

## Preprints and preprint servers as academic communication tools

### Preprints y servidores de preprints como herramientas para la comunicación académica

Ernesto Galbán Rodríguez<sup>1\*</sup> <http://orcid.org/0000-0002-2716-4799>

<sup>1</sup> Elfos Scientiae Publisher. Center for Genetic Engineering and Biotechnology. La Habana, Cuba.

\*Autor para la correspondencia. Correo electrónico: [ernesto.galban@cigb.edu.cu](mailto:ernesto.galban@cigb.edu.cu)

#### ABSTRACT

Preprints and preprint servers comprise the articulation of the ultimate and most proximal node to publishing the scientific results in academic journals. Therefore, in this review, the concept, development, advantages and limitations of preprints and preprint servers are analyzed, attending to their main function as publicly available repositories of manuscripts on the way to getting published. Moreover, an analysis of the motor forces contributing to their establishment to communicate research results among scientific communities is given, with their classification (journal, non-journal and mixed servers; subject repositories), as well as debate on the most successful (arXiv) and debated (Biology, Chemistry) servers, preprint formats, and their relationship with information phenomena such as open access, open archiving, digital information certification, information retrieval, and the added value through immediacy in availability and citation in comparison with published articles. Examples of their integration with ongoing communicational processes are discussed, such as migration of editors from journals to preprint servers to channel up manuscripts, open peer review strategies and scientific community engagement. A list of the most relevant preprint servers until 2018, their characteristics, general stats of their preprints and citation counts in Scopus is included.

**Key words:** Scientific communication; journal publishing; preprints; preprint servers; open access; open archiving.

## **RESUMEN**

Los manuscritos pre-publicación (o preprints) articulan el último eslabón previo y más próximo a la publicación de los resultados investigativos en las revistas académicas. En este artículo de revisión se analiza el concepto, el desarrollo, las ventajas y las limitaciones de los preprints y los servidores de preprints. Esto atendiendo a su función principal como repositorios públicos disponibles de manuscritos en vías de publicación. Además, se abunda sobre los motivos de su surgimiento y de su establecimiento para la comunicación de los resultados de investigación entre comunidades científicas, su clasificación (servidores de revistas científicas, no asociados a revistas científicas y servidores mixtos; repositorios temáticos). Se incluyen los debates sobre el servidor más famoso, arXiv, y los más debatidos en Biología (biorXiv) y en Química; sobre los formatos de los preprints y su interrelación con los fenómenos informacionales del acceso abierto, el archivo abierto, la certificación de información digital, la recuperación de información y el valor agregado de la inmediatez de disponibilidad y citación en comparación con los artículos publicados. También se incluyen elementos sobre su integración con los procesos comunicacionales en ejecución, entre ellos la migración de editores de las revistas a los servidores de preprints para canalizarlos hacia la publicación final, las estrategias de evaluación por pares abierta y la vinculación con las comunidades científicas. Se incluye además una lista de los servidores de preprints posicionados hasta el año 2018, con la citación de sus preprints en la base de datos Scopus, así como algunas de sus características.

**Palabras clave:** Comunicación científica; publicación de revistas; preprints; servidores de preprints; acceso abierto; archivo abierto.

Recibido: 26/10/2018.

Aprobado: 24/01/2019.

## **INTRODUCTION**

Preprints, prepublication drafts or manuscripts prior to their submission to a journal, have emerged as publishable (to be made public) digital material of increasing value for research communities and scientific systems. They are a natural step on the structuring of a scientific paper intended for validation (peer-review) and final publication in a scientific journal (the scientific record). But they have recently expanded into the node most proximal to scientific

experimentation and its structuring for technical communication in digital format prior to publishing in academic journals.

The paradigm of this movement has been the arXiv preprint server as early as in 1991, formerly a centralized automated repository and alerting system to send full texts on demand, which later became the leading example of preprint servers worldwide.<sup>(1)</sup> The implementation of this model of publication (i.e., ·to make public·, open to public discussion) in physics, astronomy and mathematics, further expanded to other disciplines and areas of scientific investigation, with a more recent diversification in the number and types of electronic systems and digital platforms.

Previous analyses have focused on the history, development and use of preprints and preprint servers, and particular aspects of preprints depending on the field of application have been analyzed. However, due to their expansion either in the number of preprints, editorial policies and platforms available in recent years, an integrative update on preprints and preprint servers was still required. Therefore, this review was aimed to update the information on preprint and preprint servers development and applications in the scientific publishing ecosystem, together with the different modalities of preprint servers and platforms, and statistics of preprints hosted on the most prestigious preprint servers at the end of 2018. Moreover, due to their importance as mediator in the scientific communication process, recent debates on preprint value, use and integration during manuscript submission to journals and the development of editorial policies for their acceptance were included.

## **METHODS**

For this purpose, the Scopus and Web of Science bibliographic databases were searched for "preprint server" in the title and keywords as on November 19th, 2018. The abstract field was omitted since there seems to be common practice in physics and other subject areas to declare the link to the preprint when present in the paper abstract. Then, pertinent records were selected, and their respective documents downloaded and subjected to classic document analysis. They included editorials and comments articles which database records were less descriptive. Cited references were inspected in the documents retrieved and downloaded for complementation. Since debate information is present in specialized web platforms other than journals, the database search was complemented with an open search for preprints on the Google search engine on the same date as databases and the first 200 results were analyzed.

Statistics on the number of preprints and their citations in the Scopus database were computed. Data on preprints posted were aggregated from each preprint server. Additionally, the Bielefeld Academic Search Engine (BASE, Bielefeld University Library; <http://www.base-search.net>) was analyzed as on December 6th, 2018, for document types hosted. Due to its largest coverage, the Scopus database was searched for citations of preprints and their servers by using the following search strategy: "WEBSITE (server website domain) OR (SRCTYPE (j) AND REFSRCTITLE (server) AND NOT REFTITLE (server))".

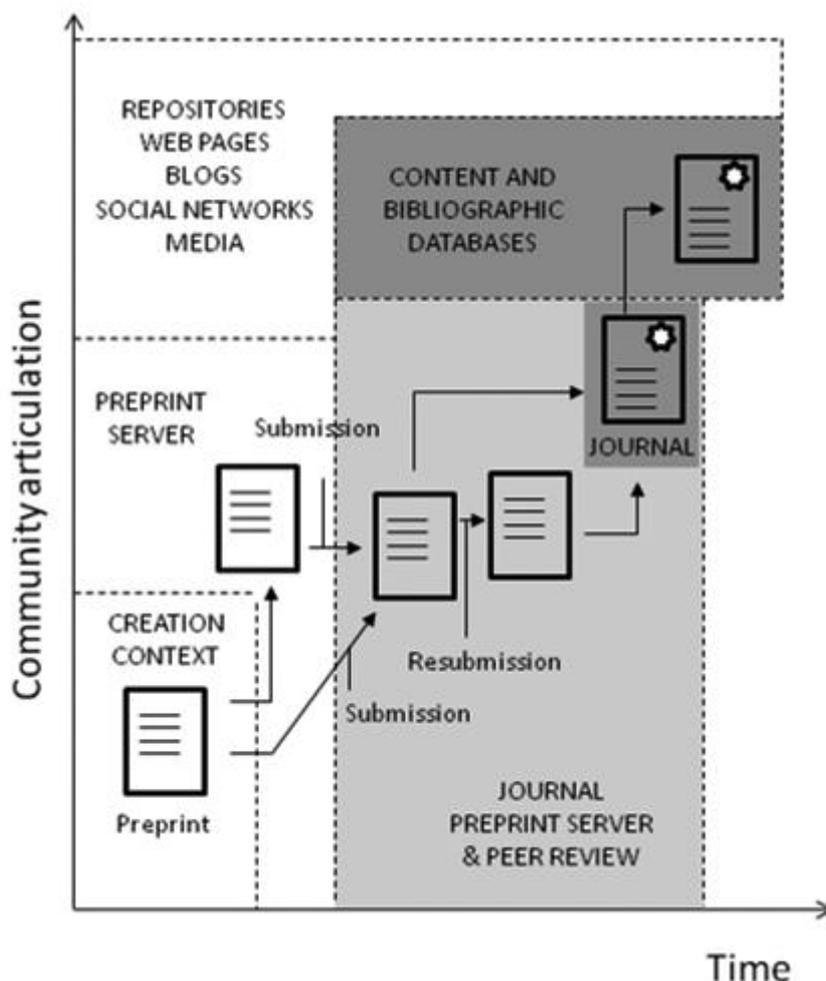
Other information pertinent to the academic publishing environment and practices was provided based on its relationship to the subject under discussion. Journal information was prioritized over open information sources, unless the latter provided unreported data and information, and information was aggregated following the timeline of preprint and preprint servers development, together with subject pertinence. Concepts and views were progressively contrasted.

