

UWM Publications as Preprints

By Svetlana Korolev, Science Librarian, UWM Libraries

skorolev@uwm.edu

Introduction

The popularity of preprints has increased remarkably since the COVID-19 pandemic outbreak had underlined the urgency for rapid dissemination of research results. The primary benefits of preprints are early publication and open access. Preprints enter the scholarly communication workflow first by the authors uploading them to a designated, usually discipline specific, server before (or at the same time) submitting them to a journal for peer review and potential article publication. The following are the secondary advantages of posting preprints:

- Ensure the originality of discovery to authors and their institutions based on the priority record with the assigned DOI (Digital Object Identifier)
- Searchable on preprint servers, Google Scholar, and on some indexing platforms
- Visible widely with options for improving the manuscripts based on feedback by peers and for updating them in subsequent versions by authors
- Citable within a longer “citation window” period.

Since there is a broad range of publication types called “preprints,” for this project I refer to this definition: “A preprint is a scientific manuscript that is uploaded by the authors to a public server...typically posted online within a day or so without peer review and can be viewed (and possibly reused depending on the license) without charge by anyone in the world.”¹

Background

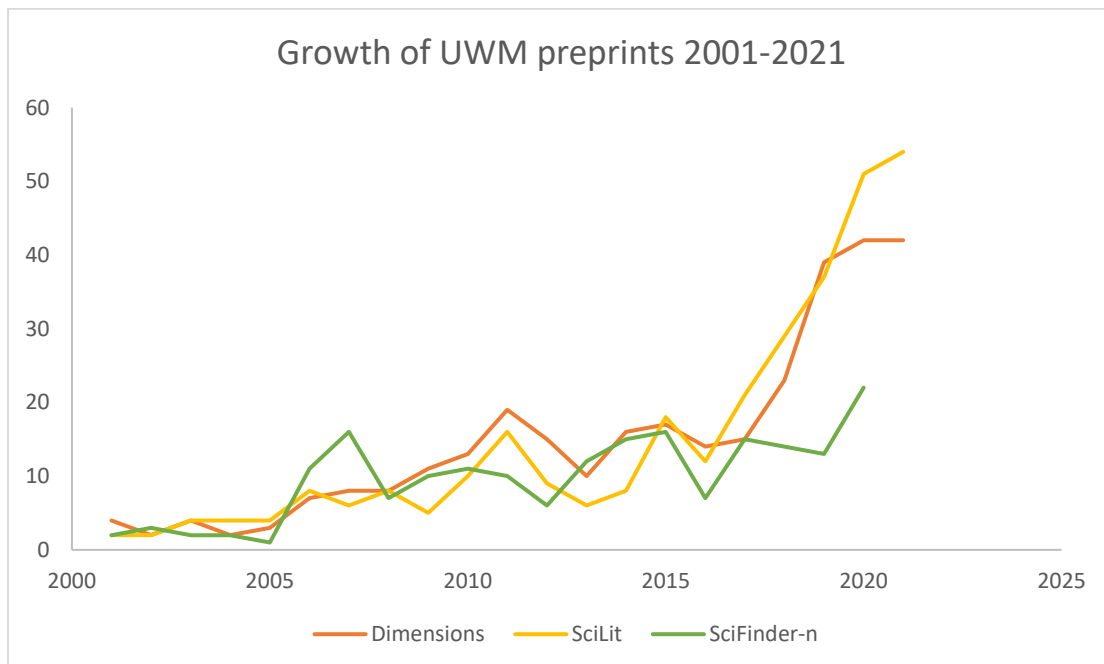
The UWM Libraries currently provide the four most popular preprint server platforms - arXiv (which marks its 30th anniversary this year), bioRxiv, ChemRxiv, and SSRN - in our Databases A-Z list. In response to the growing popularity of the preprint model, many publishers have revised their journal editorial policies encouraging posting of primary research manuscripts on preprint servers and disclosing details of preprint’s DOI and licensing terms upon submissions to a journal.² A list of publisher policies can be found on Sherpa Romeo³ or Wikipedia.⁴ It was a step forward to Open Science when many publishers agreed to share research findings relevant to the novel coronavirus outbreak in response to the call issued by the Wellcome Trust in January 2020.⁵

