MedCrave
Step into the World of Research

Journal of Human Virology & Retrovirology

Letter to the Editor

Open Access

CrossMark

Open source COVID research and the long tail effect

Keywords: open source, open research, open publishing, SARS-CoV-2, COVID-19, long tail, curiosity, innovation

Volume 8 Issue 3 - 2020

Cameron L Jones[1,2]
[1]Biological Health Services, Australia
[2]National Institute of Integrative Medicine, Australia

Correspondence: Dr. Cameron L. Jones, Biological Health Services, Level 1, 459 Toorak Rd, Toorak, Victoria, 3142, Australia, Tel +61414998900, Email info@biologicalhealthservices.com.au

Received: June 19, 2020 | **Published:** July 06, 2020

Introduction

The global pandemic caused by the SARS-CoV-2 virus currently exceeds 8.5M cases and without reliable preventative treatments or a vaccine, the mortality rate from COVID-19 has already passed 450K. In response, and over the last six months an enormous volume of academic and clinical research has emerged online. PubMed® is the largest search engine for biomedical literature and comprises more than 30M citations for literature from MEDLINE and other journals and publisher sites like PubMed Central®.[1,2] PubMed's role as the dominant resource for access to scholarly communication is widely acknowledged[3] but faces increasing competition from open source alternatives.[4] Research and development including the drive to publish is a labour-intensive activity requiring a high skill set, access to resources and time commitment in order to make scientific discoveries. Historically, the dominant and most familiar publishing model for the majority of traditional Journals listed on PubMed® involves scientists using public or private funds to pay for the research, who in turn pay for publication, and then pay again to read the final published works.[5] Preprint servers accelerate scholarly publishing through open source where content is not behind paywalls like traditional scientific publishers and feedback is encouraged prior to formal peer-review. The main benefit of this publishing pathway is immediacy and the ability to rapidly introduce new information into the academic community. Disadvantages are that search and retrieval can be more difficult, while a lack of or inadequate peer review could allow incorrect findings to be introduced or shared. These effects are likely reflected in the variable adoption pattern seen for some fields of study versus others.[6] Nevertheless, scientific publishing in preprints[7] and into lesser known Journals is increasing.[8] In fact, the trend towards publishing as open access versus the traditional Journal pathway is based on four factors including visibility, cost, prestige and speed.[9] In most cases, open source publishing offers undeniable cost-benefits and speed to publication.

In 2004 a concept called the Long Tail was popularised by the editor of Wired magazine.[10] In this famous essay, the digital future was summarized as: "Forget squeezing millions from a few megahits at the top of the charts. The future of entertainment is in the millions of niche markets at the shallow end of the bitstream". Through the lens of the long tail we can view PubMed® as the aggregator of 'hits' since they are most easily found and consumed. On PubMed®, the topic of the article as well as its publication date determine popularity.[11] But what about all the other research published elsewhere on preprint repositories – do these fulfil the criteria of 'niches' making up an ever-expanding long tail?

The purpose of this Letter is twofold. First, it is to highlight the power law scaling distribution for COVID-related publications available across different preprint servers at two time points (Figure 1). Second, it is to highlight how concepts like the long tail can be used to explore niche preprint repositories as a source of innovation that could lead to cross-disciplinary opportunities for novel discoveries. This second consideration is based on remix culture, where curiosity, knowledge sharing and reuse is a generator of new discovery.[12]

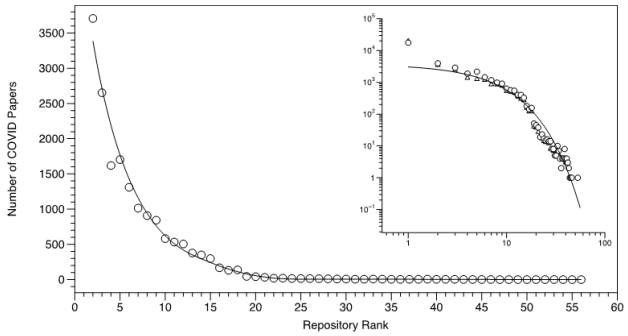

Figure 1 The main graph shows the keyword search for the term "COVID" performed over all the Paper Server.

