

The Spread and the Future of Preprint Servers that Support the Quick Sharing of Research Results

Mi-Hwan Hyun^a, Tae-Sul Seo^{b*}, ^aSenior researcher, Korea Institute of Science & Technology Information, Yuseong-gu, Korea, 34141, South Korea, ^bDirector, Korea Institute of Science & Technology Information, Yuseong-gu, Korea, 34141, South Korea, Email: mhhyun@kisti.re.kr, tsseo@kisti.re.kr

In conventional academic communication environments, Preprint servers have emerged as a solution to speeding up the publishing process. This study investigates current Preprint servers and their outlook through an analysis of overseas Preprint server cases. This study classified the publication process in a traditional academic publishing environment into six stages. In the academic publishing stage, the Preprint and Postprint versions of papers were self-archived. We examined and defined the concept of Preprint servers. This study also performed a Preprint server status and case analysis. The proliferation of open science, which shares the entire research process and results, is affecting academic publishing. As of May 2019, arXiv – the first preprint server, launched in 1991 – has over 1.5 million papers. Additionally, a total of 2.2 million papers are being serviced by 29 Preprint servers worldwide. Under the leadership of COS (Centre for Open Science) – an open science organization – various Preprint servers are being created according to different countries, languages, and academic fields. Researchers are increasingly depositing their papers at Preprint servers so that their research can be immediately published. Preprint servers are attracting attention as new alternatives to traditional academic publishing. The results of this study can be used by information service providers to formulate service strategies for researchers to share and spread their findings. This study additionally provides information for developing business strategies to academic publishers and database builders.

Key words: *Preprint Server, Preprint, Open Science, Scholarly Communication, Peer Review, Academic Publishing.*



Introduction

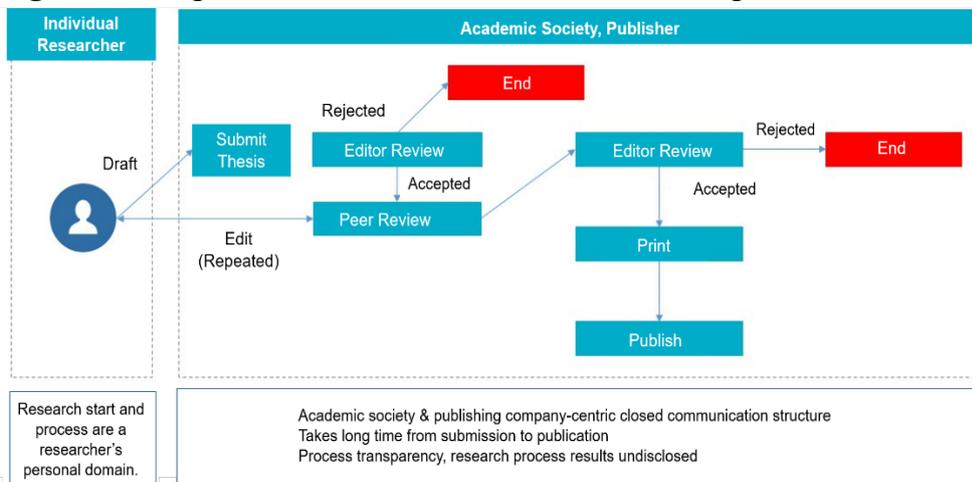
As scientists disclose their research results via R&D papers, peer review has been widely perceived as the most reliable way to assess the quality of Preprints. In fact, peer review has been the basis of many traditional academic communication environments. However, with people's growing interest 'open science', open peer review journals that allow researchers to freely participate in publication and peer review, such as PeerJ, have emerged. Furthermore, open-access Preprint servers are becoming more common. This kind of movement primarily focuses on quantitative assessment, such as 'impact factor', and the idea that it takes too long for researchers to share their research. As it is becoming more difficult to obtain an R&D subsidy, many researchers submit their papers to journals with a high impact factor. Generally, it takes between 6-12 months for a paper to be published from the initial submission date. If a researcher submits a paper to a foreign journal, and it successfully passes peer review, it can take between 3-4 months to be published. If rejected, the paper needs to be submitted to another journal. If the publication is further delayed, the paper becomes 'not new'. Even in a digital era where information is shared in real time, it can take over a year for papers to come out of research labs. While researchers want to immediately share their research results, the papers must go through multiple review processes until they can actually be published. In such conventional academic communication environments, Preprint servers have emerged as a new solution to speed up the publishing process. This study investigates current Preprint servers and their outlook through an analysis of overseas Preprint server cases and suggests a plan to utilize Preprint servers as paradigms in academic communication environments.