Concept

The purpose of this project was to explore the UWM publications as preprints and to identify the servers being used by UWM researchers, and the indexing sources for their discovery. My interest in this study is related to coordinating of the University Open Access Publication Fund that offers some support to cover article processing charges. In previous studies I looked at the “gold” open access publications by UWM authors based on records in Web of Science⁶ and MDPI.⁷ This time I attempted to explore the “green” open access route focusing on preprints posted by UWM authors at the designated servers (while excluding another “green” path such as uploading author accepted manuscripts into institutional repositories). I began this project in the spring of 2021 by activating a short complimentary trial access to Dimensions Analytics. I planned to finish it by the Open Access Week, October 25-31, 2021. This year’s theme, *It Matters How We Open Knowledge: Building Structural Equity*, is aligned with the UNESCO Recommendation draft on fostering a culture of Open Science. The Recommendation cites preprints as an example of “promoting open science from the outset of the research process and extending the principles of openness in all stages of the scientific process to improve quality and reproducibility, including the encouragement of community-driven collaboration and other innovative models, for example preprints, clearly distinguished from final peer-reviewed publications, and respecting the diversity of scientific practices, in order to accelerate dissemination and encourage rapid growth in scientific knowledge.”⁸

Discussion

My analysis of the numbers of preprints retrieved by four databases showed an increased popularity for Open Science by UWM researchers. The following chart depicts data from three databases during the specific time frame.



1. Search engines used for finding UWM preprints

I explored several indexing platforms for finding preprints as well as checked the individual hosting repositories but encountered a difficulty because those tools did not have a search option by author affiliation or organization. In the end I narrowed my choices to four multidisciplinary databases: the first offered a free version and a complimentary trial to a subscription version (Dimensions), the second was a subscription-based chemistry database (SciFinderⁿ), and the last two (SciLit and Lens) were free to use. All four databases had an option to refine the search results by a publication type: preprint. Next, I briefly describe each source with comments and provide two summary tables below.

1. Dimensions
1. SciLit
2. SciFinderⁿ
3. Lens

Dimensions (<https://www.dimensions.ai/>) is a comprehensive database of linked data from publications including preprints, data sets, grants, patents, clinical trials, and policy documents. The platform was developed by Digital Science with the goal “to provide a fresh take on research information; a more open and comprehensive data infrastructure that empowered users to explore connections between a wide range of research data.” It is available as a free version with all publications and citations for personal use and a subscription-based version called Dimensions Analytics with advanced search options including a field for an affiliation and enhanced visuals for data analysis. I used this version during a complimentary trial access in April 2021 and then updated this year’s data in September 2021. Dimensions found many of UWM preprints with the majority of records from SSRN in the research categories of Economics, Applied Economics, Econometrics, Banking and Finance, and Commerce and Management. The second group of records was from bioRxiv in the research categories of Biological Sciences, Genetics, and Medical and Health Sciences. When comparing records with other search engines, I noted that several preprints hosted by Copernicus and Research Square were not included.

SciLit (<https://www.scilit.net/>) is a free search engine of scholarly works developed and maintained by open access publisher MDPI. It crawled across a wide range of servers uniquely identifying UWM recent preprints from Copernicus, and overall, the largest number of the 2021 preprints, with an unlikely high percentage of recent preprint records (4% in 2020 and 7% in 2021), some of which could be due to versioning of the same preprint. Unfortunately, after retrieving the set of results and applying a filter for “preprint” the refine features stopped working, and the analysis of the data had to be done manually. It was also inconvenient that both sources of bioRxiv and medRxiv were compiled together for the Cold Spring Harbor Laboratory. Oddly, SciLit did not show recent UWM preprints from arXiv.

SciFinderⁿ (<https://scifinder-n.cas.org/>) is a subscription-based resource for substances and reactions produced by CAS (Chemical Abstracts Service). I trusted this source the most because of its rigorous editorial reputation and its advanced search option for organization name (University of Wisconsin-Milwaukee) and refine by publication type (preprint). Since this is a chemistry focused database, the UWM preprints were found only from two preprint servers: arXiv and bioRxiv. Among the 202 preprints, the majority of records (188) was found from

arXiv (since 2000) and 14 records were from bioRxiv (since 2016). The top research categories of UWM preprints were all in the physics sub-disciplines including astrophysics, high energy physics, and condensed matter. Per SciFinderⁿ preprints comprised 2% of UWM publications in 2020 and only under 1% in 2021. Those numbers seemed very small. In fact, I identified a few missing records after testing for individual manuscript titles uploaded by UWM researchers into bioRxiv and ChemRxiv. The records of subsequent journal articles with the same titles were retrieved in the set of UWM publications but not included in the set of UWM preprints.

