Only two out of five articles by New Zealand researchers are free-to-access: a multiple API study of access, its impact on open citation advantage, cost of Article Processing Charges (APC), and the potential to increase the proportion of open access

Richard Kenneth Alistair White[1], Anton Angelo[2], Deborah Jane Fitchett[3], Moira Fraser[4], Luqman Hayes[5], Jess Howie[6], Emma Richardson[7], Bruce Duncan White[8]

[1] Vice-Chancellor's Office, University of Otago, Dunedin, Otago, Aotearoa New Zealand
[2] Library, University of Canterbury, Christchurch, Canterbury, Aotearoa New Zealand
[3] Learning, Teaching & Library, Lincoln University, Lincoln, Canterbury, Aotearoa New Zealand
[4] Library, University of Canterbury, Christchurch, Canterbury, Aotearoa New Zealand
[5] Library, Auckland University of Technology, Auckland, Auckland, Aotearoa New Zealand
[6] Library, University of Waikato, Hamilton, Waikato, Aotearoa New Zealand
[7] Libraries & Learning Services, University of Auckland, Auckland, Auckland, Aotearoa New Zealand
[8] Office of the Provost, Massey University, Palmerston North, Manawatu, Aotearoa New Zealand

Corresponding Author:
Richard White[1]
362 Leith Street, Dunedin, Otago, 9010, New Zealand
Email address: richard.white@otago.ac.nz

Abstract

We studied journal articles published by researchers at all eight of New Zealand universities in 2017 to determine how many were freely accessible on the web. We wrote software code to harvest data from multiple sources, code that we now share to enable others to reproduce our work on their own sample set. In May 2019, we ran our code to determine which of the 2017 articles were open at that time and by what method; where those articles would have incurred an Article Processing Charge (APC) we calculated the cost if those charges had been paid. Where articles were not freely available we determined whether the policies of publishers in each case would have allowed deposit in a non-commercial repository (Green open access). We also examined average citation rates for different types of access. We found that, of our 2017 sample set, about two out of every five articles were freely accessible without payment or subscription (41%). Where research was explicitly said to be funded by New Zealand's major research funding agencies, the proportion was slightly higher at 49%. Where open articles would have

incurred an APC we estimated an average cost per article of USD1,682 (for publications where all articles require an APC, that is, Gold open access) and USD2,558 (where APC payment is optional, Hybrid open access) at a total estimated cost of USD1.45m. Of the paid options, Gold is by far more common for New Zealand researchers (82% Gold, 18% Hybrid). Where articles were not freely accessible we found that a very large majority of them (88%) could have been legally deposited in an institutional repository. In terms of average citation rates, we found Green and Hybrid open access to achieve the highest rates, higher than other forms of open access and higher still than research that is only available via payment. Given that most New Zealand researchers support research being open, there is clearly a large gap between belief and practice in New Zealand's research ecosystem, despite a clear citation advantage for open access over research that is not freely accessible.

Introduction

Researchers seek to change the world and writers seek to be read, but for many years a dysfunctional scholarly publishing system has walled off most published research findings from the majority of its potential readership. Since the transition from print to electronic publishing began in the late 1990s, various initiatives explored the potential for this digital transformation to make research literature more accessible to the public. University libraries, concerned about continuing growth in journal subscription costs, hoped an open access system would provide a more affordable alternative. At the same time they sought to advance the mission of their host institutions to create social capital through the promulgation of quality peer-reviewed research.

Three major developments in the early 2000s set the scene for the current open access environment:

- The growth of "Gold open access" journals, funded by Article Processing Charges (APCs) rather than by subscriptions.
- The adoption of "Hybrid open access" options by subscription journals, making individual papers openly accessible through the payment of APCs.
- The development of institutional repositories, providing an alternative route of "Green open access" to individual papers without publication charges.

Since then there has been considerable interest in the potential of open access to contribute to universities' own goals as a result of supporting broader society to access research outputs. This includes a growing understanding that, as a result of their enhanced availability, openly accessible papers are likely to be cited at a higher rate than those behind paywalls.

Unfortunately, open access has not produced the anticipated reduction in costs. Subscription costs of research journals continue to rise while APCs for Gold and Hybrid journals add another cost to

university budgets (Guédon et al. 2019). Furthermore, whereas subscription costs were centralised within library budgets, APC charges are paid from a variety of sources, including departmental budgets and external research funds, which makes them less visible and harder to manage (Monaghan et al. 2020). Moreover libraries have had limited success in encouraging researchers to deposit copies of their work in institutional repositories. In New Zealand, this is despite all universities having an institutional repository.

New Zealand has no specific guidance from government or major research funding agencies on open access publishing or centralised support to pay APCs. While government has established an open access framework that applies to government agencies, this does not extend to the university sector (New Zealand Government 2014). A recent government consultation document on research strategy raised the possibility that a co-ordinated approach in the research sector could be of benefit.[1] To date, New Zealand's major funding agencies have not enforced requirements on research projects they fund to release research outputs or data with open licences. Some New Zealand universities have adopted open access policies and/or guidelines following the Green open access pathway (Wikipedia contributors 2020), as there is no government funding to enable Gold open access. Three universities operate small APC funds to support researchers to publish open access when this will achieve certain goals.

In 2019 the Council for New Zealand University Libraries (CONZUL) established a project team with representatives from seven of the eight New Zealand universities to research the current environment of open access in New Zealand. A major stream of this project sought to establish how much of our universities' research outputs were open access. While other tools provide a figure for the proportion of research that is open, we wanted to extend our understanding to determine:

- how much our researchers might be spending on Article Processing Charges (APCs) on top of what libraries pay in subscriptions;
- how much of our work could be freely accessible via self-archiving but is not; and
- the impact openness has on citations or other measures of impact.

This paper reports on the findings of this project and makes our method and software code available to others to create their own sets of data and their own analyses.