Repositories listed in Table 1 spanning a 2-week period were then used to rank the number of COVID papers as the average. The long tail is evident towards the right-hand side of the distribution. Inset on log-log axes for this same data set, a power law emerges shown with an exponential least-squares fit having r^2 of 0.94 and power of -0.19. For comparison, a straight-line regression showed an r^2 of 0.90 and power of -2.84. Removing the PubMed® data point caused the exponential fit to increase with an r^2 of 0.95 and a power of -0.18; similarly for the straight line fit, the r^2 was 0.91 and power -3.08 which preferentially captured the open source papers. Open circles and triangle symbols are used for the earlier and later dates respectively for the inset graph which also emphasizes that the niches are filling quickly.

The COVID crisis presents a new tipping point for scientific publishing according to Unesco[13] and a recent analysis showed that COVID preprints are distributed at least 15-times more than non-COVID preprints.[14] Currently, there is high-motivation for authors to generate novel research and produce public-good contributions, together with an expanding pool of servers ready to digitally disseminate this information (Table 1). Since scholarship and new

idea formation is built on the work of others, when information costs reduce, the opportunity to accelerate new ideas and generate discoveries expands. It is my position that the different preprint servers act as digital aggregators which pre-filter content according to topic niche and act as post filters by assigning for example a Digital Object Identifier (DOI) or other digital tag that enables the object (the research paper) to be promoted on blogs and other social networks which is a type of recommender system. The repository is therefore the connector between supply and demand. Importantly, collaborative innovation within a Company[15] (Pfizer) has been shown to follow a power law where rank frequency plots similar to Figure 1 for new ideas showed strong straight-line power laws with exponents spanning -2.7 to -3. This observation supports the claim that COVID research and publication output, regardless of where it resides on the internet emerges as a type of collective intelligence. At this point and with a solution-focused mindset across the range of potential problems surrounding the impact of COVID on society it is worth repeating the 8 key questions about curiosity and innovation that were developed by the Defense Advanced Research Projects Agency (DARPA) to assist with the discovery process.[16] In practice, extracting and contributing new meaning into and out of the niches (or the hits) would take advantage of the 8 steps as a catalyst for novel idea generation and publication of same. These are: (1) What are you trying to do? Articulate your objectives using absolutely no jargon. (2) How is it done today, and what are the limits of current practice? (3) What is new in your approach and why do you think it will be successful? (4) Who cares? If you are successful, what difference will it make? (5) What are the risks? (6) How much will it cost? (7) How long will it take? (8) What are the mid-term and final "exams" to check for success?

Table 1 Ranked list of paper repositories showing the number of papers returned for the "COVID" search string and their respective topic focus areas. The top ranked repositories reflect content where the "rich get richer" leading to "hits" caused by centrality-based search and cumulative advantage. By contrast, niche papers published in lower ranked repositories might be obscure but will continue to attract a small fan club due to the "birds of a feather" aggregation concept caused by collaborative filtering.

Rank	Paper server repository	Number of papers 26/05/20	Number of papers 09/06/20	Site description
1	PubMed®	15385	19,824	Biomedical literature from MEDLINE, life science journals, and online books https://pubmed.ncbi.nlm.nih.gov/
2	MedRxiv	3452	3,962	The preprint server for health sciences https://www.medrxiv.org/
3	SSRN	2425	2,881	Tomorrow´s Research Today https://www.ssrn.com/index.cfm/en/
4	RePEc & IDEAS	1363	1876	Research papers in economics and finance https://ideas.repec.org/
5	EconPapers	1247	2159	The world's largest collection of on-line Economics working papers, journal articles and software econpapers.repec.org/
6	figshare	1181	1,438	Figshare is a repository where users can make all of their research outputs available in a citable, shareable and discoverable manner https://figshare.com/
7	arXiv	863	1,164	arXiv is a free distribution service and an open-access archive for scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics https://www.arXiv.org
8	bioRxiv	837	979	The preprint server for biology https://www.biorxiv.org/
9	Zenodo	774	912	Open Science https://zenodo.org/
10	Authorea	527	637	Collaborative platform to read, write, and publish research https://www.authorea.com/covid
11	Preprints	502	563	The Multidisciplinary Preprint Platform https://www.preprints.org/
12	HAL	463	547	Open archive where authors can deposit scholarly documents from all academic fields. https://hal.archives-ouvertes.fr/
13	Coronavirus Disease Research Community - COVID-19	352	401	Coronavirus Disease Research Community - COVID-19 https://zenodo.org/communities/covid-19/
14	National Bureau of Economic Research	300	400	Working papers and publications on economic research https://admin.nber.org/
15	PsyArXiv	270	327	A free preprint service for the psychological sciences https://psyarxiv.com/
16	ChemRxiv	155	181	The Preprint Server for Chemistry https://chemrxiv.org/