The Lens (<https://www.lens.org/>) is a public aggregator of metadata combining three unique sets of scholarly works, patents, and biological sequences disclosed in patents; and is supported by Cambia. It was a straightforward path on the Scholarly Works interface via Structure Search to select from the Index fields, for an institution (University of Wisconsin-Milwaukee) and for a publication type (preprint). The Lens resulted in the smallest number of only 45 UWM preprints. The top three disciplines were grouped for Biology (10 records), Psychology (10 records), and Chemistry (7 records). Among the 45 UWM preprints, most records (39) was hosted by the Cold Spring Harbor Laboratory (30 from bioRxiv and 9 from medRxiv). Overall, the Lens aggregated UWM records from five servers, including some preprints omitted by Dimensions.

Table 1: UWM preprints retrieved by databases

	Dimensions	SciLit	SciFinder ⁿ	Lens
arXiv	52	22	188	0
SSRN	158	153	0	1
bioRxiv	87	77	14	30
medRxiv	10	10	0	9
Research Square	8	20	0	4
Copernicus	0	16	0	1
ESSOAr	3	4	0	0
Authorea	3	3	0	0
PeerJ	3	3	0	0
Nature Preceedings	1	0	0	0

2. Preprint servers used by UWM authors

arXiv (<https://arxiv.org/>) is the oldest open access archive founded in 1991. It hosts the largest number of UWM preprints (188) retrieved by SciFinderⁿ since 2000 as well as a few earlier preprints retrieved by Dimensions and SciLit since the first two postings in 1995. The preprints are mostly affiliated with the UWM Department of Physics.

Authorea (<https://www.authorea.com/>) is a collaborative publishing platform and a repository of preprints founded in 2012. Reading and posting preprints is free for authors with an option to pay for premium collaborative writing and editing features. Only three recent 2021 preprints affiliated with UWM researchers in biological sciences were retrieved by Dimensions and SciLit.

bioRxiv (<https://www.biorxiv.org/>) is a preprint server in the life sciences launched by Cold Spring Harbor Laboratory in 2013. It hosts the third largest number of UWM preprints since 2016 and indexed by all four search engines. bioRxiv hosts the top highly cited preprints affiliated with UWM. The two examples are shown below (data were recorded on 10/18/21):

1. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program

- Co-authored by 174 researchers including an author at the UWM Zilber School of Public Health
- Posted to bioRxiv in March 2019 (<https://doi.org/10.1101/563866>)
- Noted in the Dimensions badge: This publication in bioRxiv has been cited 181 times. 80% of its citations have been received in the past two years, which is higher than one might expect, suggesting that it is currently receiving a lot of interest. Compared to other publications in the same field, this publication is extremely highly cited and has received approximately 50 times more citations than average
- Published open access in *Nature* in February 2021 ([10.1038/s41586-021-03205-y](https://doi.org/10.1038/s41586-021-03205-y))
- Indicated 31K accesses and 59 citations on its *Nature* webpage
- Marked this article Highly Cited in the top 1% of the academic field of Molecular Biology & Genetics as of May/June 2021 in Web of Science
- Counted 34 citing times for the preprint and 84 citing times for the article in SciFinderⁿ.

2. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study

- Co-authored by 139 researchers including two authors at the UWM Department of Psychology and an author at the Zilber School of Public Health
- Posted into bioRxiv in November 2018 (<https://doi.org/10.1101/457739>)
- Published in *NeuroImage* in November 2019 (doi: [10.1016/j.neuroimage.2019.116091](https://doi.org/10.1016/j.neuroimage.2019.116091))
- Noted in the Dimensions badge: This publication in bioRxiv has been cited 27 times. 44% of its citations have been received in the past two years. Compared to other publications in the same field, this publication is extremely highly cited and has received approximately 9.53 times more citations than average
- Indicated 98 citing articles on its *NeuroImage* webpage
- Marked this article Highly Cited in the top 1% of the academic field of Neuroscience & Behavior as of May/June 2021 in Web of Science
- Counted 31 citing times for the preprints and 73 citing times for the article in SciFinderⁿ.

The above are just two examples of UWM preprints being published as journal articles. I came across many links from preprints to journal articles on the server platforms, but I did not examine the “preprint – article” workflow systematically within the scope of this study. Preprint servers

usually automatically add a prominent link leading to the published peer-reviewed version within several weeks of journal publication.