This paper focuses on one element of the CONZUL Open Access Project. The wider project produced a full report (Fraser et al. 2019) examining the wider open access environment in New Zealand and an infographic designed to communicate its findings in a readily digestible format.

Literature Review

As the prevalence of open access publication of research results has increased over the years (Abediyarandi & Mayr 2019; Archambault et al. 2014; Archambault et al. 2013; Gargouri et al. 2010; Laakso et al. 2011; Maddi 2019; Martín-Martín et al. 2018; Piwowar et al. 2018; Wang et al. 2018), so too has the ability to gain insight into its nature and development. However, this has occurred alongside increasing complexity in the way open access levels are measured, and the resulting literature is methodologically diverse. As such, this literature review presents a brief overview of the main methodological approaches and relevant results.

Perhaps the most influential study to date was carried out by Piwowar et al. (2018). In their review of the literature, they note the paucity of studies between 2014 and the time of writing. As more automated research on open access becomes possible through Application Programming Interfaces (APIs) and enhanced indexing, sample sizes have increased (Piwowar et al. 2018). How open access development is measured depends on a number of factors, including scope and source of data. Different methods are used according to the aims of the research. Some studies focus on a given country (Abediyarandi & Mayr 2019; Bosman & Kramer 2019; Holmberg et al. 2019; Mikki et al. 2018; Piryani et al. 2019; Pölönen et al. 2019; Sivertsen et al. 2019) open access type (Wang et al. 2018) or funder (Kirkman 2018). Others aim for a global overview (Archambault et al. 2014; Laakso et al. 2011; Martín-Martín et al. 2018; Piwowar et al. 2018; Robinson-Garcia et al. 2019; Wang et al. 2018).

Because of this diversity, it is difficult to draw comparisons between results. Most recent studies point to an overall open access rate of between 45 and 55% (Bosman & Kramer 2019; Martín-Martín et al. 2018; Piwowar et al. 2018; Pölönen et al. 2019). This is significant because an open access rate of 50% is posited as a “tipping point” by some (Archambault et al. 2013). Where open access rate is calculated as total of the scholarly record or over an extended period, this figure drops dramatically – Piwowar et al. (2018) estimate the total percent of scholarly record at 28%, Maddi (2019) at 31%. Piwowar et al. (2019) hypothesize a rise to 44% by 2025. Where wider sources are included, such as Academic Social Networks (ASNs) or Google Scholar the open access rate rises (Martín-Martín et al. 2018; Nazim & Zia 2019). Unsurprisingly, older studies report lower rates (Bjork et al. 2010; Gargouri et al. 2010), reinforcing the findings of many scholars that open access is on the rise (Abediyarandi & Mayr 2019; Archambault et al. 2014; Archambault et al. 2013; Gargouri et al. 2010; Laakso et al. 2011; Maddi 2019; Martín-Martín et al. 2018; Piwowar et al. 2018; Wang et al. 2018). The growth of Gold open access in particular has been noted (Archambault et al. 2014; Martín-Martín et al. 2018; Piryani et al. 2019; Pölönen et al. 2019). Despite this, the literature clearly shows that open access development varies by discipline (Bjork et al. 2010; Bosman & Kramer 2019; Maddi 2019; Martín-Martín et al. 2018; Piryani et al. 2019; Sivertsen et al. 2019) and country (Archambault et al. 2014; Maddi 2019;

Martín-Martín et al. 2018; Robinson-Garcia et al. 2019; Sivertsen et al. 2019; Torres-Salinas et al. 2019; Wang et al. 2018).

It is important to note that all of these results can only be viewed as accurate at a given point of time. Archambault et al. (2014) point out that open access is fluid in nature – closed articles can become open after embargoes, repositories can be backfilled and publications may be open access but only at the discretion of the publisher or on limited terms. This is manifest in the fact that Piwowar et al. (2018) introduced the Bronze category, defined as "made free-to-read on the publisher website, without an explicitly open license". They noted that "It is also not clear if Bronze articles are temporarily or permanently available to read for free"(p.6). In fact, there is limited consensus on the definition of all open access types and even on the line between open access and non-open access. Similarly, the results of any study are also closely linked to its scope and source data – many studies use Web of Science or Scopus, which are known to under-represent certain areas of literature (Martín-Martín et al. 2018). As automation becomes more central to open access research, results become limited to those publications with a DOI, further excluding some important categories of research. Robinson-Garcia et al. (2019) argued that sources of open access status such as Unpaywall also need to be better understood in order to fairly represent open access rates.

The funding of open access through article processing charges (APCs) is another matter of high concern, although there is limited consensus in the literature around how these costs are to be estimated. A journal is classified as Gold if all articles are immediately open and APCs for these titles are recorded in the Directory of Open Access Journals (DOAJ), while for Hybrid journals the articles are paywalled unless an APC is charged. One method of estimating the cost of APCs to institutions is by examining financial records (Jahn & Tullney 2016; Pinfield et al. 2017; Solomon & Björk 2016) which aims to capture the actual amounts paid or by reviewing institutional agreements with publishers (Lovén 2019). The other main approach is capturing the advertised prices from DOAJ or publisher websites (Björk & Solomon 2015; Matthias 2018; Morrison et al. 2016; Solomon & Björk 2016).

