Preprints and Scholarly Communication: Adoption, Practices, Drivers and Barriers [version 1; peer review: 2 approved with reservations]

Andrea Chiarelli¹, Rob Johnson¹, Stephen Pinfield², Emma Richens¹

¹Research Consulting Limited, Nottingham, NG7 2TU, UK
²Information School, University of Sheffield, Sheffield, S1 4DP, UK

Abstract

Background: Since 2013, there has been a dramatic increase in the number of preprint servers available online. To date, little is known about the position of researchers, funders, research performing organisations and other stakeholders with respect to this fast-paced landscape. In this article, we explore the benefits and challenges of preprint posting, along with issues such as infrastructure and financial sustainability. We also discuss the definition of a ‘preprint’ in different communities, and the impact this has on further uptake.

Methods: This study is based on 38 detailed semi-structured interviews of key stakeholders based on a purposive heterogeneous sampling approach. Interviews were undertaken between October 2018 and January 2019. These were recorded, transcribed and subjected to thematic analysis to identify trends. Interview questions were designed based on Innovation Diffusion Theory, which is also used to interpret the results of this study.

Results: Our study is the first using empirical data to understand the new wave of preprint servers and found that early and fast dissemination is the most appealing feature of the practice. The main concerns are related to the lack of quality assurance and the ‘Ingelfinger rule’. We identified trust as an essential enabler of preprint posting and stress the enabling role of Twitter in showcasing preprints and enabling comments on these.

Conclusions: The preprints landscape is evolving fast and disciplinary communities are at different stages in the innovation diffusion process. The landscape is characterised by significant experimentation, which leads to the conclusion that a one-size-fits-all approach to preprints is not feasible. Cooperation and active engagement between the stakeholders involved will play an important role in the future. In our paper, we share questions for the further development of the preprints landscape, with the most important being whether preprint posting will develop as a publisher- or researcher-centric practice.

Keywords
preprints, scholarly communication, peer-review, innovation diffusion theory
Corresponding author: Andrea Chiarelli (andrea.chiarelli@research-consulting.com)

Author roles: Chiarelli A: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Visualization, Writing – Review & Editing; Johnson R: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing; Pinfield S: Conceptualization, Investigation, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; Richens E: Data Curation, Formal Analysis, Investigation, Visualization, Writing – Review & Editing

Competing interests: Stephen Pinfield is a member of the F1000 Advisory Board and was the Founding Director of SHERPA between 2002 and 2012 (he has been involved since then in providing external advice to the service). A member of the F1000Research staff was interviewed in the course of this study but had no role in the writing or production of the article.

Grant information: This study has been funded by Knowledge Exchange (KE), a group of national organisations from six European countries supporting research infrastructure and services to enable the use of digital technologies to improve higher education and research: CSC in Finland, CNRS in France, DEFF in Denmark, DFG in Germany, Jisc in the UK and SURF in the Netherlands.

Copyright: © 2019 Chiarelli A et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Chiarelli A, Johnson R, Pinfield S and Richens E. Preprints and Scholarly Communication: Adoption, Practices, Drivers and Barriers [version 1; peer review: 2 approved with reservations] F1000Research 2019, 8:971 (https://doi.org/10.12688/f1000research.19619.1)

First published: 26 Jun 2019, 8:971 (https://doi.org/10.12688/f1000research.19619.1)
Introduction

Background

The period since 2013 has seen a marked rise in the number of preprint servers set up for different communities in order to facilitate the rapid dissemination of pre-refereed research outputs. Tennant et al. (2018) list 18 servers launched between 2013 and 2018, variously set up by disciplinary communities, countries, research funders and publishers. One of the first in this new wave was the discipline-based server, bioRxiv – set up by the Cold Spring Harbor Laboratory in 2013 to cover the life sciences – which has been a focus of discussion and debate (Abdill & Blekhman, 2019; Luther, 2017; Vale, 2015). However, there are a considerable number of other disciplinary servers, including several set up by the Center for Open Science, such as SocArXiv, engrXiv and PsyArXiv (all of which were launched in 2016), as well as platforms such as ESSOAr, set up by the American Geophysical Union in 2018. At the same time, national servers have been launched, including ChinaXiv (for China), IndiaRxiv (for India) and INA-Rxiv (Indonesia) (Mallapaty, 2019). Funders of research have also set up platforms that allow the sharing of articles before peer-review, including, in 2016, Wellcome Open Research, for Wellcome-funded researchers. Finally, a number of journal publishers have added the dissemination of preprints to their workflows. The open access (OA) publishers PeerJ and MDPI both have their own preprints services; and PLOS now deposits submissions to its journal PLOS ONE in bioRxiv on behalf of authors. The F1000Research publishing platform has promoted a novel publication model involving preprints, in which immediate release of author submissions as preprints is followed by open peer review, with revised versions of a paper (alongside author responses to reviewer comments) published in the journal as they are made.

Of course, preprint servers as a venue of scholarly communication are not new. arXiv, the preprint server for physics, mathematics, computer science and related subjects, and often regarded as an exemplar preprint server, was set up as early as 1991 (Larivière et al., 2014) – it is considered by some as the origin of the open access movement (Gajdacs, 2013). RePEc, housing ‘working papers’ in economics, was launched in 1997. There have also been unsuccessful attempts in the past to develop preprints services, including for biology, in the late 1990s (Ginsparg, 2016), and chemistry, in the early 2000s (Warr, 2003).

The move to set up servers since 2013 signals a new level of interest in preprints and a number of recent studies (e.g. Abdill & Blekhman, 2019; Balaji & Dhanamjaya, 2019; Luther, 2017; Tennant et al., 2018) have provided useful overviews of what Tennant et al. (2018, p. 5) call the “explosion of preprint platforms and services”. Significantly, this new interest has often come from disciplinary communities not previously associated with adoption of preprints. Biomedical disciplines served by bioRxiv, for example, have traditionally been associated with ‘Gold’ open access (publication in journals) rather than ‘Green’ OA (deposit of copies of papers in archives or repositories), and have typically not favoured dissemination of papers in pre-refereed form (Martín-Martín et al., 2018; Wang et al., 2018).