Copernicus GmbH (<https://www.copernicus.org/>) supports a preprint server of the European Geosciences Union (EGU), *EGU Sphere* (<https://www.egusphere.net/preprints/>), collecting all submissions to the EGU events and publications in the domain of earth, space, and planetary sciences. Sixteen UWM preprints since 2010 were retrieved only by SciLit (except for one 2014 preprint also found by the Lens). The majority, 14 out of 16, preprints were posted by UWM researchers at the Atmospheric Sciences Group of the Department of Mathematics, and the other 2 preprints were affiliated with the Departments of Geography and Biological Sciences.

ESSOAr, the Earth and Space Science Open Archive (<https://www.essoar.org/>), is governed by a partnership of seven professional societies with financial and technical support by the American Geophysical Union (AGU) and Wiley. This is a relatively recent server for UWM preprints with only four records starting with a conference poster presentation by atmospheric scientists in 2018 followed by three full text papers by the Great Lakes researchers.

medRxiv (<https://www.medrxiv.org/>) covers all aspects of the health sciences. It was launched by a partnership of the Cold Spring Harbor Laboratory (who created bioRxiv earlier in 2013), Yale University, and publisher The BMJ, in summer of 2019. Ten preprints affiliated with UWM researchers at the Zilber School of Public Health, College of Nursing, and Department of Psychology were posted during the last two years 2020-2021.

Research Square (<https://www.researchsquare.com/>) is a multidisciplinary preprint and author services platform. In 2018 they added a free journal-integrated preprint sharing called *In Review* in collaboration with Springer Nature participating journals. Twenty records including *In Review* manuscripts affiliated with a wide range of UWM departments such as biological sciences, computer science, engineering, freshwater sciences, kinesiology, mathematical sciences, psychology, and public health, were posted during the last three years and mostly retrieved by SciLit.

SSRN, Social Sciences Research Network (<https://www.ssrn.com/>), in existence since 1994, is the second oldest preprint platform and was acquired by Elsevier in 2016. It consists of a series of research networks across multiple disciplines now well beyond the social sciences. The papers are classified into topical e-journals that are offered as subscriptions to individuals and institutions. It hosts the second largest number of records (about 160) affiliated with UWM including the first record dated in 1984 but posted retrospectively in 2015. Most of the preprints are affiliated with the Sheldon Lubar School of Business and some others are with the Department of Political Science.

Additionally, UWM preprints were posted on two servers that were no longer supported: **PeerJ** (<https://peerj.com/preprints/>), with three preprints (2015-18) by anthropology, biology, and public health researchers, and **Nature Precedings** (<https://www.nature.com/npre>), with two preprints (2008-09; but only one found by Dimensions) by anthropologists.

Conclusions

There is clearly a growing number of UWM preprints in the recent five years. As many as ten different preprint servers were utilized by UWM researchers beginning with submissions to arXiv in 1995. The majority of those servers are discipline specific, but two are multidisciplinary collaborative platforms. Being aware of the strengths of UWM physical scientists, particularly in the area of astrophysics, I anticipated seeing many preprints on arXiv. Thus, finding almost the same number of preprints by UWM social scientists on SSRN was impressive. There was a relatively small number of preprints in the engineering fields. In 2020, IEEE (Institute of Electrical and Electronics Engineers) announced introducing a new preprint server, TechRxiv (<https://www.techrxiv.org/>). When checking its website, I found two preprints (the 2021 preprint and the 2015 journal article postprint) uploaded by authors at the UWM Department of Electrical Engineering. Those records were not retrieved by databases when the starting inquiry was an organization. Furthermore, the same omissions happened when I compared the UWM preprints at the ChemRxiv website (<https://chemrxiv.org/>) directly to the results sets retrieved by databases. None had included UWM preprints from ChemRxiv in the “institution” and “preprint” query, but the subsequently published journal articles were in the “institution” query. Simple searching by the title of a preprint showed multiple records including preprint versions and a journal article. I noted that some databases marked the publication type as “Other” (instead of “Preprint”) in the records from ChemRxiv and TechRxiv. It was not clear why SciFinderⁿ retrieved by title of the preprint two records, one for an article and one for a preprint, but then after adding a search field for an institution, it only retrieved the journal article. Unfortunately, none of the search engines were perfect for finding all preprints affiliated with a specific institution. Such discovery through databases is currently quite fragmented. Getting a comprehensive examination of preprints affiliated with UWM will require a tedious “title by title” review of multiple versions. I estimate that the actual numbers of UWM preprints are slightly higher than those retrieved for this report. Overall, the practice of posting preprints remains marginal.