Materials and Methods

Academic Publishing Process

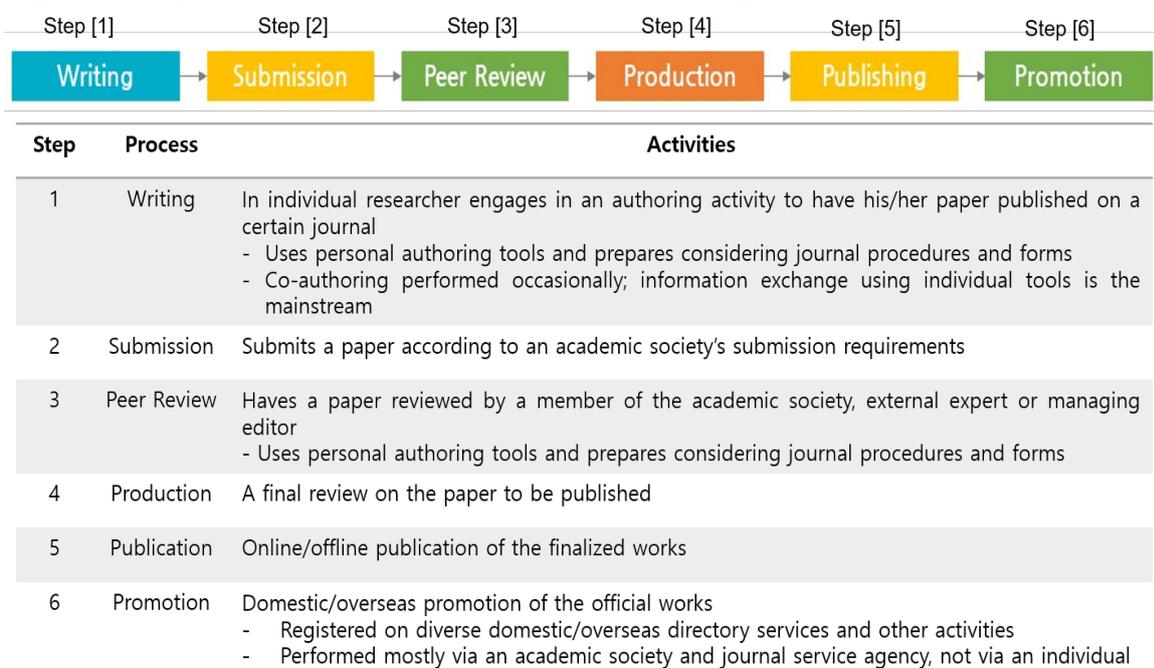
As shown in Figure 1, in conventional academic communication environments, a paper must be peer-reviewed before it is officially published in an academic journal. During this process, an editor assesses the paper based on the advice of referees and identifies problems in the research method, assumptions, and conclusions. In general, a paper is not published until such findings are properly assessed.

Figure 1. Design of a Conventional Academic Publishing Process



The ‘academic publishing process’ refers to the series of actions between the production of a paper and its spread to users through web services. This process can be divided into three categories: Creation, Archiving, and Promotion. The steps of the conventional academic publishing process, and the research activities included in each step, are described in Figure 2.

