bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Study of the Influence of Preprint in bioRχiv for Peer Review and Acceptance Time of PLOS ONE

Authors

Hiroyuki Tsunoda, Tsurumi University, Yokohama, Japan, tsunoda-h@tsurumi-u.ac.jp

https://orcid.org/0000-0002-6897-6611

Yuan Sun, National Institute of Informatics, Tokyo, Japan, yuan@nii.ac.jp

Masaki Nishizawa, National Institute of Informatics, Tokyo, Japan, nisizawa@nii.ac.jp

Xiaomin Liu, National Science Library, Chinese Academy of Sciences, Beijing, China, liuxm@mail.las.ac.cn

Kou Amano, National Institute for Materials Science, Tsukuba, Japan, AMANO.Kou@nims.go.jp

https://orcid.org/0000-0002-8079-4941

ABSTRACT

This study uses statistical methods to analyze the differences in peer review periods between the processes of preprint based on authors' order of posting published papers in a given journal. For this purpose, we developed a web crawler that downloaded metadata from bioRχiv and PLOS ONE. The average publication rate of papers posted in bioRχiv accounts for 40.67% of the total thus far. These papers were published in 1,626 academic journals. The journal that published most of these papers was PLOS ONE. Analysis of Peer Review and Acceptance Time (PRAT) of papers published in journals via preprints revealed the timing of posting papers related to these intervals. The median of PRAT of $f < s$ was 110 days, and the median of PRAT of $s < f$ was 139.50 days. If the date order of publishing a paper is the posted date of the preprint before the date of submission to the journal, the

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

PRAT tends to be shorter than when the date order of publishing a paper is the posted date of the preprint after the date of submission to the journal.

1 Introduction

Peer review is an important process that ensures the quality of a submitted paper [1][2][3][4]. However, this process has been widely criticized due to the delay introduced in publishing new findings [5][6][7][8][9]. Preprints have been attracting attention from researchers as a way of solving this issue [10]. It seems entirely possible, and perhaps inevitable, that the process of peer review will be shortened [11]. A preprint is a complete scientific manuscript uploaded by authors on a public server before formal review [12]. By posting papers on preprint servers, authors can publish draft manuscripts in a very short period after obtaining research results, and they can receive comments and advice on preprints from scientists worldwide. Preprints gained popularity around 30 years ago with the advent of arXiv, an open preprint server widely used in physics and mathematics [13][14][15][16]. Preprints were disseminated by authors over this period to bypass lengthy journal publication schedules [17][18]. Although scientists in other fields have been cautious in adopting the preprint mode of communication, the growth of preprint servers in chemistry and biology suggests that the acceptance of preprints in these may be mounting [19][20]. In recent years, some subject-specific preprint servers have been established. For instance, the website of bioRχiv, a biology preprint server, started in November 2013, it is becoming an increasingly popular platform for authors to share their work at a rate that increases almost monthly [21][22][23]. More than half of the works on the platform were published in journals with high citation frequency in the Web of Science Core Collection [24]. The journal with the most preprints is PLOS ONE [25]. It is a peer reviewed open access mega journal published by the Public Library of Science since 2006, and its

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

scope encompasses the results of scientific research from any scientific or medical discipline [26][27][28].

This study uses statistical methods to analyze the differences in peer review periods between processes of preprint based on authors' order of posting published papers in PLOS ONE.

2 Methods

First, this study investigated papers posted on bioRχiv. The web crawler visited every page on bioRχiv and downloaded metadata of the title, authors, published digital object identifier (DOI), posted date, and versions of a paper. If the paper was published, it would download the journal title and the published DOI. Next, the metadata was manually verified. The papers were matched with the PLOS ONE database using the keywords of published DOIs. Second, this study investigated articles published in PLOS ONE. The web crawler visited articles of PLOS ONE via bioRχiv and downloaded the received date, accepted date, and published date of these papers, following which metadata of bioRχiv and PLOS ONE was merged. Third, the Peer Review and Acceptance Time (PRAT) was defined as the difference in the numerical value between the accepted date and received date of published articles. The process of authors posting papers was divided into a pattern.

Variable declaration:

Four variables were declared as follows: Lowercase italic f is the date when the author should have the posted paper appear for the first time in the preprint server. Lowercase italic l is the date when the author should have the posted paper appear for the last time in the preprint server. Lowercase italic s is the date when the author should have submitted the paper to the journal. Lowercase italic a is the date to have a paper accepted by a journal.

