publications

MDPI

Article

Preprints in Scholarly Communication: Re-Imagining Metrics and Infrastructures

B. Preedip Balaji [1,*] and M. Dhanamjaya [2]

[1] Indian Institute for Human Settlements Library and School of Library and Information Science, REVA University, Bengaluru 560064, India

[2] School of Library and Information Science, REVA University, Yelahanka, Bengaluru 560064, India; registrar@reva.edu.in

* Correspondence: pbalaji@iihs.ac.in; Tel.: +91-080-6760-6661

check for updates

Abstract: Digital scholarship and electronic publishing within scholarly communities change when metrics and open infrastructures take center stage for measuring research impact. In scholarly communication, the growth of preprint repositories as a new model of scholarly publishing over the last three decades has been one of the major developments. As it unfolds, the landscape of scholarly communication is transitioning—with much being privatized as it is made open—and turning towards alternative metrics, such as social media attention, author-level, and article-level metrics. Moreover, the granularity of evaluating research impact through new metrics and social media changes the objective standards of evaluating research performance. Using preprint repositories as a case study, this article situates them in a scholarly web, examining their salient features, benefits, and futures. Moves towards scholarly web development and publishing on the semantic and social web with open infrastructures, citations, and alternative metrics—how preprints advance building the web as data—is discussed. We determine that this will viably demonstrate new metrics and, by enhancing research publishing tools in the scholarly commons, facilitate various communities of practice. However, for preprint repositories to be sustainable, scholarly communities and funding agencies should support continued investment in open knowledge, alternative metrics development, and open infrastructures in scholarly publishing.

Keywords: preprint repositories; scholarly publishing; scholarly communication; scholarly metrics; open infrastructures; scholarly web

1. Introduction

Electronic publishing has provided many benefits for sharing research materials online. Besides the mainstream publishing in books, peer reviewed journals, and conference papers, research outputs have increased in many other forms—preprints, datasets, multimedia, and software—not only for dissemination, but also for reproducibility and replication. Although outputs of publications in a variety of ways have increased, preprints stand out for their "accessibility" to early disseminated versions and "subject to review" status. They are publicly accessible and typically in line with definitions of open access, before being formally published. Preprints are scientific publications that are published online and publicly accessible before peer review in a journal publication. Growth in numbers of preprints [1] and the repositories to host them are on the rise, covering different disciplines. Specifically, they are moving beyond natural sciences to social sciences and humanities, although there is widespread skepticism [2] among scholarly communities about their acceptance of and recognition for scientific validation. Along with the growing trend for open access publishing [3], preprint repositories have grown, "while still used for small portion of papers, provided much earlier

access to scientific findings" among the scholarly communities [4,5]. As the need for access to research is felt widely, especially at an early stage to accelerate the access to new findings, many institutions and organizations have tried to establish preprint repositories alongside the scholarly publishing platforms from the late twentieth century. In one of the earliest experiments, the National Institute of Health (NIH) initiated a biological preprints circulation program called 'Information Exchange Groups' in 1961. Since journal publishers were not accepting preprints, this was shut down in 1967 [6]. Again, at the NIH's public archive platform PubMedCentral in 2000, establishing a preprint section was proposed. However, it was severely criticized by scientific publishers on the ground that "publishing preprints electronically sidesteps peer-review and increases the risk that the data and interpretations of a study will be biased or even wrong...the best way to protect the public interest is through the existing system of carefully monitored peer-review, revision and editorial commentary in journals [7]". This remains as one of the main debates since then as to why preprints cannot be accepted without peer-review or through any other feedback mechanisms.

Notwithstanding, the growth and diversity of preprint repositories in the last two decades reveal many other reasons as to how they play a vital role in the scholarly publishing ecosystem for their benefits, metrics, and risks. ArXiv was launched in 1991 and it set the trend of preprint-driven open scholarship (or e-prints server) in physics, computer science, and mathematics. In social sciences and economics, the Social Sciences Research Network (SSRN) was launched in 1994 and Research Papers in Economics (RePEc) in 1997. In 2008, the social academic networking sites Academia.edu and ResearchGate.net were launched, which had features that were more social and possessing options to accept research documents at any stage. Biology preprints, bioRxiv and PeerJ Preprints, were launched in 2013. In 2016, ChemRxiv for chemistry and SocArXiv for social sciences were launched. Earth sciences preprints, which were called ESSOAr, were launched by the American Geophysical Union in 2018 [8,9]. In addition to their principal benefits of making the scholarly content available for open access, preprint repositories break through traditional barriers—paving ways for new metrics, benefits, and research impact. For these reasons, preprint repositories emerged as a key player in scholarly publishing and they will continue to be a boon as an open infrastructure for researchers.

2. Background

In pursuit of open knowledge since the eighteenth century, scientific and scholarly communities exchanged communication without any formally integrated and holistic use of peer review [10]. Nevertheless, when the body and expanse of scholarly literature grew exponentially in the mid- to late-twentieth century with information explosion, the dissemination of current knowledge, the archiving of the canonical knowledge base, quality control of published information, assignment of priority, and credit for their work to authors, became a norm for the peer reviewing process [11]. Along the way, in scientific writing, various roles of authorship, levels of contribution, and the rules for publishing research data in the public domain, especially before the paper released was defined by journals. It was the then editor of The New England Journal of Medicine, Franz J. Ingelfinger, whose ideas on "sole distribution" in his editorial in September 1969 for scientific communities became popular [12]. Subsequently, journals became the primary mode of communication much before the widespread of peer review. With clear guidelines for authorship, academics and researchers primarily began communicating scientific research through peer reviewed journals for publishing their scholarship.

Post-1990s, in which the Internet came to be much more widespread, this did not disrupt the scholarly publishing perhaps as much as expected, as we still have the same large players that we did in the pre-digital age dominating the landscape [13]. It was thought that the Web would kill off scholarly journals, because the cost of dissemination would plummet to near zero. However, the large publishers simply shifted the offline system online—which is why we still have things like journals, issues, articles, copyright, and metrics that are designed for a pre-Web era. What this did, importantly, was to emphasize that it was publishers who were in charge, because they manage the

metrics (for evaluation/reputation) and the copyrights. Preprints challenge both of these things. This is perhaps a deeper significance that needs to be explored. Many commercial publishers and open access mega-journals consolidated their positions as large players, even as the ties between open access (OA) and incentives and power relationships between politics, publishers, and academies increased [14]. In addition, geographical heterogeneity and geopolitics play a larger part in it, as both countries in the global North and global South are attempting to address issues in open access policies and the integration of nonprofit workflows into scholarly publishing. Although preprints emerge as an equalizer to leverage its potential, the landscape is complex. In the past three decades, different regions strived for distinct things to promote open access across the national, state, institutional, and sectoral levels, as is advocated in Africa, China, and South Asia. The efforts of SciELO in Latin America; a radical open access program, Plan S, as was announced in Western Europe by research funders in 2018 and the United States of America (USA), are calling for global action towards more inclusive, open, and multilingual scholarship [15,16].

Nonetheless, open source technologies and open access movements necessitated the retooling of existing scholarly processes towards openness. Although open access publishing started to grow, some leading publishers, such as commercial, learned societies, and university press had actively opposed the growth of OA [17], until they found a way to transform it into a new business model and they were cautious to take up the OA model of publishing. Consequently, the period of 2006–2017 reported the high growth of OA mega-journals, which focused on the scientific trustworthiness and soundness, eschewing judgment of novelty or importance. Open access journals, PLOS ONE and Scientific Reports, are dominating this now [18,19].

Breaking the conventional boundaries, digital scholarly publishers have flourished innovatively with a wide variety of repository solutions and open journal publishing platforms, testing a range of open access publishing models. This is to achieve Gold open access (OA at the publishing source), Green open access (self-archiving), and Diamond open access (gold, but explicitly, with no article processing charges), as indexed by the Directory of Open Access Journals, which also lists the OA journals that charge article processing charges. As open access publishing models, licensing options, and infrastructures are getting larger, the data and resources enrich the web towards building open data. When the existing complexities of proprietary software, commercial publishers, and paywalled content is widespread, then the preprints entered to disrupt the scholarly communication system, thus making the vast amount of unpublished data and scholarly content available, regardless of peer review process. Preprints—as a leveler—enrich the scholarly web on top of the existing scholarly resources for discoverability. They do so by allowing access to not yet printed versions, timestamping ideas and findings, and adds meaning to interconnect people, concepts, and applications [20]. Therefore, preprints play a larger role in scholarly publishing strengthening the infrastructures of web through linked data, scholarly-rich content, and applications.

2.1. Rethinking Research Impact Metrics

As countries, institutions, and research communities compete on the global stage to measure and evaluate their national, institutional, and research outputs, various outcome and metrics-based research frameworks assess different research activities and performance. Some of those key areas are science and technology indicators, patents, bibliometrics, citations, rankings, research and development factors, measurements for innovation, and metrics for assessing the quality of scientific outputs. However, there is an increasing need to support research artefacts to be as inclusive as possible, going beyond research papers to preprints, software, codes, posters, media, and datasets. Scholarly activities, such as teaching and public outreach should also be included. Again, it is largely debated that the benefits of research impact should rise above academia on economy, society, public policy, human development, and the environment. This refers to the strategy, resources, and the infrastructure supporting the research, as adapted in the United Kingdom (UK) Research Excellence Framework—currently assessing the excellence of research in higher education institutions in the UK [21]. This is more important for

understanding what constitutes scholarly impact—when literature obsolescence and non-citation is rife, even with journals that maintain an impact factor of five [22,23].