Figure 2. Steps of the Conventional Academic Publishing Model



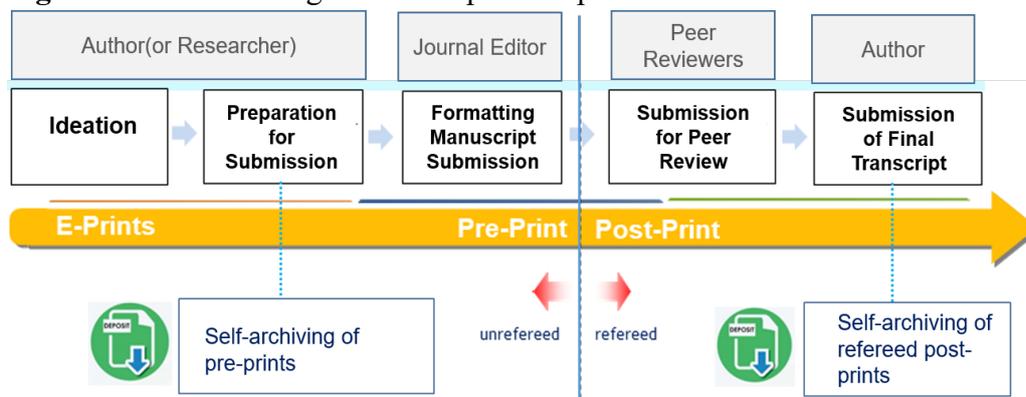
During the academic publishing process, peer review often takes a long time. Therefore, self-archiving is granted to Preprint servers so that researchers may publish their papers as

‘Preprint’ – or, before peer review. Accordingly, other scientists can express their opinions on the research results. In other words, users of Preprint servers should be aware that there may be errors in the listed papers, and that they may include unaccepted or unapproved information.

Preprint Server Definition

The word ‘Preprint’ has many meanings, including draft, incomplete version, or final version. However, there is a slight difference in this definition between the academic circle and the publishing industry. In the academic world, where peer review is perceived as critical paper editing, a Preprint refers to a draft before peer review. In contrast, in the publishing industry, a Preprint refers to a peer-reviewed paper whose publication has been finalized, even if a published form is not included. Usually, a Preprint refers to an unrefereed paper. According to Sherpa/RoMEO – which guides copyright information on open-access journals – Preprints can be problematic in copyright transfer agreements if they are used under a different name (Gadd et al., 2003). Therefore, a paper before peer review is defined as ‘pre-print’, while a paper after peer review is categorized as ‘post-print’ (Harnad, 2003). A post-print is a paper which did not go through an editing process for journal publication. In other words, a post-print’s publication is finalized after peer review and before a publishing form is reflected. The papers, therefore, are the same as the published version but different to the edited version. However, because a Preprint is a paper before peer review, content errors can exist. In other words, a Preprint might be a paper whose journal is undecided. A public server in which an author is able to archive such Preprint papers on their own is called a ‘Preprint server’. Figure 3 illustrates the concept of Preprints through self-archiving during the academic publishing stage.

Figure 3. Self-archiving and Concept of Preprint Servers



Once an academic paper is uploaded to a Preprint server, the Preprint server’s reviewer checks if the paper’s content is acceptable for a scientific environment. Then, the paper is published within a week through an online platform, generally without being peer reviewed,



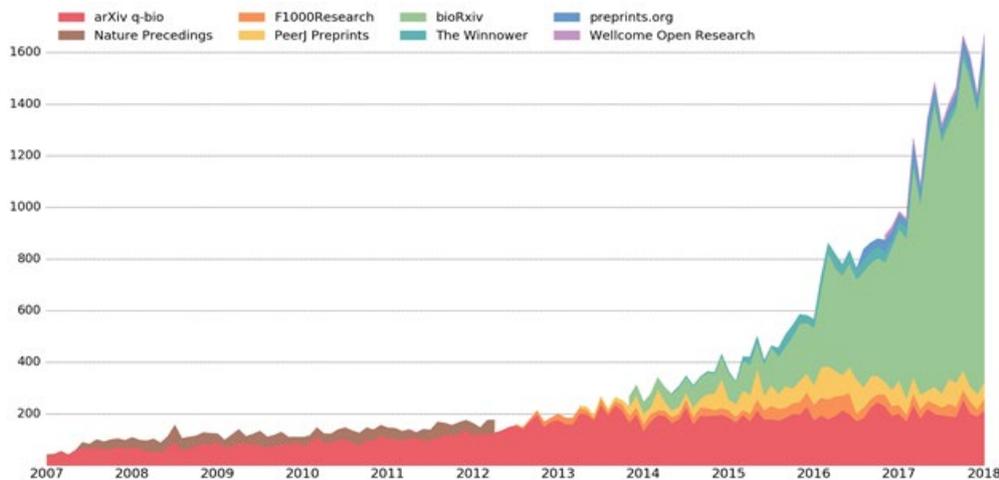
where it can be freely accessed by anyone. The edited version of the paper can be uploaded, even after it has been published through an online platform. In this case, previous versions are not deleted, and papers with multiple versions are accessible through online services. In a Preprint server, papers are self-archived before they are submitted to a journal to ensure that journal copyright laws are not infringed. According to journal service policies, some journals do not permit Preprint papers to be self-archived, which is not a legal infringement of copyright agreements. Additionally, there is a difference in self-archiving according to the research field; in physical science (physics, computer science, etc.), self-archiving is generally accepted, while in biology it is denied (Antelman, 2006). Self-archiving without following journal policies can be a violation of law. Since journal/publishing companies apply different copyright policies to post-print self-archiving, it is necessary to check the individual journal policies before self-archiving.