The launch of these new preprint servers has led to discussion and debate, and some have suggested that preprints may become a disruptive force in scholarly communication (Luther, 2017; Velterop, 2018). Green (2019) has argued for a digital transformation of publishing into a two-step process: articles would first be posted as preprints, and then invited to formal peer review only if they receive sufficient attention. He argues that this would not only represent a cost-effective model for OA and drive out predatory journals, but could also resolve the so-called ‘serials crisis’, under which growth in research budgets (which produce articles) consistently outstrips that of library budgets (which are used to purchase articles).

The case has been made for preprints in a number of disciplines, including biology (Desjardins-Proulx et al., 2013; Vale, 2015), medicine (Lauer et al., 2015) and chemistry (Carà et al., 2017). Some funders have signalled support for preprints being used in grant applications, including National Institutes of Health (NIH) and Zuckerberg Foundation in the USA, and the Wellcome Trust in the UK. However, sceptics have questioned the value of preprints and even suggested they may be dangerous – circulating versions of articles before they have been quality controlled by peer review may lead to unnecessary risk, particularly in disciplines like medicine (Sheldon, 2018).

This paper aims to explore the current and potential future role of preprints as a vehicle for scholarly communication by investigating current practices, drivers and barriers to their use. The overall objective of the study was to explore the place of preprints in the research lifecycle from the points of view of key actors, including:

- research funders;
- research performing organisations;
- preprint servers and service providers; and
- researchers (engaged and unengaged).

The topics in focus included usage of preprints, perceived benefits and challenges, policy positions, values and strategies. The research took the form of a set of 38 detailed interviews with representatives from these groups.

The study was funded by, and co-produced with, Knowledge Exchange (a group of national organisations from six European countries supporting research infrastructure), as part of their work on open-access policy and service development. It was, therefore, important that the research should not merely have a descriptive purpose but also a prescriptive one, involving setting out possible directions for future action. The study is the first using empirical data focusing on the new wave of preprint servers set up since 2013, as such it aims to make a significant contribution to knowledge in this dynamic area.

Literature review

Apart from recent discussion on the growth of preprint services (Abdill & Blekhman, 2019; Balaji & Dhanamjaya, 2019; Tennant et al., 2018), consideration of preprints in the formal academic literature, as well as in the scientific press and other online venues (such as blogs and other social media commentary), has tended to concentrate on four main issues: firstly, defining...
preprints; secondly, their benefits and dis-benefits; thirdly, disciplinary differences; and fourthly, policy developments. We discuss these in turn in what follows. There are, however, still a relatively small number of peer-reviewed studies focusing on preprints – much of the literature is still to be found in editorials and opinion pieces rather than data-driven research.

**Defining preprints.** Different definitions of preprints in the academic literature typically relate to a number of key components: (1) genre, (2) timing, (3) versioning, (4) accessibility, (5) responsibility and (6) value (see Table 1).

With regard to (1) genre, Berg et al. (2016, p. 899) state, “a preprint is a complete scientific manuscript”, and Bourne et al. (2017) observe, “typically, a preprint is a research article, editorial, review, etc.”. Whilst the latter widen the scope also to include, “a commentary, a report of negative results, a large data set and its description, and more” (p. 1), most of the discourse on preprints tends to assume they are conventional research papers and therefore follow the academic conventions of that ‘genre’.

With regard to (2) timing, the key point made by most commentators is that a preprint is made available before formal publication, which Carà et al. (2017) describe as “prepublication”.

For (3) versioning, the relationship of a preprint to peer review is central. Desjardins-Proulx et al.’s (2013, p. 1) observation is typical in stating that preprints are made available, “before, or in parallel to, submitting them to journals for traditional peer review”. Suber (2012, p. 102) points out that this is not to “bypass peer review”, but that it applies to “works destined for peer review but not yet peer reviewed”. However, Bourne et al. (2017) controversially extend the definition to include “a paper that has been peer reviewed and…was rejected, but the authors are willing to make the content public”.

Accessibility (4) is crucial, with a preprint normally defined as being (or assumed to be) openly available: it “can be viewed without charge on the Web” (Berg et al., 2016, p. 899). Therefore, the venue for distribution of preprints is often assumed to be a freely-accessible server of some kind, a point highlighted by Berg et al. (2016, p. 899), who include in their definition that a preprint “is uploaded by the authors to a public server”.

The above phrase, “by the authors” here is important and relates to (5) responsibility. Responsibility for distribution of preprints is traditionally assumed to be that of the author, a component of the definition that is often implicit in the verbs used to describe dissemination of preprints, such as, “sharing”, “posting” and “self-archiving”.

<table>
<thead>
<tr>
<th>Table 1. Components of the definition of a preprint.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
</tr>
<tr>
<td>Exemplifying quote</td>
</tr>
<tr>
<td>Other literature sources</td>
</tr>
</tbody>
</table>
The final component of (6) value is summarised by Bourne et al. (2017): “a preprint is a research output that has not completed a typical publication pipeline but is of value to the community and deserving of being easily discovered and accessed”. To say that an output is “deserving” of dissemination is, of course, a value judgement and difficult to demonstrate for each deposit as it is made, but it is one that is implicit in much of the discourse on preprints.

With ambiguities associated with each of these six definitional components, Neylon et al. (2017) are right that “no universal definition of preprints exists”. The ‘pre’ of ‘preprint’ has sometimes been defined in relation to formal publication, with a preprint characterised as “prepublication”, leading to the controversial question of whether a preprint can itself be considered a ‘publication’ in its own right (Larivière et al., 2014). More commonly, however, the ‘pre’ of ‘preprint’ specifically refers to peer-review and is contrasted with ‘postprint’, a version produced after peer review. The conflation of peer review and publication in some discussions is a reflection of their close association in scholarly communication. It is interesting that the use of the terminology of ‘postprint’ in contradistinction to ‘preprint’, and with both termed generically as ‘eprints’, has declined in recent years. However, Tennant et al. (2018) have proposed its revival for reasons of clarity. Of course, the ‘print’ part of ‘preprint’ is largely anachronistic, but like terms such as ‘paper’ and ‘manuscript’, has continued to be used even in a digital environment.