Variable definition:

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The PRAT of a paper x is days of a minus s; it is defined as equation (1).

$$PRAT = a_x - s_x \qquad (1)$$

3 Results and Discussion

3.1 Posted Papers

The number of papers posted on bioRχiv is increasing rapidly every month (43,812 from November 2013 to February 28, 2019).

Recently, more than 2,000 papers were posted each month, of which 17,818 were published. Most papers are published within six months. Figure 1 shows the status of published and unpublished papers posted on bioRχiv.

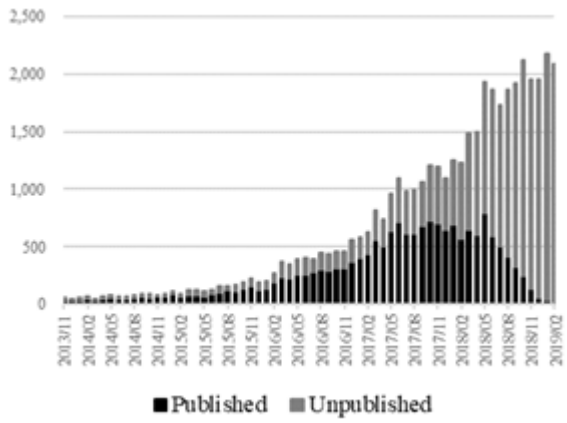

Figure 1: Published and Unpublished Papers Posted on bioRχiv

3.2 Published Journals

These papers were published in 1,626 academic journals. The journal (the number and share of these papers) that published most of these papers was PLOS ONE (902, 5.05%), followed by Scientific Reports – Nature (881, 4.94%), eLife (866, 4.86%), Nature Communications (611, 3.43%),

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Bioinformatics (509, 2.86%), PNAS (487, 2.73%), PLOS Computational Biology (375, 2.10%), PLOS Genetics (333, 1.87%), Nucleic Acids Research (285, 1.60%), Genetics (283, 1.59%), etc. Figure 2 shows the top 20 journals by papers in bioRχiv. This study focuses on PLOS ONE, because it is the journal that has published the most papers.

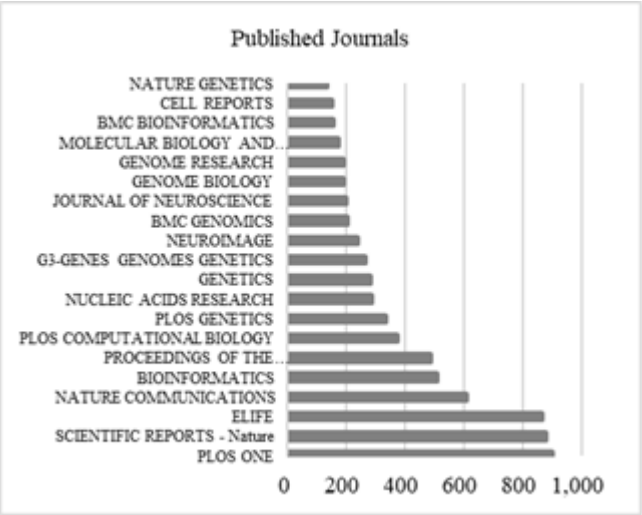

Figure 2: Top 20 Journals by Papers in bioRχiv

3.3 Analysis of Papers

These results analyzed the process of posting in PLOS ONE via bioRχiv. The average publication rate of papers posted in bioRχiv accounts for 40.67% of the total thus far. However, from 2013 to 2017, the rate was between 58% and 77%. The journal posting the most papers was PLOS ONE. PLOS ONE has published 902 papers from bioRχiv, of which the document type of two papers was a correction. The PRAT of 900 papers in PLOS ONE were calculated according to equation (1). The shortest PRAT (minimum) was 7 days, the longest (maximum) was 562 days, the middle (median) was 116 days, Quartile 1 of PRAT was 82 days, and Quartile 3 of PRAT was 166 days. Figure 3 shows the status of these PRAT. The PRAT were tested because the curve in Figure 3 does not look like a normal distribution. The p-value of the Shapiro-Wilk test was around zero (<3.73e-25) and

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

less than the 0.01 significance level. Consequently, according to the test, it cannot be stated that the analyzed data follow the normal distribution.