Journal Impact Factor, CiteScore, Scimago Journal Rank, Source Normalized Impact per Paper metrics for journals and h-index, i10-index, and s-index for authors have determined and built a reputation of scientific productivity and the research impact of digital scholarship [24]. However, there is a growing demand for other kinds of metrics, such as at the article-level and the author-level—having their own merits beyond Journal Impact Factor, which is an aggregate of citation count for a journal in which the work is published [25]. Though academics and scientometricians have developed many metrics to measure the scientific output, whether the metrics work, fair, or overused need evaluation, as citation counts have less than one percent of usage for an article [26]. Many of the metrics that exist for measuring journal quality necessitate a paradigm shift to measure author-level metrics, which essentially captures the citation-related data and the connectivity-related metrics of authors [27]. Many metrics are still focused on published, peer reviewed articles as a primary output. However, the point is—with preprints and a wider diversity of processes and outputs—this demands new metrics beyond those for traditional outputs to be developed; but also, they must be applied in a responsible manner. The Web also opens up a whole field of additional context to explore things and hence a more 'Contextualized Metrics' is required for measuring those.

Moreover, defining impact in various contexts becomes extremely challenging at the academic, economic, and societal levels—given that the way the traditional metrics used for evaluation are deeply flawed [28]. For example, that they are being misused beyond their original intention (for example: Journal Impact Factor), deeply unscientific, and mostly operated by commercial entities and often being incredibly biased in different dimensions. Citation rates, journal ranks, and impact factors are inherently hierarchical and hence the institutionalization of them as a scientific impact assessment tool has unintended consequences of negative effects [29,30]. It is further found that the methodological quality and reliability of published research works in several fields may be decreasing with increasing journal rank [30,31]. Supporting new methods in data and scholarly publishing, the open research community must encourage publishing null results or failed experiments, against a growing body of evidence, questioning the conventional forms of impact assessment, which insist on quantifying the research outputs and they cannot capture diverse, wide-ranging, and inclusive research impact [32]. Moreover, the relevancy of citations and impact factor is widely questioned for their role in problem-solving and societal impact [33]. For a long time, open access has seen a great push through academics, policy making, science communication, and so on, and preprints add to this environment as an additional layer that will further enrich the scholarly ecosystem.

Reimagining open infrastructures and metrics, this article aims to situate preprints in the emerging research ecosystem, establishing that disciplinary-centric and public preprint repositories have been on the rise in the last two decades or so. As preprints become mainstream, research publications coming out from highly to moderate novelties of incremental, supportive, or confirmatory results, and their supplementary data will more visibility benefit [34]. Research communication, academic outputs, and scholarly artifacts have diversified in many ways and they are available for various communities of practice—transcending disciplinary boundaries of research. Scholarly communication is evolving and diversifying. We need to rethink our metrics and evaluation systems based on this in the rapidly changing landscape. Research outputs are more than journal articles, and so measuring their impact should go beyond them, including prepublication outputs. Their credibility, impact, and value should be measured through heterogeneous metrics, which calls into question the whole idea of trying to measure scholarship. Are metrics appropriate? Or is qualitative assessment needed? Is such assessment even operationally better than randomness?

2.2. Growth of Preprint Repositories: From arXiv to ESSOAr

As exhibited in Table 1, the rapid growth of preprint repositories prompted the scholarly communities to define what constitutes a preprint, when there is no clear consensus on what they

are. An examination of definitions by some of the preprint repositories reveals that they are "draft, unpublished, incomplete, or unedited final versions of papers, maybe work in progress and not typeset". In one of the early attempts, Gunther [7] distinguished the preprints from an electronic publishing and e-print server perspective, referring to them as "'pre-peer-review' or 'pre-submission' documents" in a Guest Editorial in 2000. According to PeerJ Preprints, a preprint repository [35], it is described as "a draft of an article, abstract, or poster that has not yet been peer reviewed for formal publication". Many scholars have attempted to define exactly what a preprint is—distinguishing preprint as a scholarly item that is based on subject to evaluation as in pre- and postprints and preprint server as infrastructure. Neylon [36] proposed a model that distinguishes the preprints by "characteristics of the object, its 'state' from the subjective 'standing' granted to it by different communities". Rittman explains preprints as "a piece of research made publicly available before it has been validated by the research community. That is to say, some output that follows the scientific process but has not yet been peer-reviewed for journal publication [37]". However, Tennant et al. [8] propounded a definition of what is a preprint that is based around its peer review status, which is in line with the Sherpa/Romeo description:

- Preprint: Version of a research paper, typically prior to peer review and publication in a journal.
- Postprint: Version of a research paper, subsequent to peer review (and acceptance), but before any type-setting or copy-editing by the publisher. Also, sometimes called a 'peer reviewed accepted manuscript'.
- Version of Record (VOR): The final published version of a scholarly research paper after undergoing formatting (and any other additions) by the publisher.
- e-Print: Version of a research paper posted on a public server, independently of its status regarding peer-review, publication in print, etc. Preprints, postprints, and VORs are forms of e-Prints.

Publishers are accepting preprints for peer review in journals, even if they are available in preprints repositories that are submitted in parallel. They are submitted to the preprint repositories without peer review, being free of cost by authors to solicit feedback from peers, perhaps being often submitted to a journal later for peer review and subsequent publication. arXiv is a preprint repository that was established for high energy physics in 1991. However, other disciplines took more time to realize the potential of using preprint repositories and the best practices of early dissemination of research works online to maximize the research impact [8].

Table 1. Growth of preprint repositories, 1991–2018[1].

S. No.	Name of Preprints	Subject/Disciplines	Year Established	No. of Records as on 28 July 2018	Website
1	arXiv	Natural Sciences, Engineering, Economics, Finance and Computing	1991	1,421,596	https://arxiv.org
2	RePEc	Economics	1992	2,600,000	https://ideas.repec.org
3	SSRN	Social Sciences	1994	810,845	https://www.ssrn.com/en
4	E-LIS	Library and Information Science	2003	20,390	http://eprints.rclis.org
5	bioRxiv	Life Sciences	2013	25,632	http://www.biorxiv.org
6	PeerJ Preprints	Biological, Medical, Environmental and Computing Sciences	2013	4129	https://peerj.com/preprints

Table 1. *Cont.*

S. No.	Name of Preprints	Subject/Disciplines	Year Established	No. of Records as on 28 July 2018	Website
7	OSF Preprints	Natural Sciences, Technology, Engineering and Social Sciences. Arts and Humanities	2013	3170	https://osf.io/preprints
8	MDPI Preprints	Natural, Engineering, Social Sciences and Arts and Humanities	2016	5095	https://www.preprints.org
9	ChemRxiv	Chemical Sciences	2016	9910	http://www.chemrxiv.org
10	ESSOAr	Earth Sciences	2018	149	https://www.essoar.org

Figure 1 shows the growth of preprints in life sciences from 2007 to 2018, which are reporting a high number of submissions. Life sciences established more preprints, such as arXiv q-bio, which is a quantitative biology archive and it has been part of arXiv, publishing preprints since September 2003. Cold Spring Harbor Laboratory, a nonprofit, launched bioRxiv, which is a biology preprint repository, in November 2013. In April 2013, PeerJ Inc. launched its PeerJ Preprints that covered biological, medical, and environmental sciences.
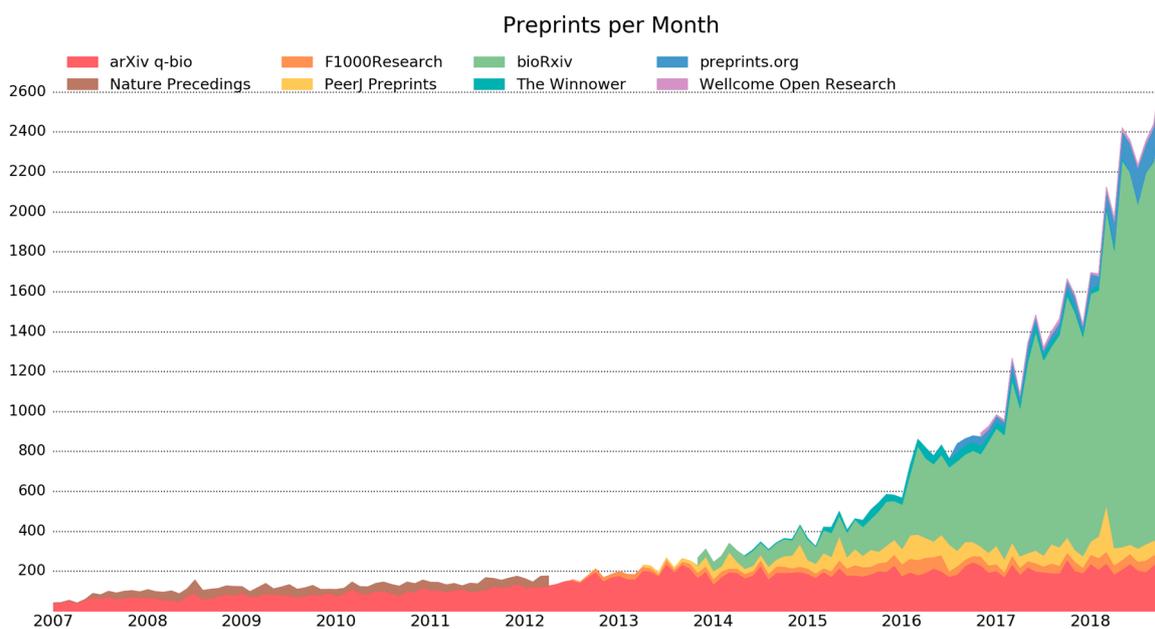

Figure 1. Monthly preprints added in November 2018 is 2509. Source: PrePubMed.org.