Benefits and dis-benefits of preprints. Of the advantages of preprints discussed in the literature, perhaps the most prominent are the early and rapid dissemination of research results (Khera, 2018). Using preprints has the potential to “accelerate” science, something that is particularly useful, for example, in combatting outbreaks of diseases (Johansson et al., 2018). The formal scientific publication process is often seen as frustratingly slow, particularly in a context where final versions of articles may be little different from preprints (Klein et al., 2016). Preprints allow authors to assert priority early – a preprint is date-stamped in a way widely recognised by many communities (Ginsparg, 2016; Mallapaty, 2019; Tennant et al., 2019). Preventing researchers being ‘scooped’ is a major priority in many fast-moving disciplines, but applies, at least to some extent, in all areas of academic research, where novelty is prized. Early dissemination can be useful to some particular members of the scholarly community, with early career researchers (ECRs) commonly being identified as particular beneficiaries, as preprints can allow them to rapidly achieve “visibility” (Desjardins-Proulx et al., 2013; Sarabipour et al., 2018; Tennant et al., 2019).

As well as being a fixed point in the scholarly discourse (date-stamped, etc.), another benefit of preprints emphasised in the literature is in the fact they are still subject to change. Authors can benefit from what Pinfield (2004) has called “informal peer review” of versions of their papers. Ginsparg calls this “crowd-sourced peer review”, in contrast to “journal-mediated peer review”, and states, “authors benefit greatly from feedback from interested readers, contributing to improved versions of articles, which are then uploaded” (Ginsparg, 2016, p. 5). There are, however, few empirical studies of such feedback and its value. There also appears to be little acknowledgement that the claim stands in tension to the one cited earlier that preprints often differ little from final published versions.

Other key advantages of preprints include wider and fairer distribution of research results, both within and beyond the academic community, something fundamental to many arguments for openness in general (Ginsparg, 2016; Sarabipour et al., 2018). Access to preprints for machine-based crawling in order to facilitate text mining is also seen as an advantage by some commentators (Chodacki et al., 2017). Preprints, partly as a result of wider dissemination, can also increase numbers of citations of papers (Davis & Fromerth, 2007) and create opportunities for collaborations (Kleinnert et al., 2018). Finally, preprint servers can sometimes usefully also house research outputs that might otherwise be ‘homeless’, including items that do not end up being published in peer-reviewed journals (Bourne et al., 2017).

Perhaps the most prominent criticism of preprints relates to this last issue: the lack of quality assurance through peer review (Sheldon, 2018). As well as a general concern about lowering quality standards, lack of quality control has been seen as potentially dangerous as “reports that have not undergone formal peer review [organised by a journal] could be misleading” (Lauer et al., 2015). Furthermore, uncertified science might be reported prematurely in the media and might even give rise to ‘fake news’ (Sheldon, 2018). Some insist that, at the very least, the opportunity to disseminate knowledge rapidly without peer review may encourage academics to produce low-quality outputs on fashionable topics (Teixeira da Silva, 2017).

Despite claims of the value of informal peer review enabled by preprints, some have pointed to the limited use of commenting and feedback features on preprint servers (Sarabipour et al., 2018). Others have gone further and questioned the value of self-appointed reviewers, as opposed to those selected by journal editors (see the issue of “self-policing” highlighted by Harnad, 1998). Preprint posting, however, is unlikely in any case to substitute for the valuable role played by selective journals in filtering content.

A number of authors report the concern that dissemination of a preprint may be considered ‘prior publication’, thereby jeopardising acceptance of the paper by a journal – the so-called ‘Ingelfinger rule’ (Nallamoithu & Hill, 2017). Whilst this convention has come under criticism and been withdrawn by some publishers, it still exists for some journals, e.g. in medicine and chemistry (Lauer et al., 2015; Teixeira da Silva & Dobrászki, 2019).

Disciplinary differences. Disciplinary differences have been a prominent feature of the literature on preprints. Neylon et al. (2017) in their seminal work conceptualising preprints usefully distinguish between the “state” and “standing” of preprints.
“State” relates to, “external, objectively determinable, characteristics” of preprints; “standing” refers to the “position, status, or reputation” of preprints. Neylon et al. (2017) discuss in detail how preprints of similar states can have very different standings in different disciplinary communities, using the example of the contrast between physics and life sciences. For example, it is not universally agreed when an output should be citable (in the literature, funding proposals or promotion cases) or when it can be used to establish a claim of precedence.

In disciplines where a preprint is not considered appropriate to establish precedence, it has also been suggested that making a preprint available may actually encourage research to be scooped by rival researchers who publish in a recognised journal before the preprint authors (the ‘flip side’ of the priority claim argument above) (Kaiser, 2017).

It is commonly observed that physics has a well-established preprint tradition unlike many other Science, Technology and Medicine (STM) disciplines. Lauer et al. (2015, p. 2448) note, “biology…has trailed behind, whereas clinical research remains well behind.” Caru et al. (2017, p. 7924) characterise chemistry as being “late in embracing preprints”. Such language (“behind”, “late”) seems to represent an assumption that all disciplines will eventually come to use preprints, but different disciplines are now simply at different points in the adoption process. Such a view has been disputed, however, with some arguing disciplinary differences in communication practices are likely to exist in the long term and therefore that preprints will not be adopted universally across disciplines (Kling & McKim, 2000).

Policy developments. Of course, disciplinary practices do not operate in a vacuum. They are influenced, amongst other things, by the policy environment in which researchers work. The ‘policy stack’ consists of several interlocking layers: publisher policies, funder policies and institutional policies. The first of these is critical, with the Ingelfinger rule and deposit embargo at its core. To this position of publishers rejecting preprints may now be added the contrasting recent development of some publishers embracing preprints, even setting up their own preprint services (Callaway, 2013). This development is not completely unprecedented, however, since it does build to some extent on well-established processes in areas like high-energy physics of integrating preprints into the journal submission process (as an example, some physics journals allow submission by simply pointing to an arXiv preprint).

Perhaps the most noticeable shift recently in terms of policy is that of funder policies. Some funders have now explicitly signalled support for use of preprints, including allowing citation of preprints in funding bids, and support their inclusion in cases for academic advancement (Berg et al., 2016; Bourne et al., 2017). Very few funders, however, currently mandate use of preprints, although this has been proposed by preprints advocates (Sever et al., 2019). Institutional policy in this area shows some limited movement, with examples of institutions rethinking (usually rather cautiously) their approaches to criteria for career advancement in relation to the shifting scholarly communication environment, with some explicitly allowing submission of preprints (ASAPbio, n.d.). It appears that, at present, many organisations still rely on metrics such as the journal impact factor when it comes to review, promotion and tenure of their staff (McKiernan et al., 2019).