Preprint Server Status

Preprint servers originate from a program designed by NIH (National Institutes of Health) in 1961 to open Preprint papers in biology. However, the program was stopped in 1967 (Cobb, 2017). Since then, the Preprint server ‘arXiv’, a repository for physics, was created in 1991. Currently, arXiv is a repository for computer science, mathematics, and many other fields, and, as of May 2019, more than 1.5 million papers are available through the website. arXiv(<https://arxiv.org/>) is the most popular Preprint server, operating 17 mirror sites in 16 countries. Its services are provided to nearly 100,000 users every day in the U.S. alone. Post-print papers are mostly found in arXiv. On the contrary, in RePEc (Research Papers in Economics), a repository developed by SSRN (Social Science Research Network) in 1997, papers in all versions are available. In 2013, bioRxiv (<https://www.biorxiv.org/>) in biology and PeerJ Preprints (<https://peerj.com/preprints/>) in computer science launched services (Desjardins-Proulx, 2013). Open source-based preprint servers are primarily distributed by COS (Centre for Open Science), a company established in multiple diverse academic fields. Preprint server data from around the globe is collected by OSF preprint sites. OSF additionally provides indexing services. As of May 2019, a total of 29 OSF Preprint servers are in operation, and 2.2 million papers are registered and available to the public (<https://osf.io/preprints/>). According to the monthly preprint production statistics announced by PrePubMed (PrePubMed, 2016), the number of papers skyrocketed since mid-2010 when the bioengineering prurient ‘bioRxiv’ emerged. In particular, the total amount of papers registered over the past 2-3 years increased more than 5 times. PrePubMed provides preprint indexing services for arXiv, bioRxiv, PeerJ Preprints, F1000Research, preprints.org, Winnower, and Welcome Open Research. PrePubMed does not save these news articles, however, it does link information to the original article on the website (Mito, 2019). According to statistics illustrated in Figure 4 below, new publication services such as bioRxiv, preprints.org, and Welcome Open Research, in addition to conventional services

such as arXiv, PeerJ preprints have gradually increased since 2014. When compared to conventional preprint servers, a difference of 8 times or more was found, confirming growth (Lin, 2018).

Figure 4. Monthly Paper Publications by Preprint Service, source: PrePubMed



As shown in Figure 5, the DOI-linked growth rate of copyrighted works can be checked. According to Jennifer (2018), preprints are 10 times greater than conventional academic journal articles in terms of growth rate (Balaji and Dhanamjaya, 2019).

Figure 5. Preprint Statistics Registered with Crossref

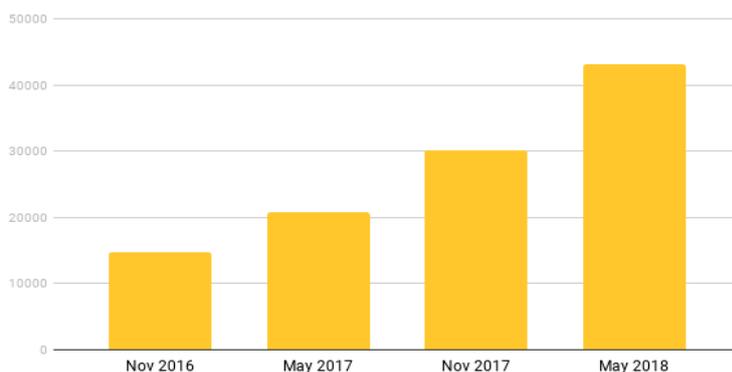


Figure 5 reveals 44,388 copyrighted works of preprints registered as of May 24, 2018. The volume and growth of preprints can be estimated by considering the total number of preprints (65,223 works). Recently, the Centre for Open Science (COS), which has promoted an open-science community through the open-source tool Open Science Framework (OSF), opened an open source-based preprint server named 'OSF Preprints' in 2016. The COS encourages researchers to collaborate and share their work (Foster and & Deardorff, 2017). In OSF