## **RESULTS**

### **The emergence of the preprint model in digital format**

The preprint movement evolved from the merge of four different initiatives: 1) a natural step on the way towards the inception of a scientific paper<sup>(2)</sup> since the pre-digital era, in the search for a common approval by colleagues before sending a manuscript to a journal; 2) the evolution and improvements of publication standards,<sup>(3)</sup> which rules permeates the very inception moment of research and its subsequent publication phase; 3) the open access to the scientific information, both as information access<sup>(4)</sup> and archiving;<sup>(5)</sup> and 4) the need for immediacy in the communication of science, to face publication delays<sup>(6)</sup> and claim for authority.<sup>(7)</sup> The latter relates to the two conditions identified as critical for participatory informational approaches: the possession of information and of the instruments necessary to deploy it.<sup>(8)</sup> Moreover, when making available the tools (i.e., preprint servers or content repositories), results are communicated to the community of interest, prior to its appropriation as valid information once corroborated by open discussion or peer-review for journal publication. Both processes, community communication and journal publication, occur under specific standards for science communication and community archiving practices. They are consistent with the community phenomenon as the fourth archival paradigm enunciated by Terry Cook,<sup>(9)</sup> as the constant and unrestrained record of scientific production which is immediately put into public scrutiny of a context wider than the producing one, and under the established consensus codes of science as a

*corpus*. Fundamentally, in this series of events, preprint server upload becomes the first archiving event of the preprint as scientific document (Fig.).



**Fig.** Community articulation vs. time in the scientific information chain from preprint inception until journal publication and further dissemination. Progression to gray implies the gain in contents certification within the corpus of science. Integration of preprint servers with open community articulation platforms is envisaged, due to their closer open boundaries and immediacy features.

Although contemporary considerations of preprints are inherently linked to digital media, evidence has been found of former preprint attempts in printed format. As recently reported by *Matthew Cobb*,<sup>(10)</sup> an initiative called Information Exchange Groups (IEG) was run at NIH as early as on 1961 until 1966. It consisted on the exchange of printed preprints in a semi-public fashion among seven subject-specific communities by conventional mail, the preprints getting

subjected to what was regarded as “informal review”.<sup>(10)</sup> Up to 2 561 circulated among 3 663 members, more than half un-reviewed by journals.<sup>(11)</sup> A more local attempt, called Physics Information Exchange (PIE) –Cobb–, was run as joint library service between the Lawrence Radiation Laboratory at Berkeley and the Stanford Linear Accelerator Center (SLAC) at Stanford since 1969.<sup>(12)</sup> Conceptually, this reinforces the idea that preprints were a latent need for scientific communities since the very grounds of academic publishing, particularly in emerging scientific areas to speedily disseminate and contrast results and clear the view on earliest developments aimed to technological empowering and advance.

The successful seminal implementation of arXiv in the fields of Physics and Astronomy and its expansion to other topics,<sup>(13)</sup> were subsequently amplified by current practices such as the social networking phenomenon, which have spread the model of preprint servers. Moreover, due to the intrinsic archiving nature of preprint servers, they are inevitably linked to, and were acknowledged to fuel up efforts as those of the Open Archives initiative,<sup>(5)</sup> while restrained to research papers. Similarly, they have to deal with journals’ self-archiving policies and digital documents interoperability standards among digital collections.

At the same time, this was backed up by an increasingly demanding technological support for content archiving. Even for scientific pre-published materials (presentations, posters, programming codes, genetic materials, clinical protocols, research workbooks and so on), such digital tools have evolved, ultimately impacting the way in which science is conceived, shaped, transmitted and archived, before or after journal publishing, either paid or not.

As any model been introduced, preprints passed through a certain number of progressive denominations. They historically circulated as letters, memos or draft copies by e-mail among researchers, for commenting prior to journal submission. In this sense, they were characterized as “type 1 electronic publication” by Eysenbach,<sup>(14)</sup> their speed, work-in-progress and transient nature being established earlier, for collaboration improvement. Later, preprints were labeled as e-prints (or more precisely e-preprints) with the advent of refereed electronic versions being published in journal management platforms. They were made available in advance to printing but regarded as accepted papers, either corrected or not. It was increasingly accepted that, except for some technical fields as in engineering or nanotechnology,<sup>(15)</sup> works began to get published prior to their presentation in congresses, due to the availability of recording devices, to protect them from piracy and plagiarism, and for in-depth validation of results as traditionally considered. However, and as early identified by *Eisenbach*,<sup>(14)</sup> data and information in pre-manuscript form began to be documented, electronically hosted and reviewed by actors of the scientific system

(project and grants evaluation boards, ethic committees and national regulatory agencies, among others).<sup>(14)</sup> They were made available for public scrutiny while not considered as prior publication by journals, but with standards and scientific formats near those of the final manuscript for uploading and permanence.<sup>(13)</sup> This resulted from the progressive gain in complexity of science as a *corpus*, and the need for sequential documentation of experimental design, conduct and results, with standards specific for each of those phases. Notably, this coincided with the character of the final publication paper as commercial product, with the advent of the international editorial houses and consortiums, and regardless of the editorial management enforced model (except for subsidized journals devoid of costs for authors' publication and access), while having connections with academic library management of archives as documentation process (e.g., Netprints in 1999).<sup>(16)</sup> In this line, the current model of preprints complies with the green access variant of open access, due to be placed as it is the pre-publication stage.