Some have argued that initiatives such as Declaration on Research Assessment (DORA), with its emphasis on the quality of the output rather than venue of publication, promote use of preprints (Polka, 2018). Another interesting area of institutional policy is the positioning of the institutional repository (IR) in relation to preprints. IR policies and practices differ in this area, with many to date having focused on versions of outputs following peer review, although this is not a universal position (Baughman et al., 2018).

Methods
Analysis framework. With preprints and preprint servers still innovative developments for most disciplines, using a qualitative research approach is an appropriate way to gain an in-depth understanding of multiple perspectives, including motivations, challenges and opportunities, and taking account of disagreements and uncertainties. We therefore chose to carry out detailed interviews of key actors in this space who could explain their perceptions, attitudes and practices in relation to preprints. Bearing in mind that perspectives of stakeholders from disciplines where preprints are established are well represented in the literature (e.g. physics, computer science, and economics), we focused our research on disciplines where preprint services are relatively new but growing rapidly in order to gauge the way in which preprint servers are now making an impact. These were biology, chemistry and psychology, corresponding to preprint servers, bioRxiv, ChemRxiv and PsyArXiv.

As a way of framing our research design, we used innovation diffusion theory (IDT), a well-established theoretical framework for explaining the way innovations are adopted in different contexts (Rogers, 2003). IDT has been tested and deployed widely and proved to be a robust explanatory model in a wide range of contexts, including OA (Hampton, 2014; Jones et al., 2006; Pinfield & Middleton, 2016; Xia, 2012). It is designed to describe “the process by which an innovation is communicated through certain channels over time among members of a social system” (Rogers, 2003) (original emphasis). An innovation is defined as “an idea, practice, or object that is perceived as new by an individual or other unit of adoption” (Rogers, 2003). Preprints are both cultural innovations, as they aim to change practices in scholarly communication, and technological innovations, in terms of changes to infrastructures and processes. IDT offers ways in which these aspects of preprints as innovation can be understood, particularly in relation to two key issues: the “innovation decision process”, and the “rate of adoption”.

The innovation adoption decision process is seen as going through a number of consecutive steps:

1. Knowledge, when the decision-making unit is exposed to the innovation’s existence and gains an understanding of how it functions;
2. **Persuasion**, when the decision-making unit forms a favourable or unfavourable attitude towards the innovation;

3. **Decision**, when the decision-making unit engages in activities that lead to a choice to adopt or reject the innovation;

4. **Implementation**, when the decision-making unit puts an innovation into use; and

5. **Confirmation**, when the decision-making unit seeks reinforcement for an innovation-decision already made but may reverse the decision if exposed to conflicting messages about it. (Rogers, 2003)

A particularly important concept to understand the success of innovations is their rate of adoption (see Table 2). This was used as the initial basis of the design of the interview questions. From the key factors associated with the rate of adoption, we selected a range of features that appear appropriate for the scope of the present investigation and that were suitable to discuss via interviews for the different stakeholder groups:

- **Perceived attributes**, are what stakeholders feel are the benefits arising from an innovation, in this case the introduction of preprints. Perceived attributes can be split into relative advantage, compatibility, complexity, trialability and observability.

- **Nature of the social system**, including “norms”, which are the established behaviour patterns for the members of a social system. They define a range of accepted behaviours and serve as a guide or standard for the behaviour of members of a social system. Norms tell individuals what behaviour they are expected to adopt and are affected by the introduction of an innovation. These relate to the level of interconnectedness or cohesiveness of the community.

- **Change agents’ promotion efforts**, which are the efforts made by individuals with influence in the system to promote the adoption of an innovation deemed desirable by a change agency (e.g. funders and institutions, service providers, publishers). Change agents often use opinion leaders in a social system as their lieutenants in diffusion activities.

- **Type of innovation decision**, which describes how the uptake of preprints is affected when individuals or communities support them, or authorities mandate their posting.

The topic of communication channels (additionally part of IDT theory on the rate of adoption) also arose organically from the discussions with our interviewees.

**Interview sampling and approach.** Interview questions were developed based on the factors outlined in Table 2 (see Extended data (Chiarelli et al., 2019a) for more information). From an initial long list of possible questions, areas for investigation were prioritised based on the different stakeholder groups involved, and our review of the literature. We also incorporated questions associated to current policy-related issues, agreed in consultation with the Knowledge Exchange steering group, taking into account the innovation adoption process. This ensured that the approach taken was both theoretically robust and sufficiently grounded in practice to be useful in generating actionable insights.

<table>
<thead>
<tr>
<th>Table 2. Variables affecting the rate of adoption of an innovation (Adapted from Rogers, 2003).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variables determining the rate of adoption</strong></td>
</tr>
<tr>
<td>Perceived attributes</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Nature of the social system</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Extent of change agents’ promotion efforts</td>
</tr>
<tr>
<td>Communication channels</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Type of innovation-decision</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
The study adopted a heterogeneous purposive sampling approach, aiming to include a wide range of perspectives from actors in the area. Participants comprised senior representatives from research funders, research-conducting organisations (universities and research institutes), preprint services, other related service providers (such as infrastructure providers), as well as researchers, both researchers demonstrably engaged with preprints (they had themselves posted a preprint) and non-engaged (there was no evidence of them having posted a preprint). Participants were based in eight countries: Denmark, Finland, France, Germany, Netherlands, Switzerland, UK and USA, all apart from the USA and Switzerland being KE member countries. Participants were identified from the literature and from their associations with relevant services or organisations. Snowball sampling was also used as the research progressed and appeals for participation on social media were also shared (particularly to identify non-engaged researchers). Participants gave their informed consent and agreed to be named as participants in any reporting on the understanding that particular views or quotations reported would not be linked to them or their organisation and that the full text of transcripts would be kept confidential. The research approach adopted by the project was given ethical approval by the University of Sheffield. A full list of participants is available in Chiarelli et al. (2019b).

We undertook 38 semi-structured interviews, with participants distributed across the targeted stakeholder groups as illustrated in Table 3. Interviews, which took place between October 2018 and January 2019, and ranged from 32–75 minutes in length, were conducted via GoToMeeting, recorded and fully transcribed using the intelligent verbatim method (including minor edits e.g. removing ‘fillers’ etc). Two interviews took the form of an email Q&A because of restrictions around the participants’ availability. The transcripts were then subjected to thematic analysis (Braun & Clarke, 2006) which took place in several stages. Initially, the research team independently read a sample of transcripts from different stakeholder interviews, including some in common, and then discussed key topics arising from the transcripts. This formed the basis of the initial coding approach then undertaken by E.R. This was reviewed as analysis proceeded, with coding being checked and validated by A.C. and S.P. as it progressed and amended as necessary in light of their comments. The codes were then grouped into themes, which form the basis of the findings reported below.