Average citation rates is another topic that has been hotly debated in the literature. Research almost always finds a positive correlation between open access and citation rate (Archambault et al. 2014; Copiello 2019; McCabe & Snyder 2014; Mikki et al. 2018; Ottaviani 2016; Piwowar et al. 2018; Piwowar et al. 2019; Wang et al. 2015). However confounding factors cast considerable uncertainty over direct causation (Gaulé & Maystre 2011; Torres-Salinas et al. 2019). It is also clear that citation advantage is not distributed evenly across all disciplines (Holmberg et al. 2019) or types of open access (Mikki et al. 2018; Piwowar et al. 2018). In fact, several studies have found a citation *dis*advantage for Gold open access (Archambault et al. 2014; Archambault et al. 2013; Piwowar et al. 2018; Torres-Salinas et al. 2019). The way citation advantage (or lack thereof) is measured can have a considerable influence on results, leading some scholars to use normalised figures such as Category Normalised Citation Impact (Torres-Salinas et al. 2019) or

Average Relative Citation (Archambault et al. 2016) rather than total citations. Others argue that quality bias from self-selection (i.e., researchers select open access for higher quality work) inflate the apparent citation advantage (Torres-Salinas et al. 2019). However, some authors have found that a citation advantage exists despite confounding factors, albeit at a lower rate (Gargouri et al. 2010; McCabe & Snyder 2014; Ottaviani 2016). Attention, as measured by views, downloads and altmetrics, are similarly positively affected (Adie 2014; Holmberg et al. 2019; Wang et al. 2015) and Wang et al. (2015) found that downloads for open access publications were sustained for longer periods of time than non-open access. On balance, the literature largely confirms the open access citation advantage but the magnitude and reasons for this remain unclear.

Materials & Methods

The CONZUL project team developed software that used Digital Object Identifiers (DOIs) to establish publications' open access status, APC price, and ability to be self-archived.

Our work depended on many open API services, the most integral being Unpaywall. As such our definition of 'open' in this study largely aligns with that of Unpaywall, including Bronze as initially proposed by Piwowar et al. (2018). Thus the openness of an article in our study is defined very broadly: "OA articles are free to read online, either on the publisher website or in an OA repository." Unpaywall does "not harvest from sources of dubious legality like ResearchGate or Sci-Hub" (Unpaywall) Table 1 shows the categories we used and an associated definition.

Table 1 Open Access Type Definition

Type of Open Access (OA)	**Definition**
Gold OA	Published version is immediate OA. APC charged.
Hybrid OA	Publication is subscription-based. APC can be paid to make individual articles OA.
Bronze OA	Currently free to read on publisher's site but licence not clear.
Green OA	Freely accessible in a reputable repository (i.e. academic social networking sites are not included). Publisher's version is paywalled.
Closed	Published version is paywalled.
Diamond OA	Published version is free. No APC charged.

Unpaywall uses a hierarchy to determine a single status for each paper. Priority is given to those statuses which imply immutability, specifically through publication in a Gold journal or through the payment of an APC in a Hybrid journal. For Gold journals no distinction is made between those that charge APCs and those that do not. For the purposes of our study, however, where the Directory of Open Access Journals (DOAJ) showed a Gold journal does not charge APCs, we re-categorised these as Diamond. As already noted, Unpaywall introduced the Bronze status for papers openly available from the publishers but without an explicit license. Perhaps unfortunately, given questions around the persistence of Bronze open access, this status was given a higher priority than Green, which was reserved for papers openly accessible from repositories rather than from publishers. The status Closed is defined as papers that are not openly available in any form.

Overlaying the access dimension is the question of authorship. The number of authors of a published research article can range from one (sole authorship) to several thousand (project participation). Multiple authorship is a significant issue when we attempt to link published research to institutions and countries, particularly when there are no established norms for allocating divisions of responsibility. Where there are, say, 200 authors in a research group the fact that one of them is employed at University A tells us very little about the behaviour and performance of that institution, although a productive project may end up crediting it with numerous publications on the basis of participation by this single team member. This may be an insoluble problem for affiliation-based bibliometric research but in a project like the present one it is advisable not to ignore it. One means of creating a "strong link" between a paper and an institution is through the “corresponding author" who takes overall responsibility for the publication process. While this is often the first-named author, this is not universal.

For our purposes, we limited our sample set to journal articles with a Digital Object Identifier (DOI) published in 2017 that included at least one author affiliated with a New Zealand university. This provided a comprehensive dataset representing a large proportion of the research outputs of all eight universities in the country. Although we were carrying out the work in 2019 we chose to use 2017 as our sample set because, firstly, the research outputs were more likely to have passed the date for embargo set by publishers for self-archiving (one of our key interests) and, secondly, citation counts would be more mature than for more recent research.

DOIs for 2017 journal articles were gathered from each university, then amalgamated into a single file of more than 12,600 journal articles. If there was a local corresponding author at any university for a given article then it was designated as having a New Zealand corresponding author. During the course of the project we found that a small percentage of articles with large numbers of authors and large numbers of citations skewed the data so articles with more than 20 authors were excluded on the grounds that they had a tenuous connection to the New Zealand University that had submitted the DOI. This reduced the sample size to 12,016. These were fed into The Program.

The Program

At the heart of our work was the 'Program', written in Python. One of our primary aims in publishing this paper is to share the code for the Program for others to use as well as detailing the results of our own work. A set of DOIs can be submitted to the Program, which uses a number of APIs to produce a set of results, whether for a single department, an institution, a discipline, a country (as in our case) or any other parameter.

DOIs are fed into the Program, written in Python, using a Comma Separated Value file (.CSV) and for each article the following information was obtained as shown in Table 2:

Table 2 Information gathered by the Program and corresponding data sources

Information	***Source***
Metadata: author(s), title, journal, ISSN, etc.	Unpaywall API
Open or closed	Unpaywall API
Type of access (Gold, Hybrid, Green, Closed, etc.)	Unpaywall API
Reprint/corresponding author	Web of Science and Scopus (CSV file)
Funders	Web of Science and Scopus (CSV file)
URLs of all repository versions	Unpaywall API
Type of 'best open version': published, preprint, postprint	Unpaywall API
APC/No APC	DOAJ (CSV file) GitHub site of Lisa Matthias (Freie Universität Berlin)[2]
Journal and publisher archiving rules and embargo periods	Sherpa/Romeo API
Citations	Crossref API

Errors are reported by the Program where information could not be found, for example if a DOI was not recognised.