Even more significantly, preprint servers, the platforms in which preprints are archived and retrieved from, were identified as articulating strong technical digital communities in a series of disciplines. This was envisaged earlier by the Keller's Delphi perceptual survey, with these servers as competitors for traditional journal models.<sup>(17)</sup> Therefore, and to some extent, preprint servers could be assumed as a publication movement amplified by digital forums and social networks, that is, the digital technological articulation of the human interaction during the inception of the final document. Moreover, they allow community articulation of research results by providing the context for consensus approval by the community, prior to journal peer-review, while providing a space for controversial studies.<sup>(18)</sup> In fact, the arXiv preprint server has been used to detect underlying scientific communities by network analysis.<sup>(19)</sup>

Importantly, the concept of preprint server considers preprints hosting and circulation prior or simultaneously to peer-review and journal publication, an action more often occurring in the reverse order, due to journal restrictions to the free circulation of the final paper. In spite of initial concerns about journal endorsement of open access policies and potential constrains to further publish the paper once made available as preprint, an increasing number of publishers and journals have backed-up the prepublication of manuscripts in preprint servers. In this regard, up to 46 % of the 2 566 publishers indexed in SHERPA RoMEO (a 10 % increase since Jan 2017<sup>(20)</sup>) support preprint archiving, distributed in green (preprint and postprint; 1 055, 41 %) and yellow (preprint; 175, 7 %) archiving policies. Other increasingly regional players, such as SciELO Network, first announced plans to operate and evaluate a preprint server for channeling

papers into journals<sup>(21)</sup> and more recently announced the joint development with PKP of a Preprint Server System, a strategy somehow reversing the cascade peer-review model of consortia journals derived from peer-review rejections.<sup>(22-24)</sup> The channeling approach attempts to redirect submissions before rather than during or after peer-review.

In fact, there were attempts to establish “channeling-up” strategies for preprints from preprint servers to journals, by proposing a modality of commissioned peer-review between preprint servers and journals, following the public scrutiny.<sup>(25)</sup> One such mechanism has been implemented recently, named “dubbed B2J”, which directly transfers preprints from bioRxiv (the Biology preprint server) to over 35 journals, including PLOS Genetics. In fact, this journal moved some of their editors to bioRxiv as preprints editors to look out for candidate preprints, who further recommend those preprints selected to the journal’s editorial board editors and invite the authors to submit the preprint to the journal for standard peer-review.<sup>(26)</sup> In a more recent development in 2018, PLOS started a partnering strategy with Cold Spring Harbor Laboratory to make possible for authors to simultaneously deploy the manuscripts during submission as preprints in bioRxiv, and to link published papers in PLOS journals to their existing preprints in that preprint server.<sup>(27)</sup>

As mentioned by *Brown* in 2003,<sup>(28)</sup> and further emphasized by *Li, Thelwall* and *Kousha*<sup>(29)</sup> concerns about the deliberate and non-certified posting of primary research data, could have certainly limited the straightforward implementation of publishing practices in preprint servers beyond the arXiv’s experience. This was evidenced in Chemistry and Life Sciences, mainly due to the lower level of consensus in the organizational structuring of research bodies and systems in areas other than Physics and Mathematics. In fact, those concerns coincide with parallel practices in Medicine journals, as the Ingelfinger’s rule.<sup>(30)</sup> This perspective got modified over time due to a careful opening of influential chemistry societies as the ACS to the preprint movement, with 20 out of the 50 ACS journals in 2016 endorsing the publication of manuscripts published as preprints.<sup>(31)</sup> Either the case, preprint submission policies are entangled at journal level and seem to be highly influenced by perception concerns of all the actors in the information chain. Particularly, perceptions could be susceptible to downstream considerations in case of preprint rejection and resubmission, with author’s irreversibly assuming the decision to publish just in journals endorsing such publication modality, to avoid further rejection due to prior public disclosure of data.

Additionally, preprint servers have indirectly contributed to foster digital archive management by the introduction of software implementations either to manage or to curate such digital document

collections (as the CDSware, CERN Document Server<sup>(32,33)</sup> and anti-plagiarism algorithms.<sup>(34,35)</sup> Moreover, they can be considered as platforms to model usage metrics, such as preprint page views and printing statistics.<sup>(35)</sup>

### **Advantages of preprint publication**

The advantages of preprint publication were noticed earlier: high speed of publication (from just 7 days to 2-4 months,<sup>(36,37)</sup> even immediate,<sup>(6)</sup> while challenged by the almost immediate peer-review model established by open access journals like PLOS ONE the inclusion of large datasets without page constraints, no publication submission charges and open access to information. These were stated for platforms including preprints, as part of the eprint concept of “any electronic work circulated by the author outside of the traditional publishing environment”, as early as in 2001.<sup>(36)</sup> They include long term archiving, mostly permanent, with some platforms allowing preprint removal by authors, and enhanced presentation, not bound to journal’s technical capabilities.

Moreover, preprint publication could provide alternatives to underreporting, incomplete or misleading reporting, as alleged in biomedicine,<sup>(18)</sup> by providing the first finished compilation of the protocol–summary results–final dataset–preprint–paper series of prepublication events. Significantly, authors retain the copyright, which could be further transferred to a journal or even retained while publishing.<sup>(38)</sup>

Regarding publication delays, there has been an increase in the complexity and amount of experimental data provided in articles, as identified by Vale for articles published in Cell, Nature and The Journal of Cell Biology, through the comparison of articles published in 1984 vs. 2014.<sup>(39)</sup> Moreover, he found that this was far more significant considering the introduction of supplementary information from 1997 onwards in digital format.<sup>(6)</sup> This coincided with the increase in the time required for peer-review as detected by Himmelstein in Pubmed papers in 1997.<sup>(7)</sup> Therefore, the delay in peer-review could be considered as one of the main forces pushing forward for the implementation of preprint servers, fostering the public availability of results and data, to claim recognition over results immediately once made public on the web and to justify financial funds,<sup>(4)</sup> particularly in those scientific areas of fierce competition for development and of limited funding.

Probably the most debated subject preprint server initiatives occurred in Chemistry,<sup>(40-44)</sup> with various attempts for preprint server implementation, either journal<sup>(41)</sup> or non-journal related.

Finally, in 2016, a new platform was announced.<sup>(42)</sup> As historically occurred with other research communities, they are responsible for putting in place the respective publication practices, and preprints are not the exception.

### **Types of preprint servers**

There are five non-exclusive types of preprint servers aside from the two former ones: 1) non-journal preprint servers (e.g., arXiv, biorXiv); 2) journal preprint servers (e.g., Netprints, currently inactive and official journals-editorial management websites); 3) mixed servers (e.g., GitHub, ResearchGate); 4) subject repositories, and 5) national and regional servers.

The first two types of servers corresponded to either non-journal or journal-bound servers. The third category includes platforms in which preprints are one of the many types of contents hosted.

Another category known as “subject repository”<sup>(29)</sup> includes platforms hosting different types of pre and postpublication stage documents with no journal boundaries under green (pre and postprint archiving allowed) or yellow (preprint only) open access. For instance, arXiv, RePec and the Social Sciences Research Network (SSRN), Pubmed Central (PMC) and even social networks as ResearchGate lay in this category.<sup>(29)</sup>

Additionally, preprint servers could be supported by non-commercial and non-editorial organizations. One of the most recent examples is the launch by the Wellcome Trust in UK of its own preprint server.<sup>(45)</sup> Furthermore, support from government and academic authorities for such platforms has succeeded. Journal content databases, scientific institutions and grant-providing organizations have started to take preprints and preprint servers seriously, beyond disciplinary attempts. For instance, governmental research funding institutions such as UK’s Medical Research Council (MRC) recently allowed the inclusion of manuscripts published on preprint servers in biomedical research grant applications.<sup>(45)</sup> This is in line with pioneering open access initiatives in medical sciences,<sup>(2)</sup> and the very recent pronouncement of NIH in March, 2017, encouraging researchers to cite preprints and other so-called “interim research products” (e.g., preregistered protocols and other unfinished research products waiting for scientific authority evaluation) for NIH funding assignment.<sup>(46)</sup> As plausible, the tag (preprint) was recommended, together with doi. This can be considered as an opportunity to evaluate ongoing research with greater immediacy than conventional publication, since such evaluation programs reside on their own technical capability for evaluation of documentation issues, either preprints or published material. There are two newer preprint platforms with less than 5000 preprints but growing fast: PrePubmed,<sup>(47)</sup> intended to publish preprints until accepted and indexed in Pubmed and cross-

indexing preprints from eight other preprint servers and the main contribution from biorXiv, and Preprints.org<sup>(48)</sup> which started to get indexed very recently by Europe PMC.