Further exploratory steps were beyond the scope of this study. Since preprints promote fast and open content that has not been peer-reviewed, it should be possible to look at the subsequent journal article publications via the links from the preprint record. I noted the existence of those links from the servers to the publishers but did not examine this matter systematically. Also, it is possible to see the usage metrics such as the numbers of downloads and citations, and whether there were comments on UWM preprints. For assessing the reputation of preprints, readers may look at the author names, their affiliations, follow the links to the author profiles on the server, Google Scholar, ORCID, (and PubMed), and peruse supplementary materials and data when provided by authors along with the manuscript.

Appendix

Table 2: UWM preprints retrieved by databases per year and per server

Database		Dimensions	SciLit	SciFinder ⁿ	Lens
Total (as of)		325 (9/25/21)	308 (9/28/21)	202 (9/29/21)	45 (9/28/21)
Oldest from		1995 arXiv	1995 arXiv	2000 arXiv	2014 Copernicus
Newest from		9/13/21 Authorea	9/24/21 bioRxiv	2/19/21 arXiv	8/12/21 medRxiv
2021		42	53	5	20
	arXiv	1	0	3	
	Authorea	3	3		
	bioRxiv	18	17	2	11
	Copernicus	0	1		
	ESSOAr	1	1		
	medRxiv	5	5		5
	Res Square	1	10		4
	SSRN	14	15		
2020		42	45	22	15
	arXiv	4	0	19	
	bioRxiv	23	21	3	10
	Copernicus	0	3		
	ESSOAr	2	2		
	medRxiv	5	5		4
	Res Square	5	5		

	SSRN	3	9		1
2019		39	36	13	5
	arXiv	0	0	12	
	bioRxiv	24	16	1	5
	Copernicus	0	1		
	Res Square	2	5		
	SSRN	13	14		
2018		23	27	14	2
	arXiv	2	0	9	
	bioRxiv	13	13	5	2
	Copernicus	0	1		
	ESSOAr	0	1		
	PeerJ	2	2		
	SSRN	6	10		
2017		15	21	15	2
	arXiv	3	1	14	
	bioRxiv	6	7	1	2
	Copernicus	0	1		
	SSRN	6	12		
2016		14	12	7	0
	arXiv	0	0	5	
	bioRxiv	3	3	2	
	Copernicus	0	2		
	SSRN	11	7		

2015		16	17	16	0
	arXiv	3	2	16	
	Copernicus	0	2		
	PeerJ	1	1		
	SSRN	12	12		
2014		16	8	15	1
	arXiv	4	3	15	
	Copernicus	0	1		1
	SSRN	12	4		
2013		10	6	12	
	arXiv	6	5	12	
	SSRN	4	1		
2012		15	9	6	
	arXiv	4	1	6	
	Copernicus	0	2		
	SSRN	11	6		
2011		19	16	10	
	arXiv	6	2	10	
	SSRN	13	14		
2010		13	10	11	
	arXiv	2	0	11	
	Copernicus	0	2		
	SSRN	11	8		
2009		11	5	10	

	arXiv	1	0	10	
	SSRN	10	5		
2008		8	8	7	
	arXiv	3	2	7	
	Nat Preced	1	0		
	SSRN	4	6		
2007		8	6	16	
	arXiv	1	1	16	
	SSRN	7	5		
2006		7	8	11	
	arXiv	1	1	11	
	SSRN	6	7		
2005		3	4	1	
	arXiv	1	1	1	
	SSRN	2	3		
2004		2	4	2	
	arXiv	0	0	2	
	SSRN	2	4		
2003		4	4	2	
	arXiv	1	0	2	
	SSRN	3	4		
2002		2	2	3	
	arXiv	1	0	3	
	SSRN	1	2		

2001		4	2	2	
	arXiv	0	0	2	
	SSRN	4	2		
2000		1	1	2	
	arXiv	0	0	2	
	SSRN	1	1		
1999	arXiv	1	0		
1998	SSRN	1	1		
1997	arXiv	1	0		
1996	arXiv	4	1		
1995	arXiv	2	2		
1984*	SSRN	1	1		