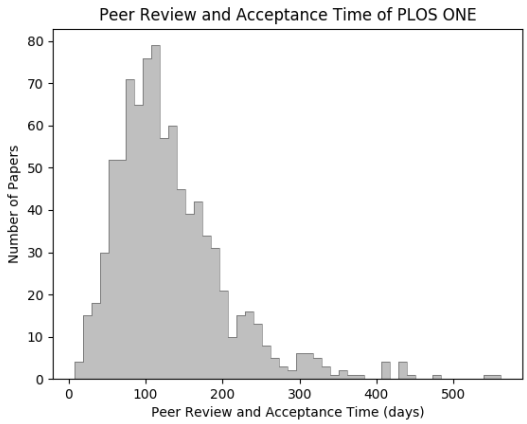

Figure 3: Peer Review and Acceptance Time of PLOS ONE

3.4 Process Order

If the difference in the numerical value between date α and date β of paper x is less than or equal to 7 days, but both days were defined as the same date, this case is expressed as α & β.

For example, when this condition is satisfied for the first posted date (f) and the submitted date (s) of paper x, it is defined as equation (2).

$$f_x - s_x \leqq 7 \Rightarrow f \,\&\, s \qquad (2)$$

If the difference in the numerical value between date α and date β of paper x is greater than 7 days, α and β were defined as a different date, and if date α is earlier than date β, the equation is expressed as $\alpha < \beta$. For example, when this condition is satisfied for the first posted date (f) and the submitted date (s), it is defined as equation (3).

$$f_x - s_x > 7 \Rightarrow f < s \qquad (3)$$

Process orders of 900 papers have been calculated in equations (2) and (3); these have been divided into 18 patterns based on variables of f, l, s, and a. These were merged into three groups. Table 1

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

shows 18 patterns and 3 groups.

Table 1: Process order of three groups

f & *s*		*f* < *s*		*s* < *f*	
Order	Papers	Order	Papers	Order	Papers
f & *l* & *s* < *a*	299	*f* & *l* < *s* < *a*	204	*s* < *a* < *f* & *l*	9
f & *s* < *a* < *l*	4	*f* < *l* & *s* < *a*	41	*s* < *a* < *f* < *l*	1
f & *s* < *l* & *a*	15	*f* < *l* < *s* < *a*	1	*s* < *f* < *a* < *l*	5
f & *s* < *l* < *a*	23	*f* < *l* < *s* < *a*	40	*s* < *f* & *l* & *a*	16
		f < *s* < *a* < *l*	5	*s* < *f* & *l* < *a*	168
		f < *s* < *l* & *a*	15	*s* < *f* < *l* & *a*	8
		f < *s* < *l* < *a*	27	*s* < *f* < *l* < *a*	19
Total:	341	Total:	333	Total:	226

The PRAT of PLOS ONE by every six months from July 2017 to Jun 2019 was 158, 171, 166, and 157 days. However, the PRAT of this study were pretty shorter than usual case. The median of PRAT of *f* < *s* was 110 days, and the median of PRAT of *s* < *f* was 139.50 days. The PRAT tends to be shorter than when the date order of publishing a paper is the posted date of the preprint after the date of submission to the journal.

Table 2: Statistical data of three groups

Groups	Number	Minimum	Quartile 1	Median	Quartile 3	Maximum
f & *s*	341	14	74.00	106.00	149.00	441
f < *s*	333	7	82.00	110.00	159.00	436
s < *f*	226	19	107.00	139.50	191.75	562

Figure 4 shows the status of three groups, which were tested because these did not appear to be a normal distribution. The p-values of the Shapiro-Wilk test of *f* & *s*, *f* < *s*, and *s* < *f* were around zero (<1.94e-14, <2.12e-13, and <6.95e-12) and less than the 0.01 significance level.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

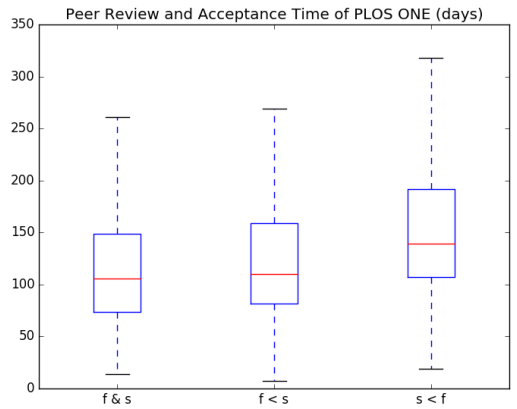

Figure 4: Peer Review and Acceptance Time of PLOS ONE (days)