And the fifth category comprises national and regional preprint servers. Examples there are the national Chinese Preprint Server Online (CSPO), which is supported by the Center for Science and Technology Development (an agency of the Ministry of Education in the Peoples' Republic of China) and the Chinese Preprint Server (CSP).<sup>(49,50)</sup> Curiously, these two servers were started in 2003, 9 years before the parallel rivalry in scientific production between China and US in 2012. More recently, in 2018, a national server was opened for Indonesia, named INA-Rxiv (<https://osf.io/preprints/inarXiv>);<sup>(51)</sup> in a partnership with the OSF Network. National preprint servers could help to increase the immediacy of national science visibility while providing a real-time inventory of scientific output for the national scientific system. Similarly, regional servers have been deployed, as the recently announced African servers AfricarXiv (in collaboration with OSF Network) and AAS Open Research (in collaboration with F1000).<sup>(52)</sup>

### **Difficulties and limitations in preprint server development and alternatives for their solution**

As any other structural and technological implementation associated to information transmission and archiving, preprints and preprint servers do not escape limitations, technical difficulties and other communication phenomena inherent to scientific communication. Among them, we could find the appearance of predatory servers (parallel to predatory journals), content plagiarism, preference for digital document formats for publication and considerations inherent to the specific main science communication model in specific disciplines. Other concerns come from the need for time-stamping (particularly in those systems not subjected to technologically enriched environments) to support immediacy claims, the need for search engines to find pertinent preprints among the varied and increasing array of emerging servers, and the highly debatable phenomenon of preprint citation.

#### **Preprint document type and the acceleration of scientific publishing**

Despite their key function in scientific communication (i.e., immediacy), it seems that full research preprints are more commonly posted than review preprints, while the full research paper is the major document type. This could lead to the misassumption that preprint servers are more suitable for publishing research rather than review papers, the former prioritizing more reporting than generalizing and elaborate thinking.

Noteworthy, preprints are not the only factor accelerating the nature of scientific practice and communication nor are their servers the only technological platform available. As emphasized by

*Hurd et al*, as early as in 2002,<sup>(53)</sup> scientific data databases have followed for years microdocumentation strategies for reporting on experimental material (DNA and protein, datasets, clinical protocols), and scholars have made intensive use of any communication technology available for speeding up research spreading (e-mail messages, web forums, teleconferences, among others). Other examples can be found in the exchange of research information both in the pre-publication (experimental interactive notebooks<sup>(54,55)</sup> and post-publication stages (data journals),<sup>(56)</sup> including the post-publication peer-review.<sup>(57,58)</sup>

### **Preprint digital format and server preference by authors**

Other difficulties come from the selection of a given preprint digital format for upload among research communities, which could influence the preference for either server. For example, arXiv seems to demand that preprints be prepared in LaTeX, more popular in mathematics, physics and engineering, while pdf upload has been somewhat troublesome. This has occasionally caused migration to other platforms with not such restrictions as ResearchGate,<sup>(59)</sup> even when it is an academic social network. In this particular, preprints, either in pdf or other document formats, coexist with data of varied arrays and purposes in platforms other than preprint specific servers (i.e., mixed servers) as FigShare or GitHub. This could be solved by future implementation of format and interoperability standards for preprints as previously happened for articles in repositories and databases. Hence, these difficulties could be viewed as issues related to the developmental stage of preprint servers, as happened with digital libraries.

### **Preprint digital ID**

A relevant aspect comes from the identification of preprints in the digital environment. This has been partially solved by encouraging authors to add a doi (digital object identifier) to the preprint, to redirect to the definitive version once published.<sup>(59)</sup> There is a feedback from the published article, by including the preprint server identifier in the published version (for instance, the arXiv identification code). Besides, there is some level of feedback between subject preprint servers, as can be found to arXiv at the CERN server, with common links for preprints relevant on related topics. Probably, the most debated concern comes from the attribution of citations at the preprint stage to the final published version. Commonly, journals endorsing free self-archiving policies for un-refereed preprints solve this through doi,<sup>(60)</sup> but there may be a certain level of fragmentation of citations, mostly for preprints at non-journal and mixed preprint servers and regardless of the model of editorial peer-review applied once they will get published. A potential solution would be that bibliographic styles and databases could evolve to include a label of (preprint) at the end of the title for preprint identification, as in the case of abstracts,<sup>(61)</sup> as previously mentioned as recognized by the NIH initiative for funding.<sup>(46)</sup> That label could help

readers to be aware of the absence of peer review until preprints get refereed and published in a journal, since the main claim of preprints is author priority over the results, while simultaneously aiding on the identification of the time of citation. Moreover, author priority is a time-stamping related issue. In fact, time stamping, the technical signature for claiming priority, has been a great deal of concern in digital platforms, both in preprint servers and digital libraries repositories. Although doi implementation solves these problems, it is still troublesome in those spaces where doi has not been afforded. For instance, a long-term time-stamping electronic signature server was successfully proposed and implemented for preprints in pdf in a digital library of mathematics in Japan,<sup>(62)</sup> as well as for a mathematical journal.

### **Preprint immediacy versus quality**

The immediacy role of preprints has led to envisage them as a mechanism to foster discovery in biomedicine, besides its contributions to manuscript improvement by open peer-review or not.<sup>(63)</sup> Moreover, preprint adoption was evidenced as influenced by policies in epidemiology, due to the intrinsic rush for speed while facing outbreaks. This was demonstrated by *Johansson et al.* while studying the preprints posted in a group of preprint servers following the successive WHO emergency declarations during Ebola and Zika outbreaks in 2014 and 2016, respectively.<sup>(64)</sup> They found that preprints and data sharing increased between outbreaks, containing new analyses made available prior to the delay of editorial peer-review, and, consequently, they formulated some recommendations for making preprints a valuable tool for earlier data exchange during outbreaks.

Due to the speed up in communicating rather than in the quality of the content of the manuscript, imposed by an increasingly competitive scientific industry, other alternatives have to be explored and established to reinforce the scientific scrutiny and the evolution of critical thinking. For instance, concerns have accompanied the progressive introduction of preprint practices on the general public perception of science, reintroducing the debate on the difference between scientific evidence and facts particularly for media, as formerly acclaimed at the time of the Ingelfinger's rule.<sup>(65)</sup> But recent debates agree on the culture of healthy scrutiny at both sides of the boundary between the private and public spheres.<sup>(66-68)</sup>

The lack of peer-review practices prior to a wide dissemination, to safeguard from the inadequate communication of untrusted or misleading scientific data<sup>(69)</sup> is a recursive limitation which has been remarked while analyzing the impact of preprints on scientific communities in Biology. As stated,<sup>(69)</sup> it can be partially compensated by expert criteria as occurs during the same publication process in journals for data deposited in data databases and also published in data journals, to

complement published content with other non-published information at the preprint stage and also information available in blogs and websites.