Like many kinds of qualitative research, this study was designed to be exploratory, in this case to map out key aspects of the policy environment. Its tentative conclusions will need testing. Interviewees selected for this study were not necessarily representative of the overall community and outlying results may be over-represented. Wider consultation, using other methods, will be needed in future. Furthermore, some stakeholder groups, such as publishers (who only have very limited representation in this study), and other groups (such as non-academic users of the research literature) could usefully be included in future studies.

Results
Overview
The analysis of the data identified nine major themes arising from the interviews, which can be grouped into four thematic areas (Table 4). These themes are used as the framework for presenting results and explored in more detail in what follows.

Definitions and roles of preprints
Definitions of preprints. In view of the ambiguities and disagreements in the literature around definitions of ‘preprints’, one key aim of the interviews was to ask our participants about their understanding of the term. Whilst all of our participants agreed with the broad definition that a preprint is a research output made available in a form before it has been peer-reviewed and published, there was considerable variation in the specifics of what that means. Some participants were themselves aware that the term was being used in different ways by different people and there were discrepancies (and in some cases confusion) about its precise meaning. One expressed dislike for the term, saying that it “presupposes…that you are in a print era”. However, other alternatives used by participants in this space, such as “manuscript” and “paper”, as has been observed, are equally anachronistic. There was evidence of participants struggling to find a clear language for the innovation being discussed.

<table>
<thead>
<tr>
<th>Stakeholder group</th>
<th>UK</th>
<th>Germany</th>
<th>France</th>
<th>Netherlands</th>
<th>Denmark</th>
<th>USA</th>
<th>Finland</th>
<th>Switzerland</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research funders</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>Research performing organisations</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Preprint server providers</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Other service providers</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>Engaged researchers</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Unengaged researchers</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Totals</td>
<td>11</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>38</td>
</tr>
</tbody>
</table>
Many saw a preprint as being in a form that was ready to be submitted to a journal, “at the point of submission” (Research performing organisation), as one participant put it. Referring to their own experience, one participant stated:

“the preprint...was only submitted or uploaded at a stage where it was essentially submission-ready for a journal.” (Engaged researcher)

Other participants saw preprints as earlier versions of outputs made available in order to receive comments (e.g. working papers in economics). One participant acknowledged that a preprint was commonly thought of as the submitted version of an article but also discussed possible earlier versions being shared:

“...in terms of quality, almost like [the] thing that will appear in the journal later on but if you consider a preprint like a working paper, you...definitely can see it as a step earlier in the whole research process in which there is still the possibility to enrich and improve the later formal publication on the same time.” (Research funder)

Another participant referred to the benefit of this: “it gets feedback from the community” before “official peer review” (Unengaged researcher). A key point made by one participant, but implicit in the comments of many, was that that a preprint is a version of an output that has not been peer reviewed but that the author is “committing to get it peer reviewed” (Other service provider).

“...the term itself includes the idea that you’re building it just towards something. That it’s only the preprint and then something will come later after from it.” (Research funder)

Another author said of their own approach to posting preprints: “I always had the intent to publish this in a...proper journal subsequently” (Engaged researcher). A preprint is part of a “continuum” of different research stages, one participant argued, and authors should deposit all versions of their papers (and data) in a repository as part of contributing to that continuum. However, one participant did question this emphasis on a work flow with the preprint being provisional, since it appeared to devalue preprints:

“...preprint means there’s something you know, there’s...a paradise afterwards, there’s a better life, and that it’s not a publication of its own.” (Other service provider)

Either as a provisional version of a forthcoming publication or as a “publication” in its own right, a preprint was seen as part of a recognisable genre of scientific output, and which was planned to be part of the formal published literature, but was made available earlier than formal publication.

“they’re not that radical, the concept is radical, but when you look at them, they look like articles.” (Other service provider)

Several participants included post-refereed versions of articles within the definition (although some acknowledged these were not “pure” preprints), and others recognised that in reality many authors deposited post-refereed versions on preprint servers, usually to enhance their accessibility. A small number of interviewees acknowledged that some papers posted as preprints might not end up being formally published, although this might raise questions about the value of the output:

“If there’s nothing to follow the preprint then I would start to wonder what did happen. Why was the work dropped and left on this preprint level?” (Research funder)

Some questioned whether a paper which did not end up being published could legitimately be considered a preprint, with one interviewee asserting, “it’s not a preprint; it’s just a manuscript” (Other service provider). That a preprint is basically a research paper was the assumption of the majority of participants, but one questioned this:

Table 4. Themes from the interviews.

<table>
<thead>
<tr>
<th>Thematic area</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitions and roles</td>
<td>Definitions of preprints</td>
</tr>
<tr>
<td></td>
<td>Disciplines, cultures and practices</td>
</tr>
<tr>
<td></td>
<td>Preprints’ position in the landscape</td>
</tr>
<tr>
<td>Potential benefits and challenges</td>
<td>Preprints as an asset (benefits)</td>
</tr>
<tr>
<td></td>
<td>Preprints as a liability (challenges)</td>
</tr>
<tr>
<td>Infrastructure and sustainability</td>
<td>Infrastructure</td>
</tr>
<tr>
<td></td>
<td>Policy</td>
</tr>
<tr>
<td></td>
<td>Financial stability and business models</td>
</tr>
<tr>
<td>Future</td>
<td>Future of preprints</td>
</tr>
</tbody>
</table>

Page 9 of 35
“You know, it’s basically anything, you know between a tweet... or a poster presentation and an actual paper...” (Engaged researcher)

Many acknowledged there was uncertainty about preprints and some researchers in particular were themselves cautious in committing to a definition. In some cases, this related to disciplinary differences – something acknowledged by several interviewees. One interviewee discussed definitional differences between his own discipline, chemistry, and that of physics, mainly in terms of community acceptance. Another participant from the humanities stated:

“I don’t think there’s any one perspective that scholars in the humanities have on preprints and I think that there’s some confusion about the terminology.” (Other service provider)