Table continued...

Rank	Paper server repository	Number of papers 26/05/20	Number of papers 09/06/20	Site description
17	Bepress Legal Repository	120	143	Working papers and pre-prints from scholars and professionals at top law schools around the world https://law.bepress.com/
18	SocArXiv	119	157	Research papers in social sciences https://osf.io/preprints/socarxiv; https://socopen.org/
19	Outbreak Science	40	51	Rapid, open review of preprints related to outbreaks https://outbreaksci.prereview.org/
20	F1000Research	37	45	An open access publishing platform supporting data deposition and sharing all types of research https://f1000research.com/
21	Advance	27	38	A SAGE preprints community for humanities and social sciences https://advance.sagepub.com/
22	ChinaXiv	19	19	Chinese scientific and technical paper pre-publishing Platform http://chinaxiv.org/
23	EdArXiv	17	23	A preprint server for the education research community https://edarxiv.org/
24	IndiaRxiv	14	17	A preprints repository service for India https://indiarxiv.org/
25	Neliti	13	15	Indonesia's research repository https://www.neliti.com/
26	e-Lis	13	16	e-prints in library & information science http://eprints.rclis.org/
27	Emerald Open Research	12	14	A platform for fast author-led publication and open peer review. Aligned with the United Nations Sustainable Development Goals https://emeraldopenresearch.com/
28	EngrXiv	8	14	The open archive of engineering https://engrxiv.org/
29	AfricArXiv	8	8	The preprint repository of African research https://info.africarxiv.org/; https://info.africarxiv.org/submit-via-osf/
30	EarthArXiv	7	8	A free preprint service for the Earth sciences https://eartharxiv.org/
31	SportRXiv	5	5	The open access subject repository for sport, exercise, performance, and hHealth research https://osf.io/preprints/sportrxiv
32	EcoEvoRxiv	5	5	A free preprint service for ecology, evolution and conservation https://ecoevorxiv.org/
33	PeerJ	5	10	the Journal of life and environmental sciences https://peerj.com/
34	FrenXiv	5	7	The French server for preprints in all the scientific fields https://frenxiv.org/
35	LawArXiv	4	4	Legal scholarship in the open (http://lawarxiv.info/; https://osf.io/preprints/lawarxiv
36	AgriXiv	4	2	Preprints for agriculture and allied sciences https://agrixiv.org/
37	NutriXiv	4	4	A free preprint service for the nutritional sciences https://osf.io/preprints/nutrixiv
38	MetaArXiv	4	4	An interdisciplinary archive of articles focused on improving research transparency and reproducibility https://osf.io/preprints/metaarxiv/
39	Thesis Commons	4	8	An open archive of theses https://thesiscommons.org/
40	BioHackrXiv	3	4	Preprints for BioHackathons https://biohackrxiv.org/
41	INA-Rxiv	3	4	The preprint server of Indonesia https://osf.io/preprints/inarxiv/
42	AgEcon	3	3	Research in agricultural and applied economics (https://ageconsearch.umn.edu/)

Table continued...