While open content can be commented and tagged in social networks and other open discussion platforms as ResearchGate and Mendeley, new instrumentations specific for preprints are appearing. One example is preLights (<https://prelights.biologists.com/>), supported by The Company of Biologists, aiding to select, highlight and comment on new reprints in biological sciences as a subject community platform.<sup>(70)</sup> Notice that this goes beyond the formal focus of standard peer-review and is intermediate between open peer-review and social media and closer to the post-peer review practices aiding in ranking already published papers (in this case, preprints). Since they are preprints, the opinions made available could certainly contribute to the subsequent peer-review process. Importantly, there is also a platform which was opened in 2017 for preprint open review, named PREreview,<sup>(71)</sup> which is supported by the collaborative writing platform Authorea developed by CERN researchers for collaborative research.

In general, it is common that journals request from authors not to communicate their findings to the media until the article becomes published and scientific methodologies and results carefully scrutinized, to avoid misleading claims or even worse outcomes<sup>(72)</sup> Ultimately, this remarks on the increasing need for an information literacy effort of the general audience, to get educated on the nature of preprints.<sup>(73)</sup> This could help to avoid sensationalisms and unexpected reactions as those that eroded earlier this year the stocks of three biotech companies in response to a preprint posted in biorXiv on the clinical application of CRISPR technology.<sup>(74)</sup>

### **Plagiarism and plagiarism detection systems**

As expected, preprint servers do not escape from plagiarism attempts,<sup>(34)</sup> and strategies have been considered for plagiarism detection. Noteworthy, it seems to be a low incidence phenomenon, at least in arXiv, partially contained by the easy detection under public scrutiny due to its open (public) nature. Moreover, papers found to be plagiarized are readily labeled or withdrawn, which puts a ban on the authors' future work credibility. In this line of development, an algorithm was implemented for detecting not just word sequences against the entire collection of documents on the server, but also the flows of ideas.<sup>(75)</sup>

### **Preprint citation**

Reconsidering, the pre-publication stage identification in the bibliographic citation format is consistent with similar tagging of other bibliographic types of scientific information disclosed prior to journal publishing, as for conference presentations and dissertations,<sup>(61)</sup> and this impacts on citation trends. In fact, preprints archived in server preprints are actively cited.<sup>(76,77)</sup> Moed, in 2007, found that preprints accelerated citations as detected in WoS for preprints in the arXiv's

Condensed Matter Section by means of immediacy rather than free access to the information.<sup>(77)</sup> Furthermore, *Larivière* et al studied the entire arXiv collection and found that, in spite of providing citations earlier, preprints had lower citation rates and citation decayed faster than for published articles indexed in the Web of Science.<sup>(76)</sup> Overall, the arising picture is consistent with the main function of preprints to make information publicly available and its transient nature on the way to publication. Moreover, preprint citations should have to be differentiated from and inherited by the published paper (see Annex for a list of some of the most relevant preprint servers, the number of hosted preprints and their citation in Scopus). This could be a general explanation for the sudden decay in preprint citations once the respective article is published, in spite of aspects inherent to citation dynamics, such as subject-specific half time citation trends<sup>(76)</sup> or negative citations,<sup>(78)</sup> among others. At the same time, uncorrected and corrected proofs in journal servers accumulate in journal platforms, waiting to be published, a time in which they got consulted and cited, with evidences on inflated impact measures.<sup>(79)</sup>

### **Predatory preprint servers**

As well as for open access journals, predatory preprint repositories and websites appeared. An example is the fake repository ChemarXiv, related to Open Academic Press predatory publisher, as difficult to identify as predatory journals.<sup>(80)</sup> Notably, this seems to have somewhat complicated the perception in the field of chemistry for successful and persistent preprint server implementation, in addition to prior unsuccessful attempts.

### **Preprint overload and preprint search engines**

Second, information overload, in addition to the rise in the number of papers published in journals.<sup>(39)</sup> In fact, the latter is further complicated by difficulties to search among the different potentially relevant preprint servers. To cope with such limitation, a new search engine named search.bioPreprint tool (<http://www.hslls.pitt.edu/resources/preprint>) was launched recently, for simultaneous multi-preprint server retrieval from arXiv, biorXiv, F1000Research, and PeerJ Preprints.<sup>(69)</sup> Notice the varied nature of the preprint servers included.

Another approach comes from aggregation of open access repositories containing a wide array of document types. One of such powerful initiatives aiding to retrieve open access documents is the Bielefeld Academic Search Engine (BASE, Bielefeld University Library; [www.base-search.net](http://www.base-search.net)), which does not allow to filter for preprints as document type while they can be retrieved by searching for preprint in the basic search interface. As on December 6th, 2018, that search

provided up to 1 186 818 documents of which 814 861 were documents (not just preprints) from RePEc, and actual preprints from CiteSeerX (69 084), PeerJ (9 832), arXiv (9 776) and Zenodo (1 307). This means that OA search engines are closer to preprint search engines than those non-OA.

### **Preprint server integration**

Preprint server integration could come not only from searching, but from the very writing process and content interoperability among servers. A promising example is that of the one-year old preprints server (<https://www.preprints.org>) which is on the way to integrating the collaborative online writing tool Fiduswriter (<https://www.fiduswriter.org>), which has been recently integrated with the Open Journal Systems (OJS) online editorial management system,<sup>(81,82)</sup> and developing tools for preprint conversion into XML format with the aim of interoperability among preprint servers. This emerged in response to the call on the need for establishing a preprint aggregator made by ASAPbio (<http://asapbio.org/>), the initiative to promote the use of preprints in life sciences. Notably, ASAPbio fostered the creation of a central preprint service in US, provided with combined capabilities as in Pubmed and the Pubmed Central repository for indexing, storage, search and retrieval of preprints in life sciences,<sup>(83)</sup> in addition to other extended capabilities, as a core hub of preprints for scientists.

## **CONCLUDING REMARKS**

Undoubtedly, preprints and preprint servers are useful implementations of scientific communication from the very same inception of scientific manuscripts. This has been made possible by the increased capabilities for archiving and retrieval of digital content at the journal pre-publication stage, in addition to the former expansion in parallel post-publication dissemination. Although uniformity and consensus are still under development, mostly for the uneven introduction among scientific disciplines, it is just a matter of time for the uniform consideration of preprints with their shortcomings for what they are: transient documents on the way to getting published, successfully or not. And irrespective of their diversity, the emergence and expansion of the preprint movement complement previous steps even more seminal for the scientific communication processes, such as experimental data sharing.

Due to the implicit step of manuscript integration into a digital technology ecosystem, aside from its writing environment, preprint servers could be potentiated as document interoperability platforms. This could include integrated tools for semantic tagging and XML-JATS format

during submission, prior to formal peer review, and possibly joining the peer-review communication as an output for online editorial management systems. Although this sort of embedding seems not to be available yet, this would pave the way for database indexing and would integrate the entire scientific communication system in terms of content enrichment and automation, especially by joining artificial intelligence developments. Significantly, they detach document structuring from methodological and scientific veracity checking during peer-review, something which could foster peer review if integrating journal forms and style check utilities within those platforms. Current demands for merging and interoperability only reinforce the view of their current stage as intermediate to their full contribution to foster science.