**Disciplines, cultures and practices.** Disciplinary differences were evident not only in perceptions of what preprints are but also in terms of their acceptance. This was important throughout the interviews. There were firstly differing levels of awareness and, following this, adoption. Some interviewees recognized physics, mathematics and computer science to have well-established preprints practices with very high levels of awareness and adoption, but in other disciplines awareness was often still low:

“I think in chemistry it’s small but growing and biology is being helped a lot by bioRxiv ...there’s certainly some areas where there is still a kind of much less awareness. I would say really outside the math and physics [communities] the awareness is much lower of what preprints are about.” (Preprint server provider)

One service provider summarized confusions (about definitions and processes) even applying to people trying to submit their work as preprints:

“So there’s definitely a growing awareness but it’s still a minority. And we still find that there are some who are confused by the process and when they submit a preprint they don’t really understand. Despite everything we try and make them aware of, they don’t really understand the process.” (Preprint server provider)

Furthermore, where there was evidence that awareness was rising, this did not necessarily result in uptake. Some unengaged researchers interviewed were aware what preprints were but had not been motivated to use them to date. Some researchers clearly saw practical barriers:

“I am not entirely sure of the process, that’s the reason why I haven’t done it. I’m sure I could work it out but I’m not entirely sure which preprint server I would use, whether one would be better for my type of work than another...” (Unengaged researcher)

However, there were signs this was beginning to change. One chemist described the situation in their discipline as moving from a position where there had previously been no use of preprints to where use was beginning to happen:

“Almost all that is changing now and also from the chemistry part, which might be the related field, there was as far as I know ChemRxiv, which is like the main repository for chemistry data. It has been going on for one or two years maybe.” (Unengaged researcher)

One participant described the process by which awareness of this new development diffuses through the community via informal channels and how they had become more personally aware of it during the course of doing a PhD:

“I didn’t get any information on anything like, you know, I found out about it myself, you know, something as simple as that. Most of it’s just through word of mouth, like as you go through a PhD, as you’re talking to people, a lot of meetings, as you hear these terms come up more and more... It’s just accidental.” (Unengaged researcher)

One participant representing a preprint server saw growing awareness and willingness to experiment in different subject communities:

“And the momentum behind [name of the preprint server] – and I would not call it a success yet, I would call it momentum – but I think that momentum has given encouragement to other groups of scholars to investigate the possibility of developing a preprint platform for their own discipline. Whether that’s in earth sciences or anthropology or psychology, sociology and so on.” (Preprint server provider)

This was confirmed by another service provider, who stated that before 2016, preprints “were not used much” in their discipline and many researchers “were unaware of what a preprint was”. But this was changing: “the popularity of preprints within the field...is rising” and this was partly attributable to the preprint server the provider represented which had been a “driving force” for change (Preprint server provider).

The subject specificity of servers was seen as a natural way for preprints services to develop as it was seen as in line with the way researchers worked. However, even within disciplines, some referred to what they saw as significant differences between sub-disciplines, with some engaging in preprints and others not. Participants were agreed, however, that in disciplines where preprints were not established, although there might be some willingness to experiment, there was still a great deal of resistance:

“I think lots of my colleagues use them just as much as I do but then I think there are some colleagues that would never post to a preprint.” (Engaged researcher)

Finally, we note that adjacency between disciplines appears to be playing an important role. For example, disciplines close to those which posted preprints (e.g. those close to some areas of computer science or physics, which use arXiv consistently)
may be more favourably disposed to the practice compared to those from other disciplines.

**Preprints’ position in the landscape.** Perceptions of the role of preprints in the scholarly communication landscape were partly derived from understandings of and sympathy for open science and open access developments. It was clear that those who were supporters of wider open science developments generally supported increased use of preprints.

We note that peer-reviewed journals are still seen as essential. As such, preprints were for most interviewees “part of a new ecosystem” (Research performing organisation), but not a radical departure from or replacement for selective journals, although their potential to prompt more fundamental change was highlighted by some (see below). The level of integration of preprints in processes associated with submission to a journal was generally seen as low, but there was some awareness of provisions, for example, where a preprint could be transferred to a journal’s submission system. For most disciplines, these were not seen as fundamentally important, however, in determining usage or take up decisions.

One researcher referred to ongoing “scepticism” in “many fields” (Engaged researcher). One of the key reasons for this was that the ‘standing’ of preprints in different fields was seen to be very different. This was reflected, for example, in different perceptions of the value of citing preprints. In some cases, researchers believed that citing preprints was not acceptable in either papers or grant proposals.

**Potential benefits and challenges of preprints**

**Benefits.** Participants highlighted a number of (potential) benefits and challenges of preprints. For the most part, these correspond to those identified in the literature, but it is useful to see how these are articulated and prioritized by our participants. Table 5 summarises the main benefits of preprints as highlighted in the interviews, comparing them with the literature. We have attempted to rank these by the prevalence of certain points across the entire dataset of interviews and have also classified them as whether they create benefits at an individual or systemic level, that is whether they benefit individual researchers who practice them or the scholarly communication system as a whole.

Of the benefits highlighted by participants, early and rapid dissemination of research was the most frequent. As one engaged researcher succinctly put it:

>“The primary purpose of preprints is to communicate scientific knowledge as early as possible to as wide an audience as possible.” (Engaged researcher)

Preprints were described by one researcher as enabling “science in real time” (Engaged researcher), particularly with reference to the lengthy period of peer review and publication that often applied to article publishing. One participant representing a preprint server described the peer review process as, “often long and tortuous” (Preprint server provider), and a university representative described it as a kind of “limbo” (Research performing organisation) in which the research was not being read or used.

