Y., Hajjem C., Larivière V., Gingras Y., Carr L., Brody T., Harnad S. Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS One*. 2010;5(10):e13636.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Хизер Пивовар*, ИмпактСтори (Impactstory);
heather@impactstory.org

Джейсон Прим*, ИмпактСтори (Impactstory);
jason@impactstory.org

Винсент Ларивьер, Школа библиотечных и информационных наук Университета Монреаля; Научно-техническая обсерватория Межвузовского научно-исследовательского центра науки и техники Университета Квебека в Монреале

Хуан Пабло Алперин, Канадский институт исследований в издательском деле Университета Саймона Фрезера; Проект «Общественные знания» (Public Knowledge Project)

Лиза Маттиас, Лаборатория научной коммуникации Университета Саймона Фрезера

Брее Норландер, Информационная школа Вашингтонского университета; ФлоришОА (FlourishOA)

Эшли Фарли, Информационная школа Вашингтонского университета; ФлоришОА (FlourishOA)

Джевин Вест, Информационная школа Вашингтонского университета

Стефани Хауштайн, Научно-техническая обсерватория Межвузовского научно-исследовательского центра науки и техники Университета Квебека в Монреале; Школа информационных исследований Университета Оттавы

Heather Piwovar*, Impactstory;
heather@impactstory.org

Jason Priem*, Impactstory;
jason@impactstory.org

Vincent Larivière, École de bibliothéconomie et des sciences de l'information, Université de Montréal; Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal

Juan Pablo Alperin, Canadian Institute for Studies in Publishing, Simon Fraser University; Public Knowledge Project

Lisa Matthias, Scholarly Communications Lab, Simon Fraser University

Bree Norlander, Information School, University of Washington; FlourishOA

Ashley Farley, Information School, University of Washington; FlourishOA

Jevin West, Information School, University of Washington

Stefanie Haustein, Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal; School of Information Studies, University of Ottawa

Перевод **Н. Н. Литвиновой**