### **Acknowledgements**

The author thanks the Center for Genetic Engineering and Biotechnology (CIGB) of Havana, Cuba, and the National Library of Medicine of Cuba for the technical capabilities provided for conducting this research. Also to Professor *Deborah Torres Ponjuán*, PhD., at the Faculty of Communication, University of Havana, Cuba, for providing useful information sources. Many thanks also to the editor and two independent reviewers who contributed to improve the submitted manuscript.

## **BIBLIOGRAPHIC REFERENCES**

1. Ginsparg P. The global village pioneers. *Learn Publ.* 2009;21(2):95-100.
2. Board NE. Nature respects preprint servers. *Nature.* 2005. Access: 17/3/2005;434:257. Available from: <https://www.nature.com/articles/434257b>
3. Graf C, Battisti WP, Bridges D, Bruce-Winkler V, Conaty JM, Ellison JM, et al. Research Methods & Reporting. Good publication practice for communicating company sponsored medical research: the GPP2 guidelines. *BMJ.* 2009;339:b4330.
4. Chan L, Cuplinskas D, Eisen M, Friend F, Genova Y, Guédon JC, La Manna M. Budapest Open Access Initiative. Declaraciones del Movimiento Internacional de Acceso Abierto; 2002. Access: 19/8/2017. Available from: <http://www.budapestopenaccessinitiative.org/>
5. Luce RE. The open archives initiative: Interoperable, interdisciplinary author self-archiving comes of age. *Serials Libr.* 2001;40(1-2):173-82.
6. Powell K. Does it take too long to publish research? *Nature.* 2016;530:148-51.
7. Himmelstein D. The history of publishing delays. *Personal Blog*; 2016. Access: 16/8/2017.

Available from: <http://blog.dhimmel.com/history-of-delays/>

8. Gak M. The public sphere: migration of normative principles and the digital construction of transnational ethics. In: Karatzogianni A, Nguyen D, Serafinelli E (Eds.). The digital transformation of the public sphere: Conflict, Migration, Crisis and Culture in Digital Networks. London: Palgrave Macmillan; 2016. p. 11-33.
9. Cook T. Evidence, memory, identity, and community: Four shifting archival paradigms. Arch Sci. 2013;13(2-3):95-120.
10. Cobb M. The prehistory of biology preprints: A forgotten experiment from the 1960s. PLOS Biol. 2017;15(11):e2003995.
11. Kaiser J. How biologists pioneered preprints-with paper and postage. Science. 2017;357(6358):1348.
12. Rosenfeld A, Wakerling RK, Addis L, Gex R, Taylor RJ. Preprints in particles and fields. SLAC-PUB-0710; 1970.
13. Ginsparg P. Preprint Déjà Vu. EMBO J. 2016;35(24):2620-5.
14. Eysenbach G. The impact of preprint servers and electronic publishing on biomedical research. Curr Op Immunol. 2000;12(5):499-503.
15. Tierney HL, Hammond P, Nordlander P, Weiss PS. Prior publication: Extended abstracts, proceedings articles, preprint servers, and the like. ACS Nano. 2012;6(9):7543-44.
16. Delamothe T, Smith R, Keller MA, Sack J, Witscher B. Netprints: the next phase in the evolution of biomedical publishing. Brit Med J. 1999;319:1515-6.
17. Keller A. Future Development of Electronic Journals: a Delphi Survey. Electron Libr. 2001;19(6):383-96.
18. Chalmers I, Glasziou P. Should there be greater use of preprint servers for publishing reports of biomedical science? F1000Research. 2016:5.
19. Gopalan PK, Blei DM. Efficient discovery of overlapping communities in massive networks. EE.UU.: Proc Nat Acad Sci. 2013;110(36):14534-9.
20. SHERPA/RoMEO. RoMEO Statistics. RoMEO; 2018. Access: 24/10/2018. Available from: <http://www.sherpa.ac.uk/romeo/statistics.php?la=en&fIDnum=|&mode=simple>
21. Packer A, Santos S, Meneghini R. SciELO Preprints on the way. SciELO in Perspective. Blog SciELO. 2017. Access: 15/8/2018. Available from: <https://blog.scielo.org/en/2017/02/22/scielo-preprints-on-the-way/>
22. PKP and SciELO announce development of open source Preprint Server system. SciELO in Perspective. 2018. Access: 24/10/2018. Available from:

<https://blog.scielo.org/en/2018/09/21/pkp-and-scielo-announce-development-of-open-source-preprint-server-system/>

23. Barroga EF. Cascading peer review for open-access publishing. *Eur Sci Ed.* 2013;39(4):90-1.
24. Kovanis M, Trinquart L, Ravaud P, Porcher R. Evaluating alternative systems of peer review: a large-scale agent-based modelling approach to scientific publication. *Scientometrics.* 2017;113:651-71.
25. Till JE. Peer review in a post-eprints world: a proposal. *J Med Internet Res.* 2000;2(3):E14.
26. Barsh GS, Bergman CM, Brown CD, Singh ND, Copenhaver GP. Bringing PLOS Genetics Editors to Preprint Servers. *PLOS Genetics.* 2016;12(12):e1006448.
27. PLOS. Power to the Preprint: An Update. *PLOS*; 2018. Access: 26/10/2018. Available from: <https://blogs.plos.org/plos/2018/09/power-to-the-preprint-an-update/>.
28. Brown C. The role of electronic preprints in chemical communication: Analysis of citation, usage, and acceptance in the journal literature. *J Am Soc Inf Sci Technol.* 2003;54(5):362-71.
29. Li X, Thelwall M, Kousha K. The role of arXiv, RePEc, SSRN and PMC in formal scholarly communication. *Aslib J Inf Manage.* 2015;67(6):614-35.
30. Relman AS. The Ingelfinger's rule. *New Engl J Med.* 1981;305:824-6.
31. Voosen P. Chemists to get preprint server of their own American Chemical Society launches ChemRxiv despite dubious precedents. *Science.* 2016;353(6301):740.
32. Cleyde S, Sitas A. CDSware CERN Document Server Software). *Library Hi Tech.* 2006;24(3):420-9.
33. Pepe A, Baron T, Gracco M, Le Meur JY, Robinson N, Simko T, Vesely M. CERN Document Server Software: The integrated digital library. *CERN-OPEN*; 2006.
34. Giles J. Preprint server seeks way to halt plagiarists. *Nature.* 2003;426(6962):7.
35. Town WG, Vickery BA, Kuras J, Weeks JR. Chemical e-journals, chemical e-preprints. *Online Inf Rev.* 2002;26(3):164-71.
36. Garner J, Horwood L, Sullivan S. The place of eprints in scholarly information delivery. *Online Inf Rev.* 2001;25(4):250-6.
37. Linden DJ. Preprint servers and the Journal of Neurophysiology. *J Neurophysiol.* 2009;102(5):2577.
38. Miranda GF, Ginestet J. The attitude of pharmaceutical industry research scientists to browsing and publishing on internet preprint and e-print servers. *Drug Inf J.* 2002;36(4):831-7.
39. Vale RD. Accelerating scientific publication in biology. *EE.UU.: Proc Natl Acad Sci.*

2015;112(44):13439-46.