Journal Impact Factor indicates the quality of journals through citation metrics, though measuring scholarly and societal impact is more important [38]. Hence, a new paradigm shift is that publications should not be subjective of impact, novelty, and interest, but that is based on scientific and methodological soundness or objective. In other words, many journals have emerged to report on what "scientific literature might gradually become less biased against negative or null results and it will be less dominated by the trends and 'hot topics' of the day [39]", for which preprints provides the access to check the prepublication of results prior to peer review. The Journal Impact Factor of journals

---

1   Record statistics are collected from their respective websites, except bioRxiv and for this, data is collected from OSF Preprints.

and their peer reviewing process are found to be excruciatingly slow (typically 85–150 days or longer) and the decision-making process is invariably slow—affecting the careers of early researchers with no recognition until the research is published, though they foster international collaboration and global reach [40]. Among the other criticisms that are widely conceded among researchers is that science and knowledge are measured by numerical ranking systems, which first makes researchers pursue the rankings over research [41]. Muller argues against these counterproductive conventions on the performance evaluation and metrics, calling it 'metric fixation' [42]. Preprints break these conventions, giving advantage to research publications for their merits and research impact, as they become openly available as early prepublication outputs, but they also are independent of any specific journal venue at the point of sharing.

Preprint repositories play a vital role in the dissemination of research artifacts for impact and making them visible to connect with their audience. This concerns the material culture of academic reading and writing, which may be transient in social media communication, but calls for reuse, credit, and replication in an open research ecosystem with data, code, citations, and software. Additionally, scholar identity has grown alongside the technological innovations for technology-influenced scholarship through participatory technologies in the public sphere. Increasingly, academics, practitioners, and researchers [43] tend to communicate their research using social media as a utility in the research landscape and lifecycle—as a digital opportunity to learn tools and techniques and then apply them for research communication in the changing research landscape. There is a growing trend in publishing for unrefereed preprint repositories; writings blog posts or field notes online; creating infographics, data visualizations, and publishing research data in data journals; making podcasts, creating videos/images, photo-essays, and overlay journals—which all diversify scholarly communication [44]. Furthermore, the scholarly communication activities and processes on informal channels boost interaction, collaboration, seeking, citing, publishing and disseminating in orthodox, moderate, and heterodox use scenarios [45]. A few examples are The Conversation Global [46] and Policyforum.net [47], the online independent news platforms that are run by research communities. Using these platforms, journalists, scientists, academicians, and researchers primarily aim to communicate scholarly information for the lay audience. In this, preprints help journalism, promoting transparency and science communication for the public.

3. Methods

For the purpose of this study, a sample of ten preprint repositories was chosen, which were based on their history, popularity, and disciplinary diversity. This was a combination of preprints (that go on to be published or not), postprints, final published articles, datasets, working papers, which were all examined of their salient features, disciplinary focus, and the number of records available between March and September 2018 (See Table 1). As a case study of preprints, the research was conducted in two stages. First, was to highlight their principal features, such as system architecture, persistent identifiers and registries, disciplinary focus, research data management, peer reviewing models, infrastructures, and metrics.

The second stage consisted of using indicators in depth for analysis: software and open source technologies used, standards and protocols adopted, knowledge organization systems applied, interoperability and open licensing options, indexing and aggregating agencies involved, metrics and peer reviewing processes, community standards, and web 3.0 applications that are available. Subject and disciplines of preprints, such as life sciences, technology, engineering, and social sciences, were included for this study. Additionally, management aspects, such as funding agencies and whether the preprints were supported by for-profit corporations or nonprofits, their advisory committees, code of conduct; management of digital object identifiers, submission guidelines, copyright policies, and publishing workflows were investigated. Subsequently, a comparative analysis at the site and record levels were performed in order to synthesize the results and discussions further.

4. Results

4.1. Comparative Features of Preprint Repositories

The results that are presented below are in eight sections. Key findings are categorized based on themes, such as: System architecture, Persistent identifiers and registries, Disciplinary focus and management, Interoperability and open licensing, Indexing and aggregators, Knowledge organization systems, authority control and subject categories, Metrics and open reviews, and Community standards.

4.1.1. System Architecture

System architecture refers to the database structures, hardware, and software that are used to set up a preprint repository. As shown in Table 2, there was a limited number of software solutions available when preprint repositories were started, so legacy preprint repositories, such as arXiv and RePEc, are migrating to use digital repository software—Invenio and EPrints, respectively, to integrate new applications, such as DOIs, ORCIDs, and Altmetric. E-LIS is a public preprint repository in library and information science run on DSpace. Though DSpace and EPrints dominate globally in repositories development, OSF Preprints and Figshare are new entrants for repository solutions. Few preprints are building application programme interfaces (APIs) to build robust features and accommodate services from other programs. It is found that, out of ten, four preprints repositories have Open APIs, which are RePEc, MDPI Preprints, OSF Preprints, and Figshare. OSF Preprints is an aggregator from across almost all of the other servers. It also links to other services, such as Figshare or GitHub, and it is virtually unlimited in scope of what can be 'attached' to preprints and offers local storage. It uses SHARE, which is a community open-source initiative suite of technologies. Out of the ten preprints that were evaluated, four preprint repositories are using custom proprietary systems, which could not be identified, as listed in column 2 of Table 2 in infrastructure. Managing research data has become an integral part of system architecture, where multiple files are supported from word processors to datasets in variety of formats, such as LaTex to Zip, for preservation, and essentially all of the preprints support that [48].

4.1.2. Persistent Identifiers and Registries

Persistent identifiers help to provide perpetual IDs for digital objects to identify and retrieve them. Most preprint repositories have identifiers, such as article IDs, URIs, and Handle system for publications, which make the records unique, identifiable, persistent, and retrievable (See Table 2). Many of them are cross-linked and directed to the DOIs of the article, where the latest version of the article is available as permalinks. An example of arXiv ID is arXiv:hep-th/9603067, where hep-th stands for High Energy Physics—Theory and 9,603,067 is the unique record number. Another example of the RePEc identifier handle is: https://ideas.repec.org/p/hhs/cesisp/0277.html, where hhs:cesisp denotes the Centre of Excellence for Science and Innovation Studies, Royal Institute of Technology, Stockholm, Sweden, followed by the unique record number: 0277. Among the ten preprint repositories analyzed, seven are found to be using Crossref's DOI services for preprint records. Crossref dominate DOIs among preprints. At OSF Preprints, each project is assigned a globally unique identifier, or GUID, though DOIs are used as well. DOIs versioning was found to be unique with ChemRxiv, MDPI Preprints, and PeerJ Preprints for version control. Further, DOIs assigned to supplementary data, file, code, and dataset enable them to be citable as well. Moreover, one of the important features found is registries, which records various projects to make them available publicly as crucial content providers and helps in avoiding the duplication of studies. OSF Registries has 274,910 registrations of research studies of systematic reviews and meta-analyses in clinical psychology and medicine that are cross-searchable with Research Registries and ClinicalTrials.gov registries.

4.1.3. Disciplinary Focus and Management

The examination of preprints history and growth reveals that disciplinary focus has been one of the major factors for establishing them. Since the need for sharing the scholarly research arose in different settings—laboratory, academic, research, and practice—preprints were created and supported by diverse disciplinary areas (see Table 1). This also ties into the social differences and norms between different research communities, wherein the replication, reproducibility, and methodological approaches vary greatly among different domains, especially when preprints have a 'state' from the subjective 'standing' granted to them by different communities of practice [36] (p. 4). arXiv was started with physics, but it soon expanded to covering quantitative biology, astronomy, computer science, and mathematics. Biology has been quite conventional, but in recent years it has been reporting high number of submissions in bioRxiv and PeerJ Preprints (see Figure 1). In the last year, there are more than 20 disciplinary-based preprint platforms that have emerged. See here the disciplinary prerints, which are backed by Centre for Open Science. Its other country-specific examples are INA-Rxiv, Arabixiv, and AfricArxiv, which are committed for Indonesia, The Arab states, and Africa, respectively, to promote open science. Moreover, managing preprint repositories are not only solely resting with public institutions or government, but also by different agencies that are funded by nonprofits and for-profit companies [8]. arXiv is hosted by Cornell University Library, RePEc by Munich University Library and consortia, E-LIS is supported by AIMS, FAO, and University of Naples Federico II, Naples–Centralino, bioRxiv is hosted by Cold Spring Harbor Laboratory, and OSF Preprints by Centre for Open Science, MDPI Preprints by MDPI, which are nonprofits. ChemRxiv is collaboratively managed by American Chemical Society, German Chemical Society (GDCh) and the Royal Society of Chemistry, UK, and ESSOAr by the American Geophysical Union are learned societies. These agencies are backing the growth and development of preprints in key disciplinary areas. However, SSRN that are acquired by RELX Group in May 2016 and both PeerJ Preprints and MDPI's preprints.org are services run by commercial publishers, meaning that preprint servers are seen as a key part of business models (see Table 2). This is part of dangerous move from some publishers into controlling the entire research workflow and it is symptomatic of a highly dysfunctional scholarly publishing market [49].