*written in 1984, posted in 2015

References

1. ASAPbio (Accelerating Science and Publication in biology). Preprint resource center. <https://asapbio.org/preprint-info>
2. Springer. Preprint sharing. <https://www.springer.com/gp/editorial-policies/preprint-sharing>
3. Sherpa Romeo. Open access policies: Publishers. <http://www.sherpa.ac.uk/romeo/journalbrowse.php?la=en&fidnum=|&mode=simple>
4. Wikipedia. List of academic publishers by preprint policy. http://en.wikipedia.org/wiki/List_of_academic_journals_by_preprint_policy
5. Wellcome Trust. (2020). Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak. <https://wellcome.org/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-ncov-outbreak>
6. Korolev, S. (2018). Finding Open Access publications by UWM authors. *UWM Libraries Other Staff Publications*. 12. https://dc.uwm.edu/lib_staffart/12

7. Korolev, S. (2020). UWM publications in MDPI journals. *UWM Libraries Other Staff Publications*, 14. https://dc.uwm.edu/lib_staffart/14
8. UNESCO. (2021). Draft text of the UNESCO recommendation on open science. <https://en.unesco.org/science-sustainable-future/open-science/recommendation>

Further Readings

1. Bourne, P., Polka, J., Vale R., Kiley, R. (2017). Ten simple rules to consider regarding preprint submission. *PLOS Computational Biology*, **13** (5): e1005473. <https://doi.org/10.1371/journal.pcbi.1005473>
2. Brainard, J. (2021). No revolution: COVID-19 boasted open access, but preprints are only a fraction of pandemic papers. *Science*, **373** (6560), 1182-1183. <https://doi.org/10.1126/science.acx9058>
3. Callaway, E. (2020). Will the pandemic permanently alter scientific publishing? *Nature*, **582**, 167-168. <https://www.nature.com/articles/d41586-020-01520-4>
4. Chiarelli, A. Johnson, R., Pinfield, S., Richens, E. (2019). Accelerating scholarly communication: The transformative role of preprints. *Zenodo*. <https://zenodo.org/record/3357727#.YRlhoIhKg2x>
5. Coudert, FX. (2020). The rise of preprints in chemistry. *Nature Chemistry*, **12**, 499-502. <https://doi.org/10.1038/s41557-020-0477-5>
6. Eisen, M., Akhmanova, A., Behrens, T., Harper, D., Weigel, D., Zaidi, M. (2020). Peer Review: Implementing a "publish, then review" model of publishing. *eLife*. <https://elifesciences.org/articles/64910>
7. Flanagan, A., Fontanarosa, P., Bauchner, H. (2020). Preprints involving medical research - do the benefits overweight the challenges? *JAMA*, **324** (18), 1840-1843. <http://doi.org/10.1001/jama.2020.20674>
8. Fraser, N., Brierley, L., Dey, G., Polka, J., Palfy, M., Nanny, F., Coates, J. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS Biology*, 19(4): e3000959. <https://doi.org/10.1371/journal.pbio.3000959>
9. Ginsparg, P. (2021). Lessons from arXiv's 30th years of information sharing. *Nature Reviews Physics*, **3**, 602-603. <https://www.nature.com/articles/s42254-021-00360-z>
10. Pagliaro, M. (2021). Preprints in Chemistry: An exploratory analysis of differences with journal articles. *Publications*, **9** (1), 5. <https://doi.org/10.3390/publications9010005>

11. Polka, J. Strasser, C., Taraborelli, D. (2021). Shared Technology needs for preprints. (Version 1.1). *Zenodo*. <https://asapbio.org/shared-technology-needs-for-preprints>
12. Polka, J et el (2021). Preprints in motion: tracking changes between posting and journal publication. <https://www.biorxiv.org/content/10.1101/2021.02.20.432090v2>
13. Schonfeld, R., Rieger, O. (2020). Publishers invest in preprints. *Scholarly Kitchen*. <https://scholarlykitchen.sspnet.org/2020/05/27/publishers-invest-in-preprints/?informz=1>
14. Sever, R., Inglis, J., Bloom, T., Rawlinson, C., Krumholz, H., Ross, J. (2021). Pandemic preprints - a duty of responsible stewardship. *The BMJ Opinion*. <https://blogs.bmj.com/bmj/2021/04/27/pandemic-preprints-a-duty-of-responsible-stewardship/>
15. Widener, A. (2020). Pandemic puts preprints first. *C&EN*, **98** (22), 16-19. <https://cen.acs.org/policy/publishing/Pandemic-puts-preprints-first/98/i22>
16. Xie, B., Shen, Z., Wang K. (2021). Is preprint a future of science? A thirty-year journey of online preprint services. arXiv. <https://arxiv.org/pdf/2102.09066.pdf>

Acknowledgment

I would like to thank Digital Science for a complimentary trial of Dimensions Analytics.