[2] github.com/lmatthia/publisher-oa-portfolios

DOIs were obtained from the research management systems of the individual universities, with one exception where Scopus was used as the source. This meant we were able to go beyond the limitations set by the use of proprietary databases which contain only a proportion of any institution's research publications. It was then possible to "chain" the data gathering. For example, Unpaywall provided ISSNs which were then submitted via API requests to Sherpa/Romeo to capture data on publisher allowances for the use of publications in institutional repositories. ISSNs were also used to capture data on APCs for individual journals.

However, not all the data used by the Program was accessible through APIs. Crossref was an excellent source of information for authors, even when these numbered in the thousands, but provides very limited data on author affiliations. On the other hand, Web of Science and Scopus provide detailed author-affiliation data, including identifying corresponding authors, but this needed to be output manually as CSV files for subsequent access by the program. A similar process was followed with APC data.

While this paper is focused on the national picture for New Zealand, for those who may be interested in utilising our code on their own DOIs we note that author affiliation data is included in the output. Therefore results can also be broken down to analyse subsets at the level of individual institutions.

Results

The Program was run on 30 May 2019. The output was analysed and the following information extracted:

- the overall percentage of open and closed papers both for all authors and the subset of New Zealand corresponding authors;
- the total percentage of papers in each of the access categories: Closed, Gold, Hybrid, Bronze, Green, Diamond (note that the “best version” is reported so there was no overlap between categories.);
- the total percentage of papers available through repositories (note that because an article can be published and in a repository there is some overlap with the other categories);
- the total percentage of open and closed papers funded by major New Zealand agencies;
- the total cost for Gold and Hybrid papers if all APCs had been charged as advertised;
- the total cost of APCs as advertised if they had been paid on papers available in repositories;
- the total number of closed papers that could be made open as Author Accepted Manuscripts (AAM) as deduced from allowances recorded in Sherpa/Romeo;
- the total cost of APCs as advertised if these papers were made open in Hybrid mode.

Overall proportion of open v closed articles

Overall 59% of all the articles in our sample set were only available behind a subscription paywall (see Table 3). This result also shows a higher average citation rate for the 41% of articles that were openly accessible of around +30%.

Table 3 Proportion of articles with at least one New Zealand author which were open

Availability of article	Count	%	Avg citations (Crossref)
Closed	7056	59%	4.53
Open	4960	41%	5.91
Total	**12016**	**100%**	**5.10**

When we performed the same analysis of those articles where the corresponding author was affiliated with a New Zealand university (as opposed to *any* of the authors being from a New Zealand university) we found the proportion of open articles was significantly less (see Table 4).

Table 4 Proportion of articles with New Zealand corresponding author which were open

Availability of article	Count	%	Avg citations (Crossref)
Closed	3502	66%	3.69
Open	1799	34%	4.88
Total	**5301**	**100%**	**4.09**

The proportion of open here reduced to 34%, meaning only 1 in 3 articles from 2017 where the corresponding author was affiliated with a New Zealand university was freely accessible. The average citation rates for open and closed remain almost the same (+32% for open) as for the larger sample set of all authors.

Articles by type of access

As seen in Table 5, Gold open access comprises the largest proportion of open articles in our sample (35% of open articles and 14% of all articles). Green is next (26% of open articles or 10% of all articles), closely followed by Bronze. Hybrid is a significantly lower percentage with only 13% of open articles or 5% of all articles. Diamond clearly is not commonly used by New Zealand researchers.

Table 5 Articles by type of access

Type of access	Count	%	Avg citations (Crossref)
Bronze	1101	9%	5.11
Closed	7056	59%	4.53
Diamond	265	2%	1.79
Gold	1706	14%	5.14
Green	1259	10%	7.52
Hybrid	629	5%	7.94
Total	**12016**	**100%**	**5.10**

The average citation rates for the different types can be used as a measure of impact. Many factors influence the number of citations an article receives but, over a large sample set, we would expect this to even out and our research was focused on the widest possible view of New Zealand's open access environment. The average citation rates for Hybrid and Green rates come out well above the others, at almost the same level (7.94 and 7.52 respectively). All types of open access except for Diamond have a higher average citation rate than Closed articles.

Again we analysed the subset of articles where a New Zealand university researcher was the corresponding author for the article (see Table 6).

Table 6 Articles by type of access for New Zealand corresponding authors

Type of access	Count	%	Avg citations (Crossref)
Bronze	423	8%	4.93
Closed	3502	66%	3.69
Diamond	95	2%	1.45
Gold	697	13%	4.74
Green	432	8%	5.13
Hybrid	152	3%	6.84
Total	**5301**	**100%**	**4.09**

The pattern is broadly similar to the dataset for all authors (as seen in Table 5). Gold is most common (39% of the open subset and 13% of all articles); Green and Bronze are near-equal at 24% and 23% respectively of the open articles and 8% of all articles; Hybrid is somewhat lower here, with 8% of open and just 3% of the total. Diamond is constant at 2% of all articles.

Similarly, open articles for New Zealand corresponding authors have higher average citation rates than Closed articles with the exception of Diamond. Here, however, Hybrid has a much higher

average citation rate than Green. As with the larger sample set, all types of open have a higher average citation rate than Closed with the exception of Diamond.

Gold and Hybrid costs

We extrapolated the total number of articles that might have incurred an APC by adding together Hybrid and Gold figures. We see that 697 Gold articles and 152 Hybrid ones were published in 2017 in our local author subset (849 in total). The Program included a calculation of APCs for each article, where this was known via publicly-available data sources. This was calculated only for New Zealand university-affiliated authors on the basis that the corresponding author is the most likely to be responsible for paying an APC.