Achieving rapid and wide dissemination could in turn “accelerate” the pace of research itself (Other service provider). This might mean being able to “see a result that somebody else did and I can start working on it” (Research funder), even though this type of behaviour was described as appropriate only in very fast-moving sub-disciplines. Research could also make progress thanks to the reduction of “redundant work”, which makes

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Focus/ significance</th>
<th>Interviews</th>
<th>Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early and rapid dissemination</td>
<td>Systemic</td>
<td>✔️***</td>
<td>✔</td>
</tr>
<tr>
<td>Increased opportunities for feedback</td>
<td>Individual</td>
<td>✔️***</td>
<td>✔</td>
</tr>
<tr>
<td>Preprint servers as an outlet for ‘homeless’ results</td>
<td>Systemic</td>
<td>✔️**</td>
<td>✔</td>
</tr>
<tr>
<td>Advantages for early career researchers</td>
<td>Individual</td>
<td>✔️**</td>
<td>✔</td>
</tr>
<tr>
<td>Preventing scooping</td>
<td>Individual</td>
<td>✔️**</td>
<td>✔</td>
</tr>
<tr>
<td>Broader access to scientific research</td>
<td>Systemic</td>
<td>✔️**</td>
<td>✔</td>
</tr>
<tr>
<td>Increased citation counts</td>
<td>Individual</td>
<td>✔️*</td>
<td>✔</td>
</tr>
<tr>
<td>Preprints can support collaborations</td>
<td>Systemic</td>
<td>✔️*</td>
<td>✔</td>
</tr>
<tr>
<td>Preprints in some formats (e.g. xml) and with open licences are easier to text and data mine</td>
<td>Systemic</td>
<td>✔️*</td>
<td>✔</td>
</tr>
<tr>
<td>Much shorter time before research can be shared, so authors remain enthusiastic about it</td>
<td>Individual</td>
<td>✔️</td>
<td>✔</td>
</tr>
<tr>
<td>Preprints may reduce predatory publishing</td>
<td>Systemic</td>
<td>✔️</td>
<td>✔</td>
</tr>
</tbody>
</table>
it more likely for researchers to identify “the next big research question” (Research performing organisation) more quickly.

The benefits discussed were mostly seen in terms of communicating with a particular disciplinary community, but some interviewees also mentioned reaching a wider audience, including policymakers or clinicians, in a timely way – research reflecting the latest thinking in an area has more “news value” (Research performing organisation). One engaged researcher also emphasized the benefit of making the latest research available in the case of an outbreak of disease.

Participants were conscious of the importance of dialogue and interchange as part of the research process and so valued the potential for preprints to create opportunity for feedback. Some participants conceived of this from the point of the researcher receiving comments on their paper in order to make corrections – a kind of “debugging” (Engaged researcher) – or other improvements:

“…feedback from others to help you with your thinking and to improve your ideas.” (Research performing organisation)

It was commented that some preprint servers facilitate this in various ways, by for example allowing authors to solicit feedback or providing commenting or annotation features. Many respondents saw this in wider systemic terms, rather than just personal: enabling community engagement and discussion – what one funder called “community-oriented discourse on research results” (Research funder). The language of ‘community’ was particularly strong amongst many participants here. Some provided stories of their experience of this, with one researcher commenting on a particular paper that,

“the feedback from the community which we received through the preprint was at least as constructive and helpful as the official reviews from the journal.” (Engaged researcher)

Many of the participants agreed that preprints servers could be a useful outlet for otherwise ‘homeless’ research outputs, even though this goes counter to the emphasis of many of preprints as early versions of outputs later formally published elsewhere. The most common sorts of outputs mentioned were null results and replication studies, which would not satisfy journal requirements of novelty or significance. However, participants also mentioned older or under-developed papers that had not been formally published but could easily be deposited on a preprint server and would be of value.

Such an approach could be particularly useful for early career researchers, who benefit in general from posting preprints in order achieve greater “visibility” relatively quickly. This could be particularly useful in the case of funding proposals or job applications in order to demonstrate productivity, although several participants were quick to emphasise that formal publications would be preferable. For all researchers, in fact, preprints were seen as possible evidence of productivity but no real substitute for formal publications.

The benefit of preventing scooping was also prominent – preprints were a way to establish priority:

“If you’re an author the benefit of having a preprint in the public domain is it identifies the ideas as being yours, so it prevents you from being scooped because the ideas are down in time stamp against your name.” (Research performing organisation)

However, a small number of participants observed that this might cut both ways, since in disciplines where preprints did not have the standing of a citable resource, making research available in this way might cause the researcher to be scooped: “if you put something online and it’s not yet published [I worry] that it would be stolen by competitors” (Unengaged researcher). The strength of the language (“stolen by competitors”) is telling of the intensely competitive culture in which many researchers work and is in marked contrast to the language of community noted above. However, other researchers were sceptical about whether this was likely to happen.

This was partly because a preprint is available so widely. Breadth of availability was certainly a benefit seen by many participants, primarily in reaching as many members of their own research community as possible. However, many participants also mentioned access by a wider readership, including the general public. Either way, the benefit of increased usage of papers was seen as particularly important. However, once again, the benefits were qualified, with greater value being placed on formal publications or accepted author manuscripts:

“I think preprints are a very good kind of second option if you can’t access the published version and you can’t find the author’s accepted manuscript.” (Research performing organisation)

A number of participants were confident making work available as a preprint would increase citations, with the paper being citable in preprint form if available on a preprint server. This applied particularly in fast-moving areas. Getting work out into the community at an early stage also, it was suggested, increased the opportunity to form new collaborations. This is in many respects the more optimistic side of the idea of competitors stealing ideas – the possibility that researchers may be encouraged to develop a collaboration as a result of seeing the developing work of peers.

Challenges. Participants also recognised problems with preprints (summarised and compared with the literature in Table 6, with an indication of prevalence). It is worth noting, however, that participants often had a nuanced view of the challenges, commonly stating potential disadvantages but then themselves qualifying their criticisms or citing possible solutions.

Of the problems, the lack of quality of assurance was most widely discussed by participants, with a set of related issues clustering around this theme. In some cases, there was a concern preprints could simply mean lower quality:
“…so my worry would be with rapid publication by preprint that there would be an increase in the amount of poor quality science that’s available…” (Unengaged researcher)

However, more commonly, there was the view that a lack of quality assurance meant greater uncertainty and that readers had a greater responsibility to approach preprints critically. Words such as “caution” and “sceptical” were often used. The filtering role of selective journals was valued by many participants. Peer review, for all of its faults (and participants were not slow to point out its possible failings), was still highly valued:

“It’s really important that good reviewers have looked at the article and at the results section.” (Research funder)

Peer review was thought of not just as a safety net but as a process which often improved the quality of a paper. Most of the supporters of preprints, did not therefore see them as an alternative to or replacement for peer-reviewed papers but as a complement to them. Whether or not they valued informal comments, participants commonly pointed out that the use of feedback functionality of preprint servers was still limited. One preprint service provider observed that only 10% of preprints received comments.