40. Garson LR. Communicating original research in chemistry and related sciences. *Accounts Chem Res.* 2004;37(3):141-8.

41. Janet EA, Boyett RE, Town WG. Communities on the Web: ChemWeb.com - The World Wide Club for the chemical community. *Trends Anal Chem.* 1998;17(2):54-8.

42. Kiessling LL, Fernández LE, Alivisatos AP, Weiss PS. ChemRXiv: A chemistry preprint server. *ACS Chem Biol.* 2016;11(11):2937.

43. Warr WA. Evaluation of an experimental chemistry preprint server. *J Chem Inf Comp Sci.* 2003;43(2):362-73.

44. Weeks JR, Kuras J, Town WG, Vickery BA. The chemistry preprint server: An experiment in scientific communication. *J Chem Inf Comp Sci.* 2002;42(3):765.

45. Hampton K. U.K.'s Medical Research Council encourages use of preprints. UK: Medical Research Council; 2017. Access: 24/10/2018. Available from:

<http://rescuingbiomedicalresearch.org/blog/u-k-s-medical-research-council-encourages-use-preprints/>

46. NIH. Reporting preprints and other interim research products. NIH: Notice Number NOT-OD-17-050; 2017. Access: 17/4/2017. Available from: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-050.html>

47. Anaya J. PrePubMed: A PubMed for preprints. PrePubMed. 2016. Access: 24/10/2018. Available from: <http://www.omnesres.com/research/prepubmed/>

48. Preprints Editorial Office. Better visibility for preprints: Inclusion in Europe PMC. Europe PMC; 2018. Access: 24/10/2018]. Available from:

<https://www.preprints.org/announcement/show/26>

49. Hu C, Zhang Y, Chen G. Exploring a New Model for Preprint Server: A case study of CSPO. *J Acad Libr.* 2010;36(3):257-62.

50. Yaokun Z, Nanqiang X. The usage and acceptance of domestic preprint servers in China. *Interlend Docum Supply.* 2008;36(3):152-7.

51. Shih I. Indonesian preprint server takes off. *Nature.* 2018;553(7687):139.

52. Nordling L. African scientists get their own open-access publishing platform. *Nature.* 2017. Access: 10/11/2018. Available from: <https://www.nature.com/news/african-scientists-get-their-own-open-access-publishing-platform-1.23018>

53. Hurd J, Brown CM, Bartlett J, Krietz P, Paris G. The role of "unpublished" research in the scholarly communication of scientists: digital preprints and bioinformation databases –

Sponsored by SIG STI, SIG BIO, SIG PUB. Proceedings of the 65th Asist Annual Meeting; 2002.

54. Bohle S. Open access: Online repository for lab notebooks. *Nature*. 2014;506(7487):159.

55. Shen H. Interactive notebooks: Sharing the code. *Nature*. 2014;515(7525):151-2.

56. Candela L, Castelli D, Manghi P, Tani A. Data journals: A survey. *J Assoc Inform Science Technol*. 2015;66(9):1747-62.

57. Hunter J. Post-publication peer review: opening up scientific conversation. *Front Comput Neurosci*. 2012;6:63.

58. Kirkham J, Moher D. Who and why do researchers opt to publish in post-publication peer review platforms? - findings from a review and survey of F1000 Research. *F1000Research*. 2018;7:920.

59. Wildberger N. The arXiv *versus* ResearchGate. *ResearchGate*; 2015. Access: 24/8/2018. Available from: <https://njwildberger.com/2015/08/30/the-arXiv-versus-researchgate/>

60. Das AK. "Peer review" for scientific manuscripts: Emerging issues, potential threats, and possible remedies. *Med J Arm Forc Ind*. 2016;72(2):172-4.

61. Youngen GK. Citation patterns to traditional and electronic preprints in the published literature. *Coll Res Libr*. 1998;59(5):448-56.

62. Namiki T, Yamaji K, Kataoka T, Sonehara N. Time stamping preprint and electronic journal server environment. In: *Towards a Digital Mathematics Library*. Brno: Masaryk University Press; 2011. p. 19-23.

63. Aquino-Jarquín G, Valencia-Reyes JD, Silva-Carmona A, Granados-Riveron JT. Preprints in biomedicine: alternative or complement to the traditional model of publication? *Gac Med Mex*. 2018;154(1):87-91.

64. Johansson MA, Reich NG, Meyers LA, Lipsitch M. Preprints: An underutilized mechanism to accelerate outbreak science. *PLOS Med*. 2018;15(4):e1002549.

65. Sheldon T. The impact of preprint on media reporting of science. *Lancet*. 2018;392(10154):1194.

66. Fraser J, Polka J. Preprints: safeguard rigour together. *Nature*. 2018;560(7720):553.

67. Sarabipour S. Preprints: good for science and public. *Nature*. 2018;560(7720):553.

68. Tennant J, Gatto L, Logan C. Preprints: help not hinder journalism. *Nature*. 2018;560(7720):553.

69. Iwema CL, LaDue J, Zack A, Chattopadhyay A. search.bioPreprint: a discovery tool for cutting edge, preprint biomedical research articles. *F1000Res*. 2016;5:1396.

70. Brown K, Pourquie O. Introducing preLights: preprint highlights, selected by the biological community. *Development*. 2018;145(4):1.
71. PREREVIEW. Read and engage with Preprint Reviews. Blog posts; 2018. Access: 24/10/2018. Available from: <https://www.prereview.org/>
72. Peralta D. Early 2018 Update: Board Members, Preprints and Special Issues. *Chemmedchem*. 2018;13(9):861-8.
73. Bourne PE, Polka JK, Vale RD, Kiley R. Ten simple rules to consider regarding preprint submission. *PLOS Comput Biol*. 2017;13(5):e1005473.
74. von Schaper E. Preprint wipes millions off CRISPR companies' stocks. *Nat Biotechnol*. 2018;36(3):211.
75. Feder T. Experimenting with plagiarism detection on the arXiv. *Physics Today*. 2007;60(3):30-1.
76. Larivière V, Sugimoto CR, Macaluso B, Milojevic S, Cronin B, Thelwall M. arXiv e-prints and the journal of record: An analysis of roles and relationships. *J Assoc Inf Sci Technol*. 2014;65(6):1157-69.
77. Moed HF. The effect of "open access" on citation impact: An analysis of arXiv's condensed matter section. *J Am Soc Inf Sci Technol*. 2007;58(13):2047-54.
78. Catalini C, Lacetera N, Oettl A. The incidence and role of negative citations in science. *EE.UU.: Proc Natl Acad Sci*. 2015;112(45):13823-6.
79. Tort AB, Targino ZH, Amaral OB. Rising publication delays inflate journal impact factors. *PLOS ONE*. 2012;7(12): e53374.
80. Widener A. Beware of a bogus ChemRxiv. *Chem Eng News*. 2016;94(46):16.
81. Sadeghi A, Wilm J, Mayr P, Lange C. Opening scholarly communication in social sciences by connecting collaborative authoring to peer review. arXiv Preprint; 2017 Access: 16/10/2018. Available from: <https://arXiv.org/abs/1703.0442.8>
82. Wilm J. OJS/Fidus Writer integration at the Open Science Conference. *Fiduswriter*; 2017. Access: 17/4/2017. Available from: <https://www.fiduswriter.org/2017/03/22/ojs-fidus-writer-integration-osc2017/>.
83. ASAPbio. Creation of a central preprint service for the life sciences. Blog ASAPbio; 2016. Access: 17/4/2017. Available from: <http://asapbio.org/summary-of-a-central-preprint-service-model>.