Preprints, sources are open so that anyone is able to use the services freely. An auto-indexing function is provided to search engines such as Google Scholar in case a preprint server is developed using OSF. Additionally, OSF offers other functions such as an integrated search function, provided by arXiv, BioRxiv and PeerJ. The OSF Preprint-based services are agriXiv (agriculture), PaleorXiv (palaeontology), and ECSasXiv (electrical science) in 2016, and SportRxiv (sports), Thesis Commons (thesis), MedArxiv (medicine), and EarthArxiv (earth science) in 2017. Additionally, preprint servers have been actively developed according to language and country; ArabiXiv (Arabic speaking countries), FrenXiv (French speaking countries), SciELO Preprints (Latin America, South Africa, Iberian Peninsula) and INArxiv (Indonesia) are all based on OSF Preprints.

Preprint Server and Academic Publishing Model

Open Science and Preprint Server roles

Open Science is a movement that helps both researchers and citizens participate in scientific research (Schmidt, 2016). Researchers are able to let people know when their research begins, set a research topic, and launch collaboration with diverse co-researchers. Additionally, since the research process and data is open, research transparency and quality is maintained. Open science also enables results to be published through traditional publishing companies and their quick sharing services via an open ecosystem such as web. Preprints play the role of a channel, through which the research processes and results can be openly promoted and shared. In COS, services are provided that allow individual researchers to directly handle the start of their research, their schedule, data disclosure, communication through social media, archive development, and publication. In fact, these services refer to the beginning of new era, in which researchers conduct research in open environment, provide open feedback, and share results directly or via a preprint service. In particular, COS' preprint services have made a contribution to the expansion of language and country and led to changes in academic communication ecosystems.

Advantages of Preprint Server

It is hard to say that a Preprint can solve all of the problems of current academic publishing environments. Even so, the diverse advantages of Preprint servers meet social and academic needs, where opening and sharing is emphasized. Firstly, Preprint enable the fast distribution of research results. The latest paper drafts or data can be quickly distributed without have to go through official, and often time-consuming, publishing procedures. Secondly, the effects of research topic pre-emption are expected; Preprint servers often reveal paper-listed dates in case lots of researchers from around the globe are researching the same topic. Therefore, Preprint servers are able prioritize research ideation and discovery. Thirdly, the level of



perfection of the research can be increased through fast and diverse feedback. A Preprint enables researchers to get expert opinions in a quick and broad fashion, offering them an opportunity to edit and improve their paper before publication. In other words, since the quality of the papers can be improved overtime, there is an increased possibility that these papers will eventually be published in a journal. Fourthly, research promotion is fast. The papers listed on a Preprint server more often than not receive DOIs, and thus have a high chance of being quoted. Additionally, since papers can be quickly spread through social media, authors can easily get attention from research and academic circles. Fifthly, the paper's validity is expected. In general, agencies granting R&D subsidies request supporting documents on the researchers' latest research activities. A Preprint server helps researchers promptly respond to such requests. Sixthly, there are increased chances of researchers being chosen to be published in journals. If thesis contributors are searched by diverse journals, academic societies, and particular publishing companies or institutes, a Preprint paper's chance of chosen by a journal would increase. This advantage encourages researchers to contribute to Preprint servers more, increasing their chances of being chosen. Lastly, Preprints can make contributions to academic development, which is often ultimate goal of research activities. Even though research results cannot contribute to journals as formal papers because they, more often than not, do not have clear conclusions, data is shared, and positive and negative results are disclosed, in turn enhancing transparency. Ultimately, then, researchers can be remembered as contributors to academic development.