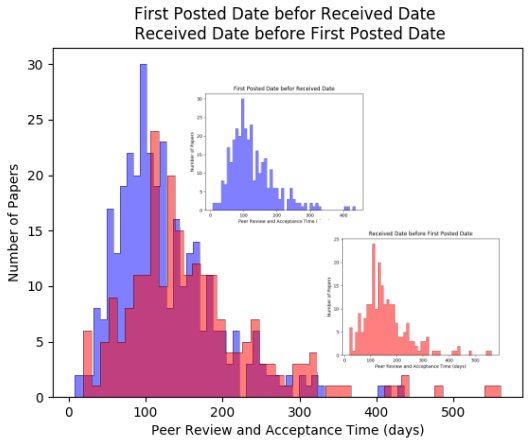

Figure 5: First Posted Date and Received Date

Between *f* < *s* and *s* < *f* was compared because the two groups of authors exhibited different behavior. If the behavior of an author is *f* < *s*, the author could expect to receive comments and advice on papers from scientists worldwide before his/her paper is submitted to journals. However, if the behavior of an author is *s* < *f*, he/she cannot expect to receive these. Figure 5 shows the PRAT of the first posted date (*f*) before the submitted date (*s*) (*f* < *s*) and the PRAT of the submitted date (*s*) before the first posted date (*f*) (*s* < *f*). If the median of *f* < *s* is compared with the median of *s* < *f*, the test must use the Mann-Whitney U test because *f* < *s* and *s* < *f* did not present as a normal distribution. The p-value of the Mann-Whitney U test was around zero (<2.07e-7) and less than the

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

0.01 significance level. It is clear that the median of $f < s$ is different from the median of $s < f$.

4 Conclusion

This study analyzed the PRAT of papers published in journals via preprints and it found that the timing of posting papers related to these intervals. If the date order of publishing a paper is the posted date of the preprint before the date of submission to the journal, the PRAT tends to be shorter than the date order of publishing a paper whose posted date of preprint is after the date of submission to the journal.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number JP19K12707, JP18K11597, and ROIS NII Open Collaborative Research 2019-(19FS02).

REFERENCES

[1] Catherine Mangan , Mark Pietroni , Denise Porter, 2016. “Being brave”: A case study of how an innovative peer review approach led to service improvement, J. Integr. Care., 24, 4, 201-213.

DOI: https://doi.org/10.1108/JICA-06-2016-0021

[2] Joshua Russell, and Jack Armitage, 2006, Peer review effectiveness: An analysis of potential loopholes within the USA Peer Review Program, Manage. Auditing J., 21, 1, 46-62.

DOI: https://doi.org/10.1108/02686900610634757

[3] Brian Paltridge, 2013. Learning to review submissions to peer reviewed journals: how do they do it?, Int. J. Res. Dev., 4, 1, 6-18

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DOI: https://doi.org/10.1108/IJRD-07-2013-0011

[4] Fytton Rowland, 2002. The peer-review process, Learn. Publ., 15, 247-258.

DOI: https://doi:10.1087/095315102760319206

[5] Jacalyn Kelly, Tara Sadeghieh, and Khosrow Adeli, 2014. Peer review in scientific publications: Benefits, critiques, & a survival guide, EJIFCC. 25, 3, 227-243.

[6] Kendall Powell, 2016. Does it take too long to publish research?, Nature, 530, 148-151.

DOI: https://doi.org/10.1038/530148a

[7] Martin Raff, Alexander Johnson, and Peter Walter, 2008. Painful Publishing. Science, 321, 5885, 36-36.

DOI: https://doi.org/10.1126/science.321.5885.36a

[8] Solomon H. Snyder, 2013. Science interminable: Blame Ben?, Proc Natl Acad Sci U S A. 110, 7, 2428-2429.

DOI: https://doi.org/10.1073/pnas.201300924

[9] Ronald D. Vale, 2015. Accelerating scientific publication in biology, Proc Natl Acad Sci U S A. 112, 44, 13439-13446.

DOI: https://doi.org/10.1073/pnas.1511912112

[10] Sarvenaz Sarabipour, Humberto J. Debat, Edward Emmott, Steven J. Burgess, Benjamin Schwessinger, and Zach Hensel, 2019. On the value of preprints: An early career researcher perspective, PLoS Biol. 17, 2, 1-12.

DOI: https://doi.org/10.1371/journal.pbio.3000151.

[11] Cecelia Brown, 2001. The e-volution of preprints in the scholarly communication of physicists and astronomers, J. Am. Soc. Inf. Sci. Technol. 52, 3, 187-200.