Rank	Paper server repository	Number of papers 26/05/20	Number of papers 09/06/20	Site description
43	MediArXiv	2	2	Open archive for media, film, and communication studies https://mediarxiv.org/
44	ArabiXiv	1	1	The Arabic open science repository (https://arabixiv.org/)
45	PubPsych	1	1	PubPsych is a free information retrieval system for psychological resources https://pubpsych.zpid.de/
46	MitoFit	1	1	The Open access preprint server for mitochondrial physiology and bioenergetics - https://www.mitofit.org/; https://www.mitofit.org/index.php/MitoFit
47	PaleoarXiv	0	0	A preprint archive for Paleontology (https://paleorxiv.org/)
48	MarXiv	0	0	Preprint repository for ocean and marine-climate research https://marxiv.org/
49	BodoArXiv	0	0	Open repository for medieval studies https://bodoarxiv.org/; https://osf.io/preprints/bodoarxiv
50	ECSarXiv	0	0	Preprint service for electrochemistry and solid state science and technology https://ecsarxiv.org/
51	FocUS	0	0	A free preprint service for the focused ultrasound research community (https://osf.io/preprints/focusarchive)
52	LISSA	0	0	Library and information science scholarship archive https://lissarchive.org/; https://osf.io/preprints/lissa/discover
53	PhilSci Archive	0	1	An archive for preprints in philosophy of science http://philsci-archive.pitt.edu/
54	MindRxiv	0	0	Open archive for research on mind and contemplative practices https://mindrxiv.org/
55	PaelorXiv	0	0	A preprint archive for paleontology https://paleorxiv.org/
56	Policy Archive	0	0	Policy archive is a comprehensive digital library of public policy research https://www.policyarchive.org/

I will conclude this letter by highlighting again from a recent article in Wired magazine reporting how virtual workspaces induced by the COVID-driven work from home is leading to not less but more research, and a greater degree of cross-disciplinary collaboration.[17] Since niche papers are accessible over the internet at low to no cost to the reader, collaborative filtering on topic areas of interest will always ensure a small fan club. It is also very probable that scientists will increasingly need to exploit social media tactics and influencers to promote discoveries (or niche papers) to expand their audience and reach.[18] The desire to contribute, innovate and communicate will always produce the usual 'hit' papers, but the long tail suggests that new and different ideas will emerge from the diversity and ambition written into the papers published in the 'niches'.

Acknowledgments

None.

Conflicts of interest

The author declares no conflicts of interest.

Funding

No funding.

References

1. PubMed. PubMed. 2020.
2. PubMed Central®. Ncbi.nlm.nih.gov. 2020.
3. Williamson PO, Minter CIJ. Exploring PubMed as a reliable resource for scholarly communications services. *J Med Libr Assoc*. 2019;107(1):16–29.
4. Gasparyan AY, Yessirkepov M, Voronov AA, et al. Comprehensive Approach to Open Access Publishing: Platforms and Tools. *J Korean Med Sci*. 2019;34(27):e184.
5. Gierasch LM. On the costs of scientific publishing. *J Biol Chem*. 2017;292(39):16395–16396.
6. Chiarelli A, Johnson R, Richens E, Pinfield S. Accelerating scholarly communication The transformative role of preprints. 2019.

7. Matthew B Hoy. Rise of the Rxivs: How Preprint Servers are Changing the Publishing Process, Medical Reference Services Quarterly. 2020;39:84–89.
8. Bohannon J. Uprising: Less prestigious journals publishing greater share of high-impact papers. *Science AAAS*. 2020.
9. Conte S. Making the Choice: Open Access vs. Traditional Journals. 2020.
10. Anderson C. The Long Tail. WIRED. 2020.
11. Smith L, Wilbur W. The Popularity of Articles in PubMed. 5: 1-7.
12. Flath CM, Friesike S, Wirth M. et al. Copy, transform, combine: exploring the remix as a form of innovation. *J Inf Technol.* 2017;32:306–325.
13. Open access to facilitate research and information on COVID-19. UNESCO. 2020.
14. Fraser N, Brierley L, Dey G, et al. Preprinting a pandemic: the role of preprints in the COVID-19 pandemic. *Biorxiv*. 2020
15. Woods T. The Long Tail of Idea Generation. *International Journal of Innovation Scienc*e. 2010;2(2):53–63.
16. Isaacs K, Ancona D. 3 Ways to Build a Culture of Collaborative Innovation. 2019.
17. Majumder M. Coronavirus Researchers Are Dismantling Science's Ivory Tower—One Study at a Time. Wired. 2020.
18. Galetti M, Costa-Pereira R. Scientists need social media influencers. *Science (New York, N.Y.)*. 2017;357(6354):880–881.