Outlook of Preprint Server

A preprint server can be developed by either an R&D institute or a researcher. In particular, if R&D authoring tools that enable open science become more common, researchers would be able to archive on their own work and the research process would become more transparent than conventional repository-based methods. Such self-archiving is the most efficient method for spreading research results, in that anyone can access self-archived work regardless of certain journal or publishing company services. Additionally, researchers are able to see the number of papers used in person and check research utility. Above all, Preprint servers can dramatically reduce the time needed to spread research results. It is more meaningful in that researchers can quickly disclose their research plans and results to research communities. Despite these advantages, Preprint servers still have limitations; Preprint servers are not widely used by researchers due to a lack of awareness and concern about the quality of Preprint works. However, case studies have confirmed that such issues can be improved by applying diverse service systems. Furthermore, even though there are some doubts about this new communication system because it stands against conventional academic communication approaches, many researchers have experienced that it is possible to change academic communication environments if they choose these new communication methods and develop open research environments (as shown in the arXiv case). A Preprint server is not just a



means to help researchers' activities, but a new movement which can represent an open academic ecosystem. Therefore, researchers need to voluntarily use Preprint servers and actively build open research environments. Furthermore, research outcome distribution & spread agencies have to be more active in responding to preprint services which can be a new solution to conventional academic publishing industries in the new movement titled 'open science'.

Conclusion

Peer review is still an essential step in the academic publishing process. It is a critical process that filters defective or low-quality papers. Furthermore, it is a safety net which can minimize scientific errors. Therefore, it is still controversial to trust Preprints posted on online archives, which have not gone through peer review. Even so, Preprint servers enable researchers to share their research results with peers immediately, get feedback from various researchers, and correct their errors through quick publication. Therefore, once a decent environment in which many researchers are involved is built, current concern about the reliability of preprints can be eliminated. Preprints and post-prints are not a matter of right or wrong – they need to coexist as complementary pairs. Preprints and post-prints can co-prosper by acquiring positive aspects from each other. More fundamentally, researchers and academic societies need to change – after all, it is researchers who produce academic research results and consume academic research publications. New technologies or concepts such as preprint, open access, and open science cannot be more than assistance tools. The current problems in academic publication caused by these new tools can be solved partially, but not completely. If researchers work hard to overcome current limitations and problems and are not afraid of new attempts and challenges for open and better research environments, it will be possible to build open research environments where researchers, not publishing companies, lead academic publication.

REFERENCES

- Antelman, K. (2006). Self-archiving practice and the influence of publisher policies in the social sciences. *Learned Publishing*, 19(2): 85-95. <https://onlinelibrary.wiley.com/doi/abs/10.1087/095315106776387011>
- Balaji, B. P. & Dhanamjaya, M. (2019). Preprints in Scholarly Communication: Re-Imagining Metrics and Infrastructures. *Publications*, 7(1): 6. <https://www.mdpi.com/2304-6775/7/1/6>
- Cobb, M. (2017). The prehistory of biology preprints: A forgotten experiment from the 1960s. *PLoS biology*, 15(11): e2003995. <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2003995>
- Desjardins-Proulx, P. (2013). The case for open preprints in biology. *PLoS Biology*, 11(5): e1001563. <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001563>
- Foster, E. D. & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association: JMLA*, 2017, 105(2): 203. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5370619/>
- Gadd, E., Oppenheim, C. & Proberts, S. (2003). RoMEO Studies 1: The impact of copyright ownership on academic author self-archiving. *Journal of Documentation*, 59(3): 243-277. <https://www.emerald.com/insight/content/doi/10.1108/00220410310698239/full/html>
- Gadd, E., Oppenheim, C. & Proberts, S. (2003). RoMEO Studies 2: How academics want to protect their open-access research papers. *Journal of information science*, 29(5): 333-356. <https://journals.sagepub.com/doi/abs/10.1177/01655515030295002>
- Harnad, S. (2003). Electronic preprints and postprints. In: *Encyclopedia of library and information science*. Marcel Dekker. <http://cogprints.org/3019/>
https://www.bioblast.at/index.php/MitoPedia:_Preprints
- Lin, J. (2018). Preprints growth rate ten times higher than journal articles. *Available at URL [Accessed 4 Feb. 2019]*. <https://www.crossref.org/blog/preprints-growth-rate-ten-times-higher-than-journal-article>
- Mito, (2019). MitoPedia: Preprints [Internet] ; [updated 19 April 2019]. Available from: PrePubmed , (2016). PrePubmed: Pubmed for preprints [Internet] ; c2016 [updated 2018 Dec]. Available from: http://www.prepubmed.org/monthly_stats/
- Schmidt, B. (2016). Stepping up open science training for European research. *Publications*, 4(2): 16. <https://www.mdpi.com/2304-6775/4/2/16>