4.1.4. Interoperability and Open Licensing

Out of ten repositories, arXiv, RePEc, and OSF Preprints are found to be interoperable and they support a whole range of integrated search features, such as cross-searching of content including abstract, full text of articles across multiple repositories, and owned by different content providers. ChemRxiv run by Figshare is a proprietary platform, but it has a unique model where all of the available content is shown on the single portal and is owned and provided by various institutions worldwide at https://figshare.com. Creative Commons license is found to be predominantly used by many of the preprint repositories for licensing to allow the reusing of the content and data. However, the degree of freedom varies across preprint repositories. arXiv uses the following license types, which goes from the most accommodative to restrictive: Attribution 1.0 Generic (CC BY 1.0), Attribution 4.0 International (CC BY 4.0), ShareAlike (SA), NonCommercial (NC), and some even have CC BY-NC, CC BY-NC-SA types [50,51] . E-LIS, PeerJ Preprints, and MDPI Preprints use Attribution 4.0 International (CC BY 4.0), which allows for the sharing and adapting of works. This implies that there is a "unrestricted use, distribution and reproduction in any medium and the original work is properly cited" [52]. ChemRxiv allows for Attribution-NonCommercial-NoDerivs CC BY-NC-ND to "download works and share them with others as long as they are credited, but they can't change them in any way or use them commercially" and applies embargo, keeps confidential files, generate private links, and reserve DOIs, and also accepts any file format up to 5 GB. ESSOAr follows the Attribution 1.0 Generic (CC BY 1.0) License. RePEc does not state any one of the above licensing options, while SSRN allows this: papers by the copyright owner or that have the copyright owner's permission are permitted to post under publishing agreement or the publisher's copyright policies or institution's license agreement or under a Creative Commons license. Preprints to peer reviewed journals portability is also worth mentioning

here for bioRxiv preprints, which become easy for authors and currently this service is available for 107 biology journals.

4.1.5. Indexing and Aggregators

It was found that all of the preprints are indexed and then aggregated by commercial, institutional, data repositories, and databases, which are bibliographic, aggregating, and depositive in nature. RePEc has its own indexing platform, called IDEAS, which is a comprehensive bibliographic database in economics, available for free, which indexes over 2,700,000 items of research, indexed in EconLit, EconStor, Google Scholar, Inomics, OAISter, OpenAIRE, and EBSCO. E-LIS provides seek option for references tht can be retrieved in Google Scholar. bioRxiv preprints are indexed by the following services: Google Scholar, CrossRef, Meta, and Microsoft Academic Search. MDPI Preprints are indexed by Europe PMC, Google Scholar, Scilit, Academic Karma, SHARE, and PrePubMed. PeerJ Preprints are indexed in Google Scholar. Crossref provides DOIs for preprints and DataCite primarily works for providing persistent identifiers for all kinds of research data and it is integrated with ChemRxiv. Though many of the abstracting and indexing databases—OpenAIRE, ResearchGate, Academia.edu OAISter—index preprints, there is no established standards available for preprints, hence no usage statistics are reported, unlike peer reviewed journals that report COUNTER-complaint usage statistics.

4.1.6. Knowledge Organization Systems, Authority Control and Subject Categories

For the authority control of authors, arXiv, bioRxiv, MDPI Preprints, and ESSOAr use the endorsement of authors through ORCID. All of the preprint repositories display author-supplied keywords and tags and the browsing of preprints by subjects/disciplines is prevalent. In addition, many of the preprint repositories display the subject/discipline category and they are based on which of the preprint categories are displayed. For example, at bioRxiv, the category of articles that are submitted are New Results, Confirmatory Results, or Contradictory Results vis-à-vis differentiate the conventional papers Research, Opinions, Reviews, Technical, Concepts, or Case Studies published in social science preprints, which are SSRN, OSF Preprints. PeerJ Preprints, arXiv, bioRxiv, and OSF have advanced features, such as article versioning, adding links, and comments. It also has faceted the browsing of its collections by manuscript type; filtering articles by entity, which are references, questions, answers, figures, and by published date and subjects. Really Simple Syndication (RSS) is popular among the preprints for having syndicated updates on new articles, subject areas, besides social media. RePEc and SSRN are using JEL Classification Codes, whereas E-LIS uses the JITA Classification of Library and information Science to classify the scholarly literature. There are no standardised metadata schema adopted by preprints, except the Dublin Core metadata schema followed in DSpace at E-LIS, and the rest of the preprint repositories use a more simplified metadata input formats.

4.1.7. Metrics and Open Reviews

Since citations data are quite distributed in various databases by their journals coverage, they need to be aggregated from multiple platforms, such as Crossref, Scopus, and Web of Science, for use. Google Scholar's citation data is essentially found to be the superset of Scopus and Web of Science databases [53]. All of the preprints provide citation tools support to export the references in multiple file formats that are supported by various platforms of reference management software. Among all of the preprints, arXiv reports a unique subject wise submissions, access, and download details—daily, monthly, and institutional wise. RePEc reports the number of citations, downloads, and abstract views; top-level metrics for institutions, regions, authors, and document types. Also, it reports statistics by research items, series and journals, authors, and institutions [54]. SSRN preprints have report on institutional level data for downloads, abstract views and rank of papers, authors, and organizations, besides integrated PlumX Metrics, which is an alternative metric platform of Elsevier. See here, an example [55]. PeerJ Preprints reports unique article-level metrics, which are grouped as social

referrals by social media and top referrals, which are essentially search engines, bookmarks, URLs, and email alerts. See the example in Figure 2 [56].

Altmetric platform is integrated with bioRxiv, MDPI Preprints, ChemRxiv, PeerJ Preprints, and ESSOAr preprints—aggregating social media metrics. PeerJ Preprints reports its visitors, downloads, and views; OSF Preprints shows the downloads count; MDPI Preprints exhibits the views, downloads, commenting options in public and private, and also provide rating options; E-LIS shows the monthly and yearly downloads in the graph at the article-level and repository-level, and also other statistics that are available are the most downloaded items, top authors. ChemRxiv shows views, downloads, and citations; ESSOAr reports the download counts. MDPI Preprints allows the viewing of reviewer comments through Publons, which is a peer-review profile platform and the only one to do so among the preprints, while PeerJ Preprints provides open feedback, Q&A, and linking options to engage with readers and reviewers.
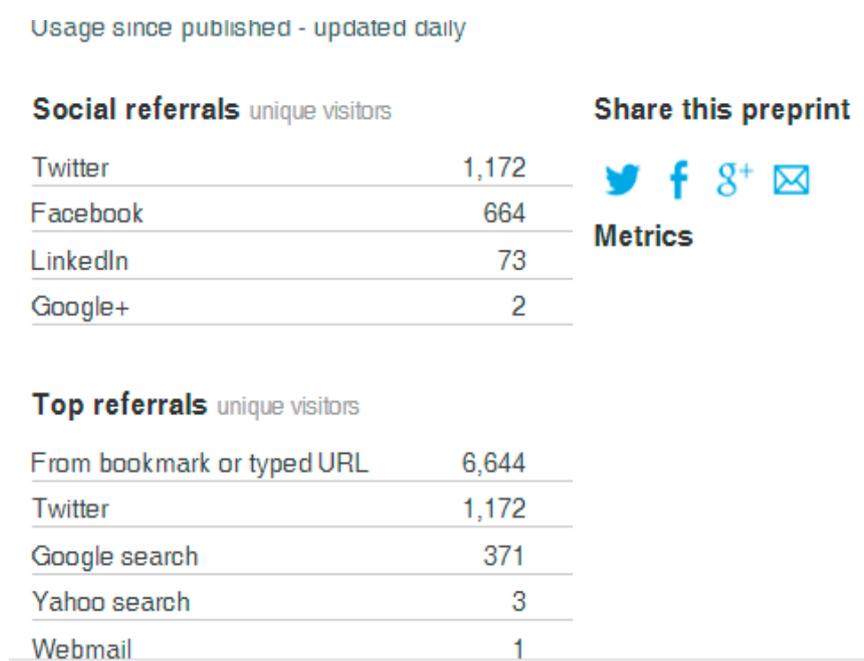

Figure 2. An example of article level metrics at PeerJ Preprints.

4.1.8. Community Standards

Community standards help to develop, integrate, and steer the standards, protocols, and codes of conduct to take the initiatives (systems, software, and programs) to the wider of community of committers, developers, and funders for the strengthening of open access initiatives, open source technologies, and scholarly publishing. This refers to the standards, which are free and open source software, projects, and communities for interoperability. One of the important metadata harvesting interoperability protocol is the Open Archives Initiative—Protocol for Metadata Harvesting—v.2.0 (OAI-PMH v2.0). arXiv, RePEC, and e-LIS are compliant to this protocol to support the harvesting of records from other digital repositories and to set the trends for community standards in building open archives [57]. For the standards of software and operating system, arXiv uses GNU and MIT License. RePEC has GNU and Guildford Protocol. E-LIS adopted Open Data Commons—Open Database License. As much as preprints operate on open community standards, managing them needs advisory boards, funding strategies, and steering committees to take the initiatives forward, which are further discussed in the Table 3. None of the preprints explicitly display code of conduct.

Table 2. Comparative features of preprint repositories[2].