**Annex:** Active preprint servers and preprint related repositories, with some of their main characteristics\*

Launch date	Preprint archives	Type	Main subjects	Website	Type of content and comments	Number of preprints	Citing journal articles in Scopus **
1991	arXiv	Preprint server	Mathematics, Physics and Astrophysics	<a href="https://arXiv.org/">https://arXiv.org/</a>	Preprints	1 461 309	118 430
1993	CERN Preprint Server	Preprint server	Nuclear physics	<a href="https://cds.cern.ch/collection/CERN%20Preprints">https://cds.cern.ch/collection/CERN%20Preprints</a>	Evolved into the CERN Library Server in 1996 Includes the CERN Document Server Software (CDSware), a suite for the in- and out-flow of documents in an autonomous digital library server. Run by the European Organization for Nuclear Research.	21 967	1 499
1994	Social Science Research Network (SSRN)	Digital library	Social sciences and humanities. Mainly Economics, Law, Finance and Political Science	<a href="https://papers.ssrn.com/sol3/DisplayAbstractSearch.cfm">https://papers.ssrn.com/sol3/DisplayAbstractSearch.cfm</a>	Partially open access Papers able to be removed by the authors from the archives Twenty-four discipline-oriented research networks	703 760 (full text papers as declared at the website)	7 984
1997	Research Papers in Economics (RePEc)	Subject Repository Network	Economics	<a href="http://repec.org/">http://repec.org/</a>	Preprints, journal articles, working papers, software components Decentralized archives network	104 303	15 172

2000	F1000Research	Journal preprint server	18 subjects, mainly biology and medicine	<a href="https://f1000research.com">https://f1000research.com</a>	Resulted from the merge of F1000Posters with the F1000 former platform Preprint non-refereed and refereed versions are independently citable Preprints, data notes	2407	10 480
2003	Chinese Science Paper Online (CSPO)	National preprint server	39 disciplines under the Chinese Natural Science Subject Classification and Codes of National Standards. Major disciplines (24 %): Electronic, communications and autocontrol technology; and Computer Science and Technology	<a href="http://www.paper.edu.cn/">http://www.paper.edu.cn/</a>	Centralized Chinese scientific preprint server Post-publication peer review and refereed certificate, regardless official publication	5 992	2 686
2003	Chinese Preprint Server (CPS)	National preprint server	Natural Sciences, Medicine, Humanities and Social Sciences, Engineering and Technical Sciences, Agricultural sciences	<a href="http://preprint.nstl.gov.cn/">http://preprint.nstl.gov.cn/</a>	Centralized Chinese scientific preprint server	13 605	3
2008	GitHub	Software repository and Internet hosting service	Software and Computer Science, mainly	<a href="https://github.com/">https://github.com/</a>	Project-oriented, collaborative and software version control repository Collaborative writing documents made on Authorea can be synchronized from this repository	ND	ND

					Include preprints		
2008	ResearchGate	Scientific social network	Multiple	<a href="https://www.researchgate.net/">https://www.researchgate.net/</a>	Published and unpublished content Postings regarded as in personal pages	ND	18 026
2011	FigShare	Mixed repository	Varied	<a href="https://figshare.com/">https://figshare.com/</a>	Supplementary data, unpublished datasets, figures and videos Replaced Nature Precedings journal preprint server	ND	ND
2011	OSF***	Preprint repository network	Arts and humanities, Education, Law, Social and behavioral sciences	<a href="https://osf.io/preprints/">https://osf.io/preprints/</a>	Preprints	ND	ND
2013	bioRxiv	Preprint server	Biology and life sciences	<a href="http://bioRxiv.org/">http://bioRxiv.org/</a>	Preprints	36 111	8 203
2013	Peer J	Non peer-reviewed journal preprint server	Medicine, Biology & Life Science, Computer Science and General	<a href="https://peerj.com/preprints-search/">https://peerj.com/preprints-search/</a>	Draft, incomplete or final version	4388	ND
2013	Zenodo****	Research data repository for datasets	Particle physics, mainly	<a href="https://zenodo.org/search?page=1&amp;size=20&amp;q=preprint&amp;type=publication&amp;subtype=preprint">https://zenodo.org/search?page=1&amp;size=20&amp;q=preprint&amp;type=publication&amp;subtype=preprint</a>	Includes publications, software, datasets, image, poster and video	721	ND

2017*	ChemRxiv	Preprint server	Chemistry	<a href="http://www.chemrxiv.org/">http://www.chemrxiv.org/</a>	Supported and collaboratively managed by the American Chemical Society (ACS), German Chemical Society (GDCh) and the Royal Society of Chemistry. Hosts preprints in 16 subject categories: Agriculture and food chemistry, Analytical chemistry; Biological and medicinal chemistry; Catalysis; Chemical education; Chemical engineering and industrial chemistry; Earth, space, and environmental chemistry; Energy; Inorganic chemistry; Materials science; Nanoscience; Organic chemistry; Organometallic chemistry; Physical chemistry; Polymer science; Theoretical and computational chemistry	812	68
2016	Preprints	Preprint server	Multidisciplinary: Arts and humanities, behavioral sciences, biology, chemistry, earth sciences, engineering, life sciences, materials science, mathematics and computer science,	<a href="https://www.preprints.org">https://www.preprints.org</a>	Provided with an advisory board of 126 experts in the different subject areas.	7 081	3 622

			medicine and pharmacology, physical sciences, social sciences				
2016	Engrxiv	Preprint server	Engineering	<a href="https://osf.io/preprints/engrxiv">https://osf.io/preprints/engrxiv</a>	Preprints	ND	0 0
2016	PsyarXiv	Preprint server	Psychology	<a href="https://osf.io/preprints/psyarXiv">https://osf.io/preprints/psyarXiv</a>	Preprints	ND	89
2016	SocarXiv	Preprint server	Social Sciences	<a href="https://osf.io/preprints/socarXiv">https://osf.io/preprints/socarXiv</a>	Preprints	ND	42
2016	Welcome Open Research	Journal preprint server	General preprint server	<a href="https://wellcomeopenresearch.org/">https://wellcomeopenresearch.org/</a>	Welcome Trust funded research's preprint repository, powered by F1000	296	232

\* The list of servers was structured on February 16<sup>th</sup>, 2017 and data updated on November 19<sup>th</sup>, 2018. Servers were checked for functionality and the number of preprints computed from their respective websites.

\*\* Identified in Scopus by searching with the strategy 'WEBSITE (server website domain) OR (SRCTYPE (j) AND REFSRCTITLE(server) AND NOT REFTITLE(server))'. In the case of journal preprint servers, the counts of citing documents in Scopus include those of papers in the respective journal. Data was collected in November 19<sup>th</sup>, 2018. Search was not determined (ND) for repositories hosting other contents in addition to preprints, in the case of GitHub (codes), and for PeerJ stats overlap with standard journal citation. In the case of ResearchGate, results were inspected for verification.

\*\*\* Platform for the open dissemination of science, backed-up by the Center for Open Science in US. Includes the preprint servers Engrxiv, PsyarXiv, SocarXiv in their respective disciplines, and is on the way to launching AgriXiv in Agricultural sciences and PaleorXiv on Paleontology.

\*\*\*\* The URL is the one corresponding to preprints, due to the varied nature of contents.

### Conflicts of interest statement

The author declares that there are no potential conflicts of interest.