Thus Table 7 shows the average APC costs, US$2558 for Hybrid and US$1682 for Gold. Hybrid also has a higher average citation rate. In other words, the average cost for a Hybrid article is 52% more while achieving an average citation rate that is 44% higher than that for Gold.

Table 7 Gold and Hybrid articles: estimated APC costs in US dollars

Type of paid OA	**Count**	**%**	**Avg citations (Crossref)**	**Known APCs**	**Known APC cost**	**Known APC avg**
Gold	697	82%	4.74	697	$1,172,029	$1,682
Hybrid	152	18%	6.84	110	$281,378	$2,558
Total	**849**	**100%**	**5.11**	**807**	**$1,453,407**	**$1,801**

We were also able to estimate the total APCs paid. Most publishers provide information on publishing charges and this data has been collected by Lisa Matthias of the Freie Universität Berlin (Matthias 2018). The ‘Known APC cost’ is a notional amount because:

- it is not possible to know where APCs may have been waived or whether they were paid from research funding, institutional funds, researchers’ own money or another source; and
- this information is not available for all journals.

The APC costs in our tables are effectively a total of the ‘list price’ for each article based on APC information that is publicly-available. Accordingly, the total amount for both categories was US$1.45 million at 2017 prices.

Embargo periods and self-archiving

Sherpa/Romeo data let us examine which of the closed articles could be self-archived according to publishers’ policies. Table 8 shows, for all New Zealand-affiliated authors, when a closed article may be deposited in an institutional repository after an embargo set by the publisher.

Table 8 Articles that could be archived in a repository in line with publishers' policies

Embargo period	**Publisher policy allows accepted manuscript in repository**	**%**
Immediate self-archive	580	17%
3 months	3	0%
4 months	1	0%
6 months	73	2%
12 months	2115	60%
18 months	318	9%
Total archivable by mid-2019	**3090**	**88%**
24 months or more	213	6%
Not self-archivable	199	6%
Total closed articles	**3502**	**100%**

We ran the Program in mid-2019, meaning any embargo period of 18 months or less would have expired. 3090 articles could have been archived but were closed, representing 88% of all the closed articles (n=3502) in our sample set. A further 213 articles have an embargo period of two years or more. It is worth noting that 12 months is by far the most common length of embargo period but also that for almost one-fifth there is no embargo.

As a result we were also able to estimate a 'theoretical' cost of APCs under the Hybrid option for papers that could have been made open as accepted manuscripts. The total comes to just under US$8 million.

Also of interest is that 114 of the 3090 articles that could have been deposited in a repository (3.7%) the publisher allowed the published version to be used, as opposed to the accepted manuscript.

Articles funded by New Zealand's major funding agencies

Funder information from Web of Science and Scopus enabled us to determine how much research funded by our major funding agencies is openly available. As indicated in our section on the context for the study, there has been no attempt by the government or major funding agencies to adopt a co-ordinated approach to open access in universities or to provide dedicated funds to support the payment of APCs.

Table 9 Proportion of open articles funded by New Zealand's major funding agencies

Funders (sorted by number of articles)	No.	closed	% closed	% open	Bronze	Gold	Diamond	Hybrid	Green
Health Research Council of New Zealand	281	128	46%	54%	38	82	2	9	22
Royal Society of New Zealand	221	120	54%	46%	20	40	2	13	26
Ministry of Business Innovation and Employment	204	137	67%	33%	17	29	2	6	13
Medical Research Foundation	100	49	49%	51%	19	25	1	1	5
Auckland Medical Research Foundation	59	28	47%	53%	11	15	1	1	3
Rutherford Discovery Fellowship	58	28	48%	52%	7	9	0	8	6
Heart Foundation	52	26	50%	50%	10	6	0	4	6
Ministry of Health	50	22	44%	56%	5	15	1	4	3
Medical Research Council	47	14	30%	70%	6	14	1	4	8
Total	**1072**	**552**	**51%**	**49%**	**133**	**235**	**10**	**50**	**92**

Overall, slightly over half (51%) of articles in our 2017 sample that were funded by our largest research funders are behind a paywall – that is, this research is inaccessible to the government agencies that funded it as well as to the New Zealand public. Therefore this subset of articles was more likely to be open than the total sample (51% open against 41% open). The Ministry for Business, Innovation and Employment has the lowest rate of open research at 33%; the Medical Research Council has the highest rate of open research at 70%.

We can also see in Table 9 how the funded articles that are freely accessible have been made open. Gold was by far the most common means of making a work open (235 of the 520 open articles or 45%); Hybrid accounted for just 50 articles (or 10% of all open). Of the funded research, Green made up 18% of the open works (92 articles) and Diamond just 2% (10 articles). Bronze means, by definition, that the permanence of the remaining open works (133 articles, 10%) is uncertain. Looking more closely at the figures for Gold and Hybrid, 55% of freely accessible research funded by these agencies theoretically incurred a fee. We calculated this to be US$529,000 if the 'list price' was paid in each instance. If we look at Gold and Hybrid as a proportion of all the articles in Table 9 (i.e. open and closed), it comes to 27% (285 Gold and Hybrid articles out of 1072 total). This compares to 19% of our total sample being made open by those means, meaning where work was specifically funded by one of these agencies an APC was more likely to have been paid.

Discussion

We found that three out of five articles with an author from a New Zealand university were only available by paying for access (59%). This figure increases to nearly two-thirds of all articles being closed when the corresponding author is a New Zealand university researcher (66%).

For validation of our results we looked at the Leiden ranking measure for openness. The Leiden Ranking (Centre for Science and Technology Studies 2018) uses a different method to ours, including using data from 2014-17 and including only 5 of the 8 New Zealand universities, but produces a similar result (see Table 10).