Preprint	Infrastructure			Host/Funding Agency	Open Technologies/Protocols Used	License Name	Knowledge Organization Systems	Web 3.0 Applications	Metrics	Nonprofit/for Profit Body
	Software Name	Open Source/Proprietary Software	Identifier/Managing Agency							
arXiv	GNU/Invenio	Open source	arXiv:1806.07477/arXiv	Cornell University Library, Simons Foundation and by the member institutions	MIT License. OAI_PMH v2.0 (OAI2)	Non exclusive-distrib/1.0/. (CC BY 4.0), (CC BY-SA 4.0), (CC BY-NC-SA 4.0), (CC0 1.0)	Keywords, subjects and authority records.	RSS, Twitter, Bookmarks, Email alerts, annotation, Blog, Citation tools	Subject wise submission, access and downloads details—daily, monthly, institutional-wise	Nonprofit
RePEc	GNU/EPrints	Open source	RePEc:hhs:cesisp:0277. RePEc short ID for authors: pzi1/RePEc	Munich University Library and members from 99 countries. Research Division of the Federal Reserve Bank of St. Louis	Guildford Protocol. OAI-PMH.	-	JEL Classification	RSS, Twitter, Facebook, G+, Reddit, StumbleUpon, Delicious, Email Alerts, Blog	Citations, downloads, and abstract views. Top-level metrics for institutions, regions, authors and document types	Nonprofit
SSRN	Custom	Proprietary	10.2139/ssrn.1926431/Crossref	RELX Group	-	-	JEL Classification	Facebook, Twitter, CiteULike, Permalink, Blog	Downloads, abstract views, PlumX metrics. Ranks for paper, author and organizations	For-profit
e-LIS	DSpace	Open source	http://hdl.handle.net/10760/32727/Handle	AIMS, FAO and University of Naples Federico II, Naples—Centralino, Italy	Open Data Commons Open Database License. The Open Archives Initiative and OAI 2.0	-	JITA Classification	-	Downloads	Nonprofit
bioRxiv	HighWire	Proprietary	/10.1101/328724/Crossref	Cold Spring Harbor Laboratory, Cold Spring Harbor, NY	-	CC-BY 4.0 International license	Subjects	RSS, Twitter, Facebook, G+, Alerts, digg, reddit, CiteULike, Google bookmarks, Comment system, Citation tools	Altmetric	Nonprofit

Table 2. *Cont.*

Preprint	Infrastructure			Host/Funding Agency	Open Technologies/Protocols Used	License Name	Knowledge Organization Systems	Web 3.0 Applications	Metrics	Nonprofit/for Profit Body
	Software Name	Open Source/Proprietary Software	Identifier/Managing Agency							
PeerJ Preprints	Custom	Proprietary	10.7287/peerj.preprints.26954v1/Crossref	PeerJ, Inc.	-	CC BY 4.0	Keywords and discipline wise browsing	Twitter, Facebook, G+, Alerts, Citation tools, versions of record	Visitors, downloads, views and Altmetric	For-profit
OSF Preprints	OSF/SHARE	Open source	10.31219/osf.io/zuwnr/Crossref	Center for Open Science	-	CC-By Attribution 4.0 International; CC0 1.0 Universal	Disciplines and tags	Twitter, Facebook, LinkedIn, Alerts, Citation tools, Annotation, Highlights	Downloads	Nonprofit
MDPI Preprints	Custom	Proprietary	10.20944/preprints201805.0375.v1/Crossref	MDPI	-	CC BY license	Disciplines	Facebook, Twitter, LinkedIn and Email alerts. Bookmarks in CiteULike. BibSonomy, Mendeley, Reddit, Delicious, Citation tools and Publons	Views, downloads, comments and Altmetric	For-profit
ChemRxiv	Figshare	Proprietary	10.26434/chemrxiv.6744440.v1/Crossref	American Chemical Society, German Chemical Society (GDCh) and the Royal Society of Chemistry	OpenAPI Initiative. MIT, GPL, GPL 2.0+, GPL 3.0+.	CC BY-NC-ND 4.0, CC BY 4.0, CC0	Subject categories and keywords	Facebook, Twitter, LinkedIn, G+, Email alerts	Views, downloads, citations and Altmetric	Nonprofit
ESSOAr	Atypon	Proprietary	10.1002/essoar.10500004.1/Crossref	American Geophysical Union	-	CC-BY-NC-ND, CC-BY-NC, or CC-BY	Keywords	Facebook, Twitter, LinkedIn, Google+, Reddit, Email alerts	Altmetric and downloads	Nonprofit

---

[2] Technical features of open infrastructures and metrics used at preprint repositories are examined at the article and site level on the respective website.

Table 3. Management of preprint repositories.

Preprint Name	Managed by Individuals/Organizations	Steering Committee/Advisory Board	Submission Guidelines	Subscription/Membership	Forum/Q&A	Companion Website/Social Media
arXiv	Cornell University Library with arXiv Scientific Advisory Board and the arXiv Sustainability Advisory Group	Member Advisory Board	Yes	No subscription required, but runs on voluntary contributions with active institutions	Yes	Yes
RePEc	Munich University Library and members from 99 countries. Research Division of the Federal Reserve Bank of St. Louis	RePEc coordinators and volunteers for editing, hosting and support	Yes	No	Yes	Yes
SSRN	RELX Group	Network Directors	Yes	Free to use, however, subscription is available	Yes	Yes
E-LIS	AIMS, FAO and University of Naples Federico II, Naples–Centralino	E-LIS Admin Board and Country Editors	Yes	No	Yes	Yes
bioRxiv	Cold Spring Harbor Laboratory	Advisory Board	Yes	No	Yes	Yes
PeerJ Preprints	PeerJ, Inc.	Academic Boards, Advisors, Editors	Yes	No	Yes	Yes
OSF Preprints	Center for Open Science	Advisory Group	Yes	No	Yes	Yes
MDPI Preprints	MDPI	Advisory Board	Yes	No	Yes	Yes
ChemRxiv	American Chemical Society, German Chemical Society (GDCh) and the Royal Society of Chemistry	No	Yes	No	Yes	Yes
ESSOAr	American Geophysical Union	Advisory Board/Editorial Board	Yes	No	Yes	Yes

5. Discussion

5.1. Preprints for Building Scholarly Infrastructures and Metrics

Preprint repositories are becoming pivotal at the intersection of scholarly web and open infrastructures, adopting the role of developing their pathways towards a dynamic research ecosystem in the advent of open technologies, such as persistent identifiers, open data harvesting and protocols, integrated data aggregators, and various discovery layers. Since the preprints make the content available, building infrastructures around them is central to the build, scale, and measure of such projects. Interoperability and crosswalking between them is critical for discoverability and citability of scholarly data. Though, some funders have guidelines, for example, at NIH, there is no general standards or established principles for preprints publishing. This is important for researchers, publishers, infrastructures, and service providers to have coherent workflows and the integration of multiple data sources and open infrastructures into unifying platforms to collect evidence regarding research impact, which will improve the demonstrated reliability [27]. Building novel metrics upon preprint infrastructures help with the quality assurance of scientific outputs, however, has its limitations. For example, alternative metrics say little about the quality of a paper and the kinds of impact, but more about its popularity [58]. Hence, the alternative metrics for alternative scholarly infrastructures need to be designed wisely to prevent adverse effects, as in how some of the conventional metrics are misused, such as the Journal Impact Factor.

Embracing Findable, Accessible, and Interoperable, Reusable (FAIR) principles for scientific data management and stewardship focuses on the reuse of scholarly data, specifically enhancing the ability of the machines to automate the reusability of data. The potential impact and good practices of using FAIR principles amongst the UK academic research community has been found to exist and be continually improving, despite disciplinary differences. However, it is found that there is lack of understanding of FAIR data and principles; need for investments in the development of data tools, services, and processes to support open research; adopting FAIR principles across the broad coordinating activities and policy development at cross-disciplinary, national, and international levels [59,60]. DataCite has been steering on persistent identifiers for research data citation, discovery, and accessibility, while also emphasizing the measurement of grants and the impact that is made by funding agencies [61,62]. Hypothesis has been experimenting with open annotation use cases on preprints and discussed the burden of moderating (editorial and site), identity, and versioning among the preprint repositories [63]. OSF Preprints has been experimenting on open annotations. At the nexus of building open scholarly infrastructure-metric in the broader scholarly communication system, preprints push for developing and integrating evidences of the impact for evaluating research and researchers with the emergent systems below:

1. Data Infrastructures and Metrics—curates resources, metadata, and datasets that make the data of scientific publications discoverable, reusable, and citable involving the seamless integration between data and researchers across the research lifecycle, connecting human and technical infrastructure for open research. Some examples include Dryad Digital Repository, DataCite, and institutional repositories.

2. Persistent Identifiers (PIDs)—connects not only digital objects, but also people, events, organizations, and vocabulary terms to achieve the persistence of digital resources. Persistent identifier infrastructure facilitates the scientific reproducibility and the discovery of open data, providing long-term access to research artifacts (software, preprints, and datasets) and interoperability. For PIDs to grow, building and strengthening legacy PIDs, provenance, preservation, and linking of scholarly works and an ecosystem of co-existence are critical. Few cases of PIDs are Digital Object Identifiers, Archival Resource Keys, RRIDs, IGSNs, and ISBNs.

3. Authority Files—build and control the names of authors and organizations to share and validate the published data for vocabulary control. International Standard Name Identifiers, ORCIDs, ResearcherID and Virtual Authority International Files, and International Registry of Authors-Links to Identify Scientists are some examples.

4. OA Applications—includes a set of open applications that facilitate free, accessible, and reusable scholarly research by building layers of new functionalities, such as programs, extractions, extensions, and link resolvers to find open access and a full text of scholarly resources. Examples, including Unpaywall, Open Access Button, Kopernio, and Lazy Scholar help to find full text of publications. There are also platforms for showing the research impact of articles, authors, and software. Some examples are Impactstory and Depsy.

5. Open Citations Databases—create and expand on open repository of scholarly citation data for reuse, which mainly include citation links, citation metrics, and cited resources under open licenses. Some examples are OpenCitations, Dimensions.ai, and Lens.org.