DOI: https://doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASI1586> 3.0.CO;2-D

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[12] Jeremy M. Berg, Needhi Bhalla, Philip E. Bourne, Martin Chalfie, David G. Drubin, James S. Fraser, Carol W. Greider, Michael Hendricks, Chonnettia Jones, Robert Kiley, Susan King, Marc W. Kirschner, Harlan M. Krumholz, Ruth Lehmann, Maria Leptin, Bernd Pulverer, Brooke Rosenzweig, John E. Spiro, Michael Stebbins, Carly Strasser, Sowmya Swaminathan, Paul Turner, Ronald D. Vale, K. VijayRaghavan, and Cynthia Wolberger, 2016. Preprints for the liciences, Science 352, 6288, 899-901.
DOI: https://doi.org/10.1126/science.aaf9133.

[13] Paul Ginsparg, 2011. ArXiv at 20, Nature 476, 145-147.
DOI: https://doi.org/10.1038/476145a

[14] Stacy S. Wykle, 2015. Enclaves of anarchy: Preprint sharing, 1940-1990, Proc. Am. Soc. Inf. Sci. Technol. 51, 1, 1-15.
DOI: https://doi.org/10.1002/meet.2014.14505101036

[15] Xuemei Li, Mike Thelwall, and Kayvan Kousha, 2015. The role of arXiv, RePEc, SSRN and PMC in formal scholarly communication, Aslib J. Inform. Manag. 67, 6, 614-635.
DOI: https://doi.org/10.1108/AJIM-03-2015-0049

[16] Gerry McKiernan, 2000. arXiv.org: the Los Alamos National Laboratory e-print server, Int. J. Grey Lit. 1, 3, 127-138.
DOI: https://doi.org/10.1108/14666180010345564

[17] Patricia A. Kreitz, Louise H. Addis, Hrvoje Galic, and Thomas Johnson, 1997. The virtual library in action: Collaborative international control of high-energy physics pre-prints, Publ. Res. Q. 13, 24-32.
DOI: https://doi.org/10.1007/s12109-997-0016-x

[18] Michael A. Johansson, Nicholas G. Reich, Lauren A. Meyers, and Mark Lipsitch, 2018.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Preprints: An underutilized mechanism to accelerate outbreak science, PLoS Med. 15, 4, 4-8.

DOI: https://doi.org/10.1371/journal.pmed.1002549

[19] Cecelia Brown, 2003. The role of electronic preprints in chemical communication: Analysis of citation, usage, and acceptance in the journal literature, J. Am. Soc. Inf. Sci. Technol. 54, 5, 362-371.

DOI: https://doi.org/10.1002/asi.10223

[20] Leslie M. Loew, 2016. Peer review and bioRxiv, Biophys. J. 111, 3, 1-2.

DOI: https://doi.org/10.1016/j.bpj.2016.06.035

[21] Richard J. Abdill and Ran Blekhman, 2019. Meta-research: Tracking the popularity and outcomes of all biorxiv preprints, Elife 8, 8, e45133.

DOI: https://doi.org/10.7554/eLife.45133.001

[22] Leslie M. Loew, 2016. Peer Review and bioRxiv, Biophys J. 111, 3, E01–E02.

DOI: https://doi.org/10.1016/j.bpj.2016.06.035

[23] Marina Picciotto, 2017. Direct Submissions from bioRxiv, J Neurosci. 37, 2, 237.

DOI: https://doi.org/10.1523/JNEUROSCI.3812-16.2016

[24] Hiroyuki Tsunoda, Yuan Sun, Masaki Nishizawa, Xiaomin Liu, and Kou Amano, 2019. An analysis of published journals for papers posted on bioRXiv, Proc. Assoc. Inf. Sci. Technol. 56, 1, 783-784.

DOI: https://doi.org/10.1002/pra2.175

[25] Kent R. Anderson, 2019. bioRxiv: Trends and analysis of five years of preprints, Learn. Publ.

DOI: https://doi.org/10.1002/leap.1265

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.28.920546; this version posted February 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[26] Joel Cummings, 2007. PLoS One, Reference Rev. 21, 7, 41-42.

DOI: https://doi.org/10.1108/09504120710821767

[27] Valerie Spezi, Simon Wakeling, Stephen Pinfield, Claire Creaser, Jenny Fry, Peter Willett, 2017. Open-access mega-journals: The future of scholarly communication or academic dumping ground? A review, J. Documentation 73, 2, 263-283.

DOI: https://doi.org/10.1108/JD-06-2016-0082

[28] Simon Wakeling, Valerie Spezi, Jenny Fry, Claire Creaser, Stephen Pinfield and Peter Willett, 2019. Academic communities: The role of journals and open-access mega-journals in scholarly communication J. Documentation 75, 1, 120-139.

DOI: https://doi.org/10.1108/JD-05-2018-0067