Table 10 Leiden ranking proportion of open articles by country, 2014-2017

University	**# Papers**	**# OA papers**	**Percentage OA**
New Zealand	29,091	11,266	38.7%
Canada	281,304	117,247	41.7%
Australia	273,486	113,789	41.6%
UK	454,802	322,827	71.0%
US	1,876,219	1,013,502	54.0%
Norway	42,608	23,109	54.2%
Ireland	26,548	12,966	48.8%
Germany	397,439	190,543	47.9%

We also used the Leiden Ranking tool to measure New Zealand's proportion of open articles against a selection of other countries. We clearly see that New Zealand's proportion of research that is openly available is below that of all the others in this selection, nearly half the figure of the highest-ranked nation, the United Kingdom.

Results of the Program also showed that average citation rates for open, with the exception of Diamond, were higher than closed. Perhaps most significantly, Green achieves an average citation rate comparable to Hybrid and higher than Gold. While some may believe that an accepted manuscript is 'less valid' than the published version of record, this certainly does not manifest itself in terms of citations by academic peers.

A huge proportion (88%) of the closed articles could be self-archived in line with publishers' policies and thereby made open. Our findings suggest that New Zealand researchers do not self-archive as often as researchers elsewhere and/or that the systems for ensuring work is archived are not effective. In any case, our researchers are missing out on the potential citation advantage conferred by Green open access. This is despite the fact that 87% of New Zealand researchers

believe that, at a policy level, publicly-funded research should be free to access (Ithaka S+R 2018). Our work identifies a clear gap between belief and practice.

When it comes to paid open access (Gold & Hybrid articles), New Zealand researchers are far more likely to use the Gold route (82% of paid open access articles were Gold). One reason for this may be the higher average APC for Hybrid, which may be seen by researchers as a luxury and opted for when publishing work in a prestigious journal that will garner interest within the discipline and/or from the public. This would require further analysis that was beyond the scope of the present project.

For our methodology, using the Unpaywall categorisation of openness means Bronze articles pose something of a quandary. Bronze was introduced by Unpaywall to be able to include papers openly accessible at a given point in time, but lacking definitive licensing information. With our Program this meant, however, that later iterations using the same DOIs (not reported on in the present paper) revealed that many papers categorised as Bronze in May 2019 had reverted to Closed or had switched to Green. The Unpaywall hierarchy places Bronze above Green, since it is the published version, but there is no way of knowing which papers will become Closed if publisher paywall restrictions are reimposed and which will continue to remain accessible through repositories. Fortunately, because Unpaywall provides repository locations in addition to the primary status it is possible to identify these Bronze/Green articles, which, for our sample, constituted 29% of all Bronze Papers.

Also of note, articles that listed a major funding agency achieved a higher overall rate of openness than the whole sample set (51% as opposed to 41%). However, at 51% this is still low considering such projects are funded specifically because they are deemed to be socially or economically valuable research to pursue and therefore worthy of targeted public funding. It should also be noted that there is a good deal of variance within the individual agencies (as low as 33% to as high as 70% open), which again evidences the lack of co-ordination amongst funders, including the government, in New Zealand. We can also see that Gold and Hybrid account for 27% of all these research outputs supported by our major funding agencies. This compares to only 19% of all articles being made open via the two paid open access routes; indeed, this figure reduces to 16% when we consider only articles with a corresponding author from New Zealand. Thus, researchers with this kind of funding are more likely to publish by paying an APC.

As we have seen, 3090 articles that were closed could have been deposited in a repository. This number will have increased in the time that has elapsed since we conducted our analysis and the publication of this paper, since 24-month embargoes will have also expired. This represents an interesting consideration for universities: can we harness the higher citation rates achieved by Green open access and what would we need to invest to do so, for example through a mass deposit project? A 2015 study found that the processing cost of depositing an article in an

institutional repository, including the time of the author, was £33 (or about US$43) (Johnson et al. 2016). Using this figure the 3,090 articles that are closed but could be open would cost US$132,870. This compares to the US$1.45 million identified in our project as potentially paid in Gold and Hybrid APCs and the amount reported by CONZUL as spent in 2017 by university libraries on subscription to electronic resources, NZ$68.5 million (around US$45 million) (Universities New Zealand 2019).

Limitations of this research

Finally we note some limitations of our work and possibilities for further research or strategic implications for New Zealand universities.

We reiterate that the programmatic nature of our method means this does not represent all research, only articles with a DOI. Thus there will be disciplinary skews to the sample set, since journal articles and DOIs are more prevalent in certain disciplines. The research could easily be expanded to incorporate book chapters or other types of work that have a DOI. Nevertheless, not all research falls within the scope of our analysis.

As we have noted in our discussion above, Unpaywall, upon which much of our data gathering depends, updates its database constantly, including the repositories it sources information from. Thus any time the Program is run the results depend on the current state of the Unpaywall database. This can result in fluctuations in results even using the same set of DOIs when the Program is run at different times. Bronze access articles may, by their nature, change status over time. This is not, in itself, problematic, but is noted here only because any set of data produced by the Program is a snapshot of a moment in time. We do intend to re-run the Program and do an analysis each year to track trends over time.

Another limitation is that our calculations of the amount spent on APCs is a maximum amount based on published prices, as noted in the section on our findings. Actual amounts paid will almost certainly be less because there will have been waivers or discounts applied.

Conclusions

In May 2019 we ran our specially-developed software to discover that about two out of every five articles authored by New Zealand researchers in 2017 were freely available on the web (41%). This is the first time we have an evidence-based picture of access to research by New Zealand universities with such detail since, as a result of our work, we have far more than a simply overall proportion: we can investigate the ways in which work has been made accessible, we can compare the average citation rates for these different modes of access, we can quantify the

volume of works that are closed access and could be made open and we can estimate how much paid forms of open access have cost.

Since our code is publicly available, anyone can run their own set of DOIs to perform their own analysis of these aspects.