6. Open Peer Review Systems—displays the pre- and post-publication track of reviews and comments made for peer-reviewed publications that are openly accessible. Peerage of Science, PubPeer, ScienceOpen, and Publons are a few examples where the reviews and comments of peer review is open for recommendation and social sharing.

Table 4 shows the common features that are found across all of the platforms. arXiv is cross-linked with the SAO/NASA Astrophysics Data System and INSPIRE—High Energy Physics databases. RePEc has many mirror sites that are hosted in 99 countries. SSRN has recommendations for related e-journals and papers while browsing. Though all of the preprints have embedded references in PDFs, RePEC collects citations data for all of its holdings through CitEc, which is a citation database and E-LIS has on site display of the references. Being at the nascent stage, preprint repositories are developing and integrating with some of the common infrastructures. Altmetric platform integration, having identifiers with Crossref's DOIs and open references, are the most implemented features in open infrastructures and metrics.

Table 4. Common features of open infrastructures and metrics.

Preprint Name	Google Scholar Integration	Publons/Open Reviews	Altmetric/PlumX Metrics	Crossref DOIs	Open References	Recommendations (Browsing Related Research)	Additional Site Integration/Final Publication Display
arXiv	Yes	No	No	No	Yes	No	Yes
RePEc	No	No	No	No	Yes	No	Yes
SSRN	No	No	Yes	Yes	Yes	Yes	Yes
E-LIS	Yes	No	No	No	Yes	No	No
bioRxiv	No	No	Yes	Yes	Yes	No	Yes
PeerJ Preprints	Yes	Yes	No	Yes	Yes	No	Yes
OSF Preprints	No	No	No	Yes	Yes	No	Yes
MDPI Preprints	No	Yes	Yes	Yes	Yes	No	No
ChemRxiv	No	No	Yes	Yes	Yes	No	No
ESSOAr	No	No	Yes	Yes	Yes	No	No

As listed in the Table 4 open references column, all of the preprints have references that are open in preprints; however, building open citations data is not freely available. Since not all of the citations are openly accessible for preprints that are being cited, building open citations remains the biggest challenge, as peer reviewed publications and their citations are invariably distributed within the Google Scholar, Crossref, Web of Science, and Scopus databases. Dimensions.ai and Lens.org are the

few open citations databases, facilitating the measurement of citations data, while remaining widely distributed to be discovered on the scholarly web.

5.2. Towards Building Sustainable Open Infrastructures with Preprints

Preprints drive demand for new scholarly metrics and infrastructures, having been part of the scholarly outputs, reporting preliminary results. Preprint repositories that are designed with open source software, technologies, and infrastructures become essentially sustainable [64]. Commenting upon the needs of open development as a socio-technological innovation towards open access, Chan [65] noted that "the term is a broad proposition that open models and peer-based production, enabled by pervasive network technologies, non-market based incentive structures and alternative licensing regimes, can result in greater participation, access and collaboration across different sectors... A key understanding of 'open development' is that while technologies are not the sole driver of social change, they are deeply embedded in our social, economic and political fabric. We therefore need to understand 'openness' within the context of a complex socio-technical framework". The collective action for scholarly communication necessitates the funding for infrastructure services to be interoperable, scalable, open, and community-based for open infrastructures as the potential funders and organizations look for demonstrable community-based services, like preprints supporting open research. SCOSS and the CoKo Foundation are notable here as promising initiatives in this space.

Hence, developing conceptual frameworks to support investors in infrastructures for open scholarship and in developing community capacity through the OA Sustainability Index becomes important. This is to take on initiatives, like preprints development, which are in hitherto under-represented disciplines and extending frontiers of open knowledge [66]. Sustainability of research ecosystem with research, education, and knowledge production components are crucial, as the implementation of preprint policies relies on the development of a fully-functioning OA infrastructure [67]. In order to build resilient open infrastructures that are inclusive and sustainable systems, creating, sharing, and disseminating knowledge is important in scholarly publishing for workflow integrations, metadata reuse, and publisher integration with the research lifecycle. In support of open and collaborative science, Chan [65] further argues that "open approaches to knowledge production have the potential to radically increase the visibility, reproducibility, efficiency, transparency, and relevance of scientific research, while expanding the opportunities for a broad range of actors to participate in the knowledge production process... openness is not simply about gaining access to knowledge, but about the right to participate in the knowledge production process, driven by issues that are of local relevance, rather than research agendas set elsewhere or from the top down". This is where preprint repositories are proven to be a disruptive development towards building public science. Scientific publishers, research enterprises, and funding agencies are at a deflecting point where research systems should be built, designed, and disseminated inherently openly, and developing preprint services provides just that opportunity for scientific communities [68].

We need to strengthen and expand the community and institutional role in managing preprints and their development. For that, we should redefine frameworks to overcome barriers and challenges in establishing open infrastructures for scholarly communication networks, so that open research principles are inbuilt in our research ecosystem, production processes, and in scientific publishing. The Open Science by Design report that was released by the United States (US) National Academies of Sciences, Engineering, and Medicine is a step towards that [69]. Research is global and scholarly communities need interoperable hubs, interlinking data, and infrastructures supporting information exchange across repositories with standards, metadata schema, and semantic interoperability, as there is lack of standards for aggregating data that is used across platforms [70]. Preprints are disrupting the scholarly communication system and many leading publishers are slowly participating in the process—supporting, accepting of, and archiving in preprint repositories. However, some of the important challenges are inconsistent metadata schema in data harvesting, supporting multilingual systems, a lack of standards in integration, and protocols for aggregating data and implementing

them across platforms in version control, deduplication, and digital preservation. In strengthening the open infrastructures and metrics, preprints add to the ever-growing repository types and artifacts that are indexed and mined by indexers, aggregators, and search engines; built into registries, authority files for authors and organizations, and vocabulary control of subject terminologies. In this, all of the stakeholders—publishers, governments, funders, organizations, authors, and institutions—will shape the preprint repositories growth as they are accepted, developed, and available. According to Johnson and Fosci [67], the key priority areas for immediate action for open infrastructures are below, which also resonates for preprint repositories:

- Interoperable, community-led preprints with strong open access initiatives and programmes should adopt sound governance structures with a greater representation from funders and policy makers, promoting the wider use of crucial identifiers and standards for preprints with maximum community participation, like open access repositories.
- Ensure the financial sustainability of critical services, particularly the DOAJ and SHERPA, strengthening coalitions and funders, like SCOSS for preprint services, and balancing different disciplines and their representation fairly.
- Take into the account the rapid growth of preprints and create an integrated infrastructure for them, which is based on roadmaps and strategies for mainstreaming them across other modes of scholarly communication.
- Invest strategically in preprint repositories and services in order to create a coherent OA infrastructure that is efficient, integrated, and representative of all stakeholders.

5.3. Preprints for Open Science and Public

With its ability to promote open, ethical, and transparent research workflows and processes, preprints promote building open infrastructures and symbiotic services as—the web of data where reproducibility is at its core—mutually supporting and growing along with other research artifacts [71]. As more and more preprint repositories grow, this is going to consolidate the research ecosystem towards a resilient, transparent, and open research environment for the public in promoting scientific temper and awareness as a public good.

Preprint repositories as public good initiatives offer enormous opportunities for researchers to manage the life cycle of research production, data management, access and collaboration control, project analytics, version control, and centralized access in a distributed environment [72]. They allow for researchers to disseminate preliminary work or draft papers to a wider global community of researchers, before formally submitting to peer reviewed journals to obtain feedback or comments. It also helps in speeding up the communication of research results and fostering collaborations. Currently, many journals accept preprint submissions. Nature and Science have been accepting preprints for long time, since they publish physics papers. At the American Chemical Society, 20 of the 50 journals accept preprints unconditionally [73]. Fostering scholarly commons, such as preprints, will open up opportunities for scientists and the public to solve some of our pressing problems from climate change to drug discovery, and it is possible through open science. Without limits and no embargos, preprints pose no great threats than if they remain inaccessible and restricted for the public [74].

5.4. Peer Review in Preprints: Revisiting for Present Times

The peer review process exists to enable nominally disinterested experts to assure the quality of academic publications, but preprint servers usually host articles that have not yet been subject to peer review. The question of peer review at this juncture is—for open science—will the scientific communities accept preprints without peer review when this process itself has been entangled with a lack of incentives, credits, and recognition for peer reviewers [36,75,76]. Since preprints are not necessarily peer reviewed and explicit about that, this remains to be discussed. This offers enormous potential for establishing processes like the open review mechanism and new models of peer review.

Open science through preprints promote transparency and secure provenance, time, and integrity of scientific data in an open and distributed infrastructure documenting every step of the research process and data for public. As bibliometric measures are not the indicator of achievement, there is a need to evaluate what needs to change in our culture, who are all involved, what are the best effective ways, and how it can be measured [77,78]. The challenges of maintaining unbiased review systems without gender bias in authorship and peer review, keeping the diversity of gender, racial, and ethnic communities, and the high quality of ethics and transparency calls for attention and cultural change in scholarly communication. ASAPbio's initiative is worth mentioning for accelerating scholarly communication in life sciences through preprints. There is also an equal emphasis on standards, research integrity and ethics, quality, and credibility to navigate through the peer review process with scope for new initiatives having potential issues and advantages disrupting scholarly communication both in systems and as a process with incentives in place of fostering open research environments and open access publishing [76,79].

Hence, reforming scholarly communication system to overcome barriers in legal framework, information technology infrastructures, business models, indexing services and standards, the academic reward system, marketing, and critical mass to integrate subject-specific, institutional, and data repositories into the main channels of scientific publications is critical, in which preprints development is a key component [80]. Though long established as a standardized practice with no other viable options for scientific communities, the peer review process is crucially invaluable and unquestionable, and for preprints, this process calls for openness. Moreover, it should broaden the approaches to accommodate open rewards, incentives, and other non-monetary benefits, as they advance scientific communication [75] to solve social problems, make sense for policy makers, and push forward scholarship to the advancement of humanity.

6. Conclusions

Preprint repositories are gaining momentum in becoming active partners of the scholarly research ecosystem and they contribute to open scholarship as a new model of scholarly publishing, as discussed in this article. Nevertheless, the dangers of the commercialization of preprints does not augur well for open science. This necessitates questions regarding the sustainability of preprint repositories and to what degree commercial business models interfere with open science. Without embargos, preprints pose no risks to the public understanding of science and hence imposing limits is against the public interest [74]. Preprints apparently add to the existing complexities in scholarly publishing; however, its plethora of models, scale, and form give rise to opportunities to embrace it on one hand and on the other hand may take time for mainstreaming in scholarly publishing [81–85]. Nonetheless, what constitutes them and whether they will stand out in the constructs of scholarly communication remains to be seen in the wake of diverse open data, open access-publishing models, open infrastructures, and web 3.0 technologies [86]. These factors are central for scholarly communication to enrich and strengthen scholarly web with search engines, indexing systems, semantic technologies, and social software analytics to maximize the research impact and build reputation systems through open infrastructures and metrics for authors and institutions. Going forward, on the landscape of preprints and metrics, perhaps overlay systems could be implemented, based on repositories using new metrics as overlay journals emerge. Preprint repositories have emerged as movement and they are implemented in different ways; approached in heterogeneous forms and seeing them along with conventional journals may be a possibility or whether they will change the scholarly communication landscape fundamentally, as hubs of early-research output have important caveats for open science [6,37]. However, the trade-offs, such as the questions of conflict of interests, risks, and research ethics with which preprints are published, need to be addressed for the public in the public domain and in understanding science [87,88].

References

1. PrePubMed. Monthly Statistics for October 2018. Available online: http://www.prepubmed.org/monthly_stats/ (accessed on 14 December 2018).
2. Teixeira da Silva, J.A. The Preprint Wars. *AME Med. J.* **2017**, *2*, 74. [CrossRef]
3. Piwowar, H.; Priem, J.; Larivière, V.; Alperin, J.P.; Matthias, L.; Norlander, B.; Farley, A.; West, J.; Haustein, S. The State of OA: A Large-Scale Analysis of the Prevalence and Impact of Open Access Articles. *PeerJ* **2018**, *6*, e4375. [CrossRef]
4. Peiperl, L. Preprints in Medical Research: Progress and Principles. *PLOS Med.* **2018**, *15*, e1002563. [CrossRef] [PubMed]
5. Severin, A.; Egger, M.; Eve, M.P.; Hürlimann, D. Discipline-Specific Open Access Publishing Practices and Barriers to Change: An Evidence-Based Review. *F1000Research* **2018**, *7*, 1925. [CrossRef]
6. Cobb, M. The Prehistory of Biology Preprints: A Forgotten Experiment from the 1960s. *PLoS Biol.* **2017**, *15*, e2003995. [CrossRef]
7. Eysenbach, G. The Impact of Preprint Servers and Electronic Publishing on Biomedical Research. *Curr. Opin. Immunol.* **2000**, *12*, 499–503. [CrossRef]
8. Tennant, J.; Bauin, S.; James, S.; Kant, J. The Evolving Preprint Landscape: Introductory Report for the Knowledge Exchange Working Group on Preprints. Available online: https://osf.io/preprints/bitss/796tu/ (accessed on 31 July 2018).
9. Wikipedia. Preprint. Available online: https://en.wikipedia.org/wiki/Preprint (accessed on 18 November 2018).
10. Bornmann, L. Scientific Peer Review: An Analysis of the Peer Review Process from the Perspective of Sociology of Science Theories. *Hum. Arch. J. Sociol. Self-Knowl.* **2008**, *6*, 23–38.
11. Rowland, F. The Peer-Review Process. *Learn. Publ.* **2002**, *15*, 247–258. [CrossRef]
12. Ingelfinger, F.J. Definition of Sole Contribution. *N. Engl. J. Med.* **1969**, *281*, 676–677.
13. Larivière, V.; Haustein, S.; Mongeon, P. The Oligopoly of Academic Publishers in the Digital Era. *PLoS ONE* **2015**, *10*, e0127502. [CrossRef]
14. Van Noorden, R. Open Access: The True Cost of Science Publishing. *Nature* **2013**, *495*, 426–429. [CrossRef]
15. Shen, C. Open Access Scholarly Journal Publishing in Chinese. *Publications* **2017**, *5*, 22. [CrossRef]
16. Else, H. Radical Open-Access Plan Could Spell End to Journal Subscriptions. *Nature* **2018**, *561*, 17–18. [CrossRef]
17. Smith, A. Alternative Open Access Publishing Models: Exploring New Territories in Scholarly Communication. In *Report on the Workshop Held on 12 October 2015 at the European Commission Directorate-General for Communications Networks, Content and Technology*; European Commission: Brussels, Belgium, 2015.
18. Björk, B.-C. Evolution of the Scholarly Mega-Journal, 2006–2017. *PeerJ* **2018**, *6*, e4357. [CrossRef]
19. Spezi, V.; Wakeling, S.; Pinfield, S.; Creaser, C.; Fry, J.; Willett, P. Open-Access Mega-Journals. *J. Doc.* **2017**, *73*, 263–283. [CrossRef]
20. Berners-Lee, T.; O'Hara, K. The Read-Write Linked Data Web. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2013**, *371*, 20120513. [CrossRef]
21. Research Excellence Framework. REF 2014: Key Facts. Available online: http://www.ref.ac.uk/2014/media/ref/content/pub/REFBriefGuide2014.pdf (accessed on 5 October 2018).
22. Garg, K.C.; Kumar, S. Uncitedness of Indian Scientific Output. *Curr. Sci.* **2014**, *107*, 965–970.
23. Hu, Z.; Wu, Y. A Probe into Causes of Non-Citation Based on Survey Data. *Soc. Sci. Inf.* **2018**, *57*, 139–151. [CrossRef]

24. Flatt, J.; Blasimme, A.; Vayena, E. Improving the Measurement of Scientific Success by Reporting a Self-Citation Index. *Publications* **2017**, *5*, 20. [CrossRef]

25. Martín-Martín, A.; Orduna-Malea, E.; López-Cózar, E. Scholar Mirrors: Integrating Evidence of Impact from Multiple Sources into One Platform to Expedite Researcher Evaluation. In Proceedings of the *STI 2017* Conference: Science, Technology and Innovation Indicators. "Open Indicators: Innovation, Participation and Actor-Based STI Indicators", Paris, France, 6–8 September 2017.

26. Buschman, M.; Michalek, A. Are Alternative Metrics Still Alternative? *Bull. Am. Soc. Inf. Sci. Technol.* **2013**, *39*, 35–39. [CrossRef]

27. Martín-Martín, A.; Orduna-Malea, E.; Delgado López-Cózar, E. Author-Level Metrics in the New Academic Profile Platforms: The Online Behaviour of the Bibliometrics Community. *J. Informetr.* **2018**, *12*, 494–509. [CrossRef]

28. Tennant, J.P.; Waldner, F.; Jacques, D.C.; Masuzzo, P.; Collister, L.B.; Hartgerink, C.H.J. The Academic, Economic and Societal Impacts of Open Access: An Evidence-Based Review. *F1000Research* **2016**, *5*, 632. [CrossRef]

29. Seglen, P.O. Citation Rates and Journal Impact Factors Are Not Suitable for Evaluation of Research. *Acta Orthop. Scand.* **1998**, *69*, 224–229. [CrossRef]

30. Brembs, B.; Button, K.; Munafò, M. Deep Impact: Unintended Consequences of Journal Rank. *Front. Hum. Neurosci.* **2013**, *7*, 291. [CrossRef]

31. Brembs, B. Prestigious Science Journals Struggle to Reach Even Average Reliability. *Front. Hum. Neurosci.* **2018**, *12*, 37. [CrossRef]

32. Rau, H.; Goggins, G.; Fahy, F. From Invisibility to Impact: Recognising the Scientific and Societal Relevance of Interdisciplinary Sustainability Research. *Res. Policy* **2018**, *47*, 266–276. [CrossRef]

33. Weale, A.R.; Bailey, M.; Lear, P.A. The Level of Non-Citation of Articles within a Journal as a Measure of Quality: A Comparison to the Impact Factor. *BMC Med. Res. Methodol.* **2004**, *4*, 14. [CrossRef]

34. Chaddah, P. Evaluation of Research Output. *Curr. Sci.* **2017**, *113*, 1814–1845.

35. PeerJ Prints. What Is a Preprint? Available online: https://peerj.com/about/preprints/what-is-a-preprint/ (accessed on 15 April 2018).

36. Neylon, C.; Pattinson, D.; Bilder, G.; Lin, J. On the Origin of Nonequivalent States: How We Can Talk about Preprints. *F1000Research* **2017**, *6*, 608. [CrossRef]

37. Rittman, M. Preprints as a Hub for Early-Stage Research Outputs. *Preprints* **2018**, 1–16. [CrossRef]

38. Fabry, G.; Fischer, M.R. Beyond the Impact Factor—What Do Alternative Metrics Have to Offer? *GMS J. Med. Educ.* **2017**, *34*. [CrossRef]

39. Meadows, A. Journals Peer Review: Past, Present, Future. Available online: https://scholarlykitchen.sspnet.org/2017/09/14/journals-peer-review-past-present-future/ (accessed on 20 April 2018).