Overall, we see that more New Zealand research from 2017 is behind a paywall than is freely accessible (41% freely accessible, 59% closed). However, when the corresponding author was a New Zealand researcher the open figure drops to around a third (34%). When our major funding agencies have specifically funded the research the proportion of articles that is accessible is higher but still just under half is accessible without a subscription. These figures are rather sobering, especially in the context of our finding that closed articles, on average, receive far fewer citations than all forms of open access bar the seldom-used free-to-publish-free-to-read Diamond option.

Where work is freely accessible, Gold is the most likely means of achieving this at an average cost of USD1,682 per article; while Hybrid is used significantly less often it comes with a higher average price (USD2,558). In all the two paid methods of making research accessible comes with an estimated price tag – on top of library subscription costs, of course – of USD1.45m. Hybrid's cost may be seen as "worth it", given that our findings suggest it achieves a higher average citation rate than all forms except Green. This would also lend weight to the idea that researchers are unaware of their Green options. These theories would be a fruitful avenue to explore further by studying researchers' publication choices in the New Zealand context.

Green open access accounted for about one-quarter of our open articles. One further avenue we can investigate is where this work was archived, whether in our own university repositories or in public ones like PubMed. Significantly, Green articles achieved an average citation above other forms of open access and well above closed articles; the average rate was on par with the more expensive paid option (Hybrid). However we found that this proportion could be greatly increased if our authors utilised the rights afforded to them by publishers to make versions of their work freely accessible in non-commercial repositories. Fully 88% of closed articles could be made available in this way.

These findings beg several questions worthy of further research. What are the barriers to self-archiving? The most likely reasons – of which we are aware from our own anecdotal experiences – are lack of time, lack of awareness of the possibility of self-archiving, confusion about copyright and embargo periods, negative perceptions of the status of author accepted manuscripts, and the lack of user-friendliness of software used to deposit works in a repository. Why do our researchers choose one mode of publication over another? Which publishers do our researchers favour when choosing open? What influences them to choose to pay a Hybrid APC?

Does journal impact factor play a role in decisions or in average citation rates? Are there disciplinary differences?

What we do know is that New Zealand research is less likely to be open than research of other countries. Our overall proportion of open work lags behind other countries, our corresponding authors are less likely to make research open than corresponding authors from other countries and, clearly, we could be taking advantage of Green open access, and its apparent citation advantage, to a far greater extent than we are. This last point in particular suggests there are important policy and systemic issues that should be considered by New Zealand's research community. Despite the fact we know most authors support open access to research in principle, there is a very large gap between this belief and their practices in making New Zealand's research outputs free to access.

Acknowledgements

This research formed one part of a larger project under the auspices of the Council of New Zealand University Librarians. The authors acknowledge CONZUL's guidance and support.

References

Abediyarandi N, and Mayr P. 2019. The state of open access in Germany: An analysis of the publication output of German universities. International Conference on Scientometrics & Informetrics. Rome, Italy.

Adie E. 2014. Attention! A study of open access vs non-open access articles. Altmetric blog.

Archambault É, Amyot D, Deschamps P, Nicol A, Provencher F, Rebout L, and Roberge G. 2014. Proportion of open access papers published in peer-reviewed journals at the European and world levels—1996–2013. Available at https://digitalcommons.unl.edu/scholcom/8/2014).

Archambault E, Amyot D, Deschamps P, Nicol A, Rebout L, and Roberge G. 2013. Proportion of open access peer-reviewed papers at the European and world levels—2004-2011. Available at https://www.science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf2013).

Archambault É, Côté G, Struck B, and Voorons M. 2016. Research impact of paywalled versus open access papers. Available at https://digitalcommons.unl.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1028&context=scholcom2016).

Björk BC, and Solomon D. 2015. Article processing charges in OA journals: Relationship between price and quality. *Scientometrics* 103:373-385. DOI:10.1007/s11192-015-1556-z

Bjork BC, Welling P, Laakso M, Majlender P, Hedlund T, and Gudnason G. 2010. Open access to the scientific journal literature: Situation 2009. *PLoS One* 5:e11273. DOI:10.1371/journal.pone.0011273

Bosman J, and Kramer B. 2019. Publication cultures and Dutch research output: A quantitative assessment. Available at https://zenodo.org/record/2643360#.Xid4X8gzaUk2019).

Centre for Science and Technology Studies. 2018. CWTS Leiden Ranking 2018. Available at https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:970852018).

Copiello S. 2019. The open access citation premium may depend on the openness and inclusiveness of the indexing database, but the relationship is controversial because it is ambiguous where the open access boundary lies. *Scientometrics* 121:995-1018. DOI:10.1007/s11192-019-03221-w

Fraser M, White R, Richardson E, Hayes L, Howie J, White B, Angelo A, Broughton S, and Fitchett D. 2019. Open Access in New Zealand universities: An environmental scan. Available at https://www.universitiesnz.ac.nz/about-universities-new-zealand/unz-committees-and-working-groups/council-new-zealand-university2019).

Gargouri Y, Hajjem C, Lariviere V, Gingras Y, Carr L, Brody T, and Harnad S. 2010. Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS One* 5:e13636. DOI:10.1371/journal.pone.0013636

Gaulé P, and Maystre N. 2011. Getting cited: Does open access help? *Research Policy* 40:1332-1338. DOI:10.1016/j.respol.2011.05.025

Guédon JC, Kramer B, and Laakso M. 2019. Future of scholarly publishing and scholarly communication: Report of the Expert Group to the European Commission. Available at http://digitalcommons.unl.edu/scholcom/972019).

Holmberg K, Hedman J, Bowman TD, Didegah F, and Laakso M. 2019. Do articles in open access journals have more frequent altmetric activity than articles in subscription-based journals? An investigation of the research output of Finnish universities. *Scientometrics* 122:645-659. DOI:10.1007/s11192-019-03301-x

Ithaka S+R. 2018. CONZUL faculty survey: Aggregate report of findings.