40. Gölitz, P. Preprints, Impact Factors, and Unethical Behavior, but Also Lots of Good News. *Angew. Chem. Int. Ed.* **2016**, *55*, 13621–13623. [CrossRef]

41. Nicholas, D. Editorial: Thematic Series on Scholarly Communications in the Digital Age. *FEMS Microbiol. Lett.* **2018**, *365*. [CrossRef]

42. Muller, J.Z. *The Tyranny of Metrics*; Princeton University Press: Princeton, NJ, USA, 2018.

43. Gu, F.; Widén-Wulff, G. Scholarly Communication and Possible Changes in the Context of Social Media. *Electron. Libr.* **2011**, *29*, 762–776. [CrossRef]

44. Mahesh, G. The Changing Face of Scholarly Journals. *Curr. Sci.* **2017**, *113*, 1813–1814.

45. Shehata, A.; Ellis, D.; Foster, A.E. Changing Styles of Informal Academic Communication in the Age of the Web. *J. Doc.* **2017**, *73*, 825–842. [CrossRef]

46. The Conversation Global. The Conversation. Available online: https://theconversation.com/global (accessed on 18 May 2018).

47. Asia and the Pacific Policy Society. Policyforum.net. Available online: https://www.policyforum.net/ (accessed on 18 May 2018).

48. Brochu, L.; Burns, J. Librarians and Research Data Management- A Literature Review: Commentary from a Senior Professional and a New Professional Librarian. *New Rev. Acad. Librariansh.* **2018**. [CrossRef]

49. Tennant, J.; Brembs, B. RELX Referral to EU Competition Authority. *Zenodo* **2018**. [CrossRef]

50. Commons, C. Licensing Types. Available online: https://creativecommons.org/share-your-work/licensing-types-examples/ (accessed on 12 October 2018).

51. Commons, C. What Our Licenses Do. Available online: https://creativecommons.org/licenses/ (accessed on 9 October 2018).

52. Commons, C. Attribution 4.0 International (CC BY 4.0). Available online: https://creativecommons.org/licenses/by/4.0/ (accessed on 10 October 2018).

53. Martín-Martín, A.; Orduna-Malea, E.; Thelwall, M.; Delgado López-Cózar, E. Google Scholar, Web of Science, and Scopus: A Systematic Comparison of Citations in 252 Subject Categories. *J. Informetr.* **2018**, *12*, 1160–1177. [CrossRef]

54. RePEc. RePEc/IDEAS Rankings. Available online: https://ideas.repec.org/top (accessed on 8 October 2018).

55. Evans-Cowley, J.S. There's an App for That: Mobile Applications for Urban Planning. *SSRN Electron. J.* **2011**. [CrossRef]

56. Broman, K.W.; Woo, K.H. Data Organization in Spreadsheets. *Am. Stat.* **2018**, *72*, 2–10. [CrossRef]

57. Lagoze, C.; Van de Sompel, H.; Nelson, M.; Warner, S. The Open Archives Initiative Protocol for Metadata Harvesting. Available online: https://www.openarchives.org/OAI/openarchivesprotocol.html (accessed on 15 December 2018).

58. Costas, R.; Zahedi, Z.; Wouters, P. Do "Altmetrics" Correlate with Citations? Extensive Comparison of Altmetric Indicators with Citations from a Multidisciplinary Perspective. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 2003–2019. [CrossRef]

59. Allen, R.; Hartland, D. *FAIR in Practice—Jisc Report on the Findable Accessible Interoperable and Reuseable Data Principles*; JISC: Bristol, UK, 2018.

60. Bruce, R.; Cordewener, B. Open Science Is All Very Well but How Do You Make It FAIR in Practice? Available online: https://www.jisc.ac.uk/blog/open-science-is-all-very-well-but-how-do-you-make-it-fair-in-practice-12-jul-2018 (accessed on 28 July 2018).

61. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]

62. Robinson-Garcia, N.; Mongeon, P.; Jeng, W.; Costas, R. DataCite as a Novel Bibliometric Source: Coverage, Strengths and Limitations. *J. Informetr.* **2017**, *11*, 841–854. [CrossRef]

63. Staines, H. Preprint Services Gather to Explore an Annotated Future. Available online: https://web.hypothes.is/blog/preprint-services-gather-to-explore-an-annotated-future/ (accessed on 23 May 2018).

64. Shewale, N.A.; Balaji, B.P.; Shewale, M. Open Content: An Inference for Developing an Open Information Field. In *Open Source Technology: Concepts, Methodologies, Tools, and Applications*; IGI Global: Hershey, PA, USA, 2015; pp. 902–917.

65. Chan, L. What Role for Open and Collaborative Science in Development? Available online: http://www.universityworldnews.com/article.php?story=20140905132026576 (accessed on 25 June 2018).

66. JISC. *OA Sustainability Index*; JISC: Bristol, UK, 2015.

67. Johnson, R.; Fosci, M. *Putting down Roots: Securing the Future of Open-Access Policies*; JISC: Bristol, UK, 2016.

68. Ali-Khan, S.E.; Jean, A.; MacDonald, E.; Gold, E.R. Defining Success in Open Science. *MNI Open Res.* **2018**. [CrossRef]

69. National Academies of Sciences, Engineering, and Medicine. *Open Science by Design: Realizing a Vision for 21st Century Research*; National Academies Press: Washington, DC, USA, 2018.

70. Hudson-Vitale, C.R.; Johnson, R.P.; Ruttenberg, J.; Spies, J.R. SHARE: Community-Focused Infrastructure and a Public Goods, Scholarly Database to Advance Access to Research. *D-Lib Mag.* **2017**, *23*. [CrossRef]

71. Capadisli, S.; Guy, A.; Lange, C.; Auer, S.; Greco, N. Linked Research: An Approach for Scholarly Communication. Available online: http://csarven.ca/linked-research-scholarly-communication (accessed on 15 May 2018).

72. Foster, E.D.; Deardorff, A. Open Science Framework (OSF). *J. Med. Libr. Assoc.* **2017**, *105*, 203.

73. American Chemical Society. ACS Launches Chemistry Preprint Server. Available online: https://cen.acs.org/articles/94/web/2016/08/ACS-launches-chemistry-preprint-server.html (accessed on 28 April 2018).

74. Sarabipour, S.; Wissink, E.M.; Burgess, S.J.; Hensel, Z.; Debat, H.; Emmott, E.A.; Akay, A.; Akdemir, K.; Schwessinger, B. Maintaining Confidence in the Reporting of Scientific Outputs. *PeerJ Prepr.* **2018**. [CrossRef]

75. Tennant, J.P. The State of the Art in Peer Review. *FEMS Microbiol. Lett.* **2018**, *365*. [CrossRef]

76. Tennant, J.P.; Dugan, J.M.; Graziotin, D.; Jacques, D.C.; Waldner, F.; Mietchen, D.; Elkhatib, Y.; Collister, B.L.; Pikas, C.K.; Crick, T.; et al. A Multi-Disciplinary Perspective on Emergent and Future Innovations in Peer Review. *F1000Research* **2017**, *6*, 1151. [CrossRef]

77. Ma, L.; Ladisch, M. Scholarly Communication and Practices in the World of Metrics: An Exploratory Study. In Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives through Information & Technology, Copenhagen, Denmark, 14–18 October 2016; Volume 53, p. 132.

78. Meadows, A. Changing the Culture in Scholarly Communications. Available online: https://scholarlykitchen.sspnet.org/2017/08/07/changing-culture-scholarly-communications/ (accessed on 15 April 2018).

79. Allahar, H. Is Open Access Publishing a Case of Disruptive Innovation? *Int. J. Bus. Environ.* **2018**, *10*, 35–51. [CrossRef]

80. Björk, B.-C. Open Access to Scientific Publications—An Analysis of the Barriers to Change? *Inf. Res.* **2004**, *9*, 170.

81. Calne, R. Preprint Servers: Vet Reproducibility of Biology Preprints. *Nature* **2016**, *535*, 493. [CrossRef]

82. Da Silva, J.A.T. Preprints Should Not Be Cited. *Curr. Sci.* **2017**, *113*, 1026–1027.

83. Inlexio. The Rising Tide of Preprint Servers. Available online: https://www.inlexio.com/rising-tide-preprint-servers/ (accessed on 30 May 2018).

84. Hoyt, J.; Binfield, P. Who Killed the PrePrint, and Could It Make a Return? Available online: https://blogs.scientificamerican.com/guest-blog/who-killed-the-preprint-and-could-it-make-a-return/ (accessed on 8 May 2018).

85. Luther, J. The Stars Are Aligning for Preprints. Available online: https://scholarlykitchen.sspnet.org/2017/04/18/stars-aligning-preprints/ (accessed on 29 April 2018).

86. Balaji, B.P.; Vinay, M.S.; Shalini, B.G.; Raju, M.J.S. An Integrative Review of Web 3.0 in Academic Libraries. *Libr. Hi Tech. News* **2018**, *35*, 13–17. [CrossRef]

87. Da Silva, J.A.T. Intellectual Phishing, Hidden Conflicts of Interest and Hidden Data: New Risks of Preprints. *J. Advocacy Res. Educ.* **2017**, *4*, 136–146.

88. Sheldon, T. Preprints Could Promote Confusion and Distortion. *Nature* **2018**, *559*, 445. [CrossRef]