Jahn N, and Tullney M. 2016. A study of institutional spending on open access publication fees in Germany. *PeerJ* 4:e2323. DOI:10.7717/peerj.2323

Johnson R, Pinfield S, and Fosci M. 2016. Business process costs of implementing “gold” and “green” open access in institutional and national contexts. *Journal of the Association for Information Science and Technology* 67:2283-2295. DOI:10.1002/asi.23545

Kirkman NS. 2018. A study of open access publishing by NHMRC grant recipients Master of Philosophy. Curtin University.

Laakso M, Welling P, Bukvova H, Nyman L, Bjork BC, and Hedlund T. 2011. The development of open access journal publishing from 1993 to 2009. *PLoS One* 6:e20961. DOI:10.1371/journal.pone.0020961

Lovén L. 2019. Monitoring open access publishing costs at Stockholm University. *Insights* 32. DOI:10.1629/uksg.451

Maddi A. 2019. Construction of a Normalized Open Access Indicator (NOAI). Available at https://hal.archives-ouvertes.fr/hal-02328158/document2019).

Martín-Martín A, Costas R, van Leeuwen T, and Delgado López-Cózar E. 2018. Evidence of open access of scientific publications in Google Scholar: A large-scale analysis. *Journal of Informetrics* 12:819-841. DOI:10.1016/j.joi.2018.06.012

Matthias L. 2018. publisher-oa-portfolios. Available at https://github.com/lmatthia/publisher-oa-portfolios2018).

McCabe MJ, and Snyder CM. 2014. Identifying the effect of open access on citations using a panel of science journals. *Economic Inquiry* 52:1284-1300. DOI:10.1111/ecin.12064

Mikki S, Gjesdal Ø, and Strømme T. 2018. Grades of openness: Open and closed articles in Norway. *Publications* 6:1-12. DOI:10.3390/publications6040046

Monaghan J, Lucraft M, Allin K, Graaf Mvd, and Clarke T. 2020. 'APCs in the wild': Exploring funding streams for an accelerated transition to Open Access. Available at https://figshare.com/articles/_APCs_in_the_Wild_Could_Increased_Monitoring_and_Consolidation_of_Funding_Accelerate_the_Transition_to_Open_Access_/119881232020).

Morrison H, Mondésir G, Salhab J, Villamizar C, Calvé-Genest A, and Desautels L. 2016. Open Access Article Processing Charges (OA APC) longitudinal study 2015 preliminary dataset. 2016).

Nazim M, and Zia S. 2019. Acceptance and adoption of open access publishing by researchers in India. *Global Knowledge, Memory and Communication* 68:148-158. DOI:10.1108/gkmc-09-2018-0077

New Zealand Government. 2014. New Zealand Government Open Access and Licensing framework (NZGOAL). Available at https://www.data.govt.nz/assets/Uploads/nzgoal-version-2-december-2014.pdf2014).

Ottaviani J. 2016. The post-embargo open access citation advantage: It exists (probably), its modest (usually), and the rich get richer (of course). *PLoS One* 11:e0159614. DOI:10.1371/journal.pone.0159614

Pinfield S, Salter J, and Bath PA. 2017. A "Gold-centric" implementation of open access: Hybrid journals, the "Total cost of publication," and policy development in the UK and beyond. *Journal of the Association for Information Science and Technology* 68:2248-2263. DOI:10.1002/asi.23742

Piryani R, Dua J, and Singh VK. 2019. Open access levels and patterns in scholarly articles from India. *Current Science* 117:1435-1440.

Piwowar H, Priem J, Lariviere V, Alperin JP, Matthias L, Norlander B, Farley A, West J, and Haustein S. 2018. The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ* 6:e4375. DOI:10.7717/peerj.4375

Piwowar H, Priem J, and Orr R. 2019. The Future of OA: A large-scale analysis projecting open access publication and readership. DOI:10.1101/795310

Pölönen J, Guns R, Kulczycki E, Laakso M, and Sivertsen G. 2019. Open access challenge at the national level: Comprehensive analysis of publication channels used by Finnish researchers in 2016-2017. International Conference on Scientometrics & Informetrics. Rome, Italy.

Robinson-Garcia N, Costas R, and Leeuwen TNV. 2019. Indicators of open access for universities. International conference on scientometrics & informetrics. Rome, Italy.

Sivertsen G, Guns R, Kulczycki E, and Pölönen J. 2019. The use of Gold open access in four European countries: An analysis at the level of articles. International Conference on Scientometrics & Informetrics. Rome, Italy. p 1600-1605.

Solomon D, and Björk B-C. 2016. Article processing charges for open access publication: The situation for research intensive universities in the USA and Canada. *PeerJ* 4:e2264. DOI:10.7717/peerj.2264

Torres-Salinas D, Robinson-Garcia N, and Moed HF. 2019. Disentagling Gold open access. In: Glanzel W, Moed HF, Schmoch U, and Thelwall M, eds. *Springer handbook of science and technology indicators*, 129-145.

Universities New Zealand. 2019. Submission on review of the Copyright Act 1994: Issues Paper.

Unpaywall. Frequently Asked Questions. *Available at https://unpaywall.org/faq*

Wang X, Cui Y, Xu S, and Hu Z. 2018. The state and evolution of Gold open access: A country and discipline level analysis. *Aslib Journal of Information Management* 70:573-584. DOI:10.1108/ajim-02-2018-0023

Wang X, Liu C, Mao W, and Fang Z. 2015. The open access advantage considering citation, article usage and social media attention. *Scientometrics* 103:555-564. DOI:10.1007/s11192-015-1547-0

Wikipedia contributors. 2020. Open access in New Zealand. *Available at https://en.wikipedia.org/w/index.php?title=Open_access_in_New_Zealand&oldid=953400278*.