Particular concerns around quality related to the media or members of the public latching onto unreliable findings, and on the harm a preprint containing errors could create in sensitive areas, especially associated with medicine or perhaps law. Several participants observed that people from non-academic contexts may not know the difference between a preprint and peer-reviewed article, and could therefore be more easily misled. Science journalists might potentially play a negative role. One participant said it was common for journalists to report findings from a peer-reviewed paper but “misinterpreting the results…and spreading the news without actually…understanding” the research (Engaged researcher). The risks associated with this were perceived to increase with preprints. Other participants acknowledged the responsibility of journalists, but were more optimistic, citing evidence of journalists using preprint servers responsibly and adding appropriate qualifiers to reports on preprints.

Several participants commented researchers themselves need to be aware of these problems and be cautious about how they published on controversial issues. Peer review did not mean that research papers were immune from such problems in any case.

There was a view expressed that basic screening provided by preprint servers, which was seen as very important by many participants, could address some of these concerns, but was still limited. There was consciousness that such basic checks themselves differed across different preprint servers and also that they would need strengthening in the case of servers dealing with medical outputs. One service provider working in this area suggested that preprint servers in the medical field might have to consider, among other things, “conflict of interest, financial disclosures, assurance that the work reported has been cleared by appropriate ethical review boards and assurance that data underpinning the article is available in an appropriate repository.” (Preprint server provider)

This is a very interesting development which would involve preprint servers undertaking additional quality checks than just the current basic screening, potentially blurring the boundaries between them and peer-reviewed journals.

### Table 6. Challenges of preprints. Mentions across the entire dataset: ✔️ = over 20 mentions; ✔ = between 10 and 20 mentions; ✔️️ = fewer than 10 mentions.

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Focus/ Significance</th>
<th>Interviews</th>
<th>Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of quality assurance</td>
<td>Systemic</td>
<td>✔️️️</td>
<td>✔️️️</td>
</tr>
<tr>
<td>Limited use of commenting/feedback features on the servers</td>
<td>Both</td>
<td>✔️️️</td>
<td>✔️️</td>
</tr>
<tr>
<td>Risk of the media reporting incorrect research</td>
<td>Systemic</td>
<td>✔️️️</td>
<td>✔️️</td>
</tr>
<tr>
<td>Possible harm in the case of sensitive areas</td>
<td>Systemic</td>
<td>✔️️️</td>
<td>✔️️</td>
</tr>
<tr>
<td>Questionable value of self-appointed reviewers</td>
<td>Both</td>
<td>✔️️️</td>
<td>✔️️</td>
</tr>
<tr>
<td>Information overload</td>
<td>Systemic</td>
<td>✔️️️</td>
<td>✔️️</td>
</tr>
<tr>
<td>The Ingelfinger rule – journals rejecting submissions if they have been posted as preprints</td>
<td>Individual</td>
<td>✔️️️</td>
<td>✔️️</td>
</tr>
<tr>
<td>Possible reputational damage to the depositor if the preprint is not good enough</td>
<td>Individual</td>
<td>✔️️️</td>
<td>✔️️</td>
</tr>
<tr>
<td>Possible ‘preprints wars’ in which the findings in one preprint are quickly attacked in another</td>
<td>Systemic</td>
<td>✔️️️</td>
<td>✔️️</td>
</tr>
<tr>
<td>There may be a rush to post low-quality research about popular topics</td>
<td>Systemic</td>
<td>✔️️️</td>
<td>✔️️</td>
</tr>
</tbody>
</table>
Participants were conscious of preprints creating what might be called a trust barrier. Whilst some of the determinants of trust for peer-reviewed papers might transfer to preprints (such as the overall shape of the paper, its authors, etc), some of the key determinants, the brand of the journal and its associated peer review processes, were missing. Participants were clear that preprints should be clearly marked as “not peer reviewed” and therefore treated with caution and handled responsibly. However, interviews also indicated possible contributors to trust that might apply to preprints, including: the preprint being widely discussed on social media, receiving comments online (e.g. on the preprint server), being cited, reported on by a magazine or newspaper, recommended by a colleague, and housed in a recognised preprint server.

Whilst there was enthusiasm amongst many participants for “community” based review and commentary on papers facilitated by preprint servers, there was some scepticism about the value of reviewers who “self-select themselves” (Unengaged researcher). This could either be because an author might invite comments from people who are expected to be positive about their work, or people commenting not understanding the area or use commenting to pursue personal agendas. The view was also expressed by some participants that the practice of commenting on preprints could cause difficulty with the formal peer review process, since people who had made public comments may be barred from undertaking blind peer review because of a conflict of interest.

There was concern also about information overload expressed by some participants, some of whom believed that preprints may inflate the number of papers being made available. However, other participants expressed scepticism of this, emphasising that the same number of papers were simply being made available earlier. There was also an acknowledgement that the filtering role played by selective journals was being removed from the process: some participants suggested that this could be at least partially solved with technology-based solutions, improving discoverability and filtering of content:

“I think there’s always the concern that people are worried that if I put something up there then it restricts where they can submit their paper.” (Engaged researcher)

“I have a duty to make sure that the work is peer reviewed in the best journals that we can get it into and if I rule certain journals out because I’ve submitted it as a preprint then I’ve kind of done a disservice to myself and those who I work with.” (Unengaged researcher)

Such “fears” themselves acted as a considerable barrier to uptake. Interestingly, one engaged and one unengaged researcher did show an awareness of SHERPA RoMEO (a database of copyright and OA self-archiving policies of academic journals) as a source of information, but this was not common.

Apart from permissions barriers, there were also fears that reputational damage could arise from premature release of preprints. The consequence of this was that sharing a preprint would be delayed until the authors were confident in it to avoid the possibility of any reputational damage. One researcher commented:

“I don’t think people in my field would just post off stuff that’s…terrible…because you’re still being judged on what’s going up there.” (Engaged researcher)

In addition, where work was produced by a team of co-authors, gaining agreement from the team was itself an important quality barrier to overcome before disseminating the research in preprint form.

We note that some of the challenges arising from the posting of preprints were only perceived to be substantial in cases where researchers or re-users behave unprofessionally or unethically; however, this does not just apply to preprints, but also to other areas of scholarship and publishing. A key feature of many of the interviews was that participants were rarely able to cite empirical evidence of either benefits or challenges of preprints. At best, personal experience or anecdotes about the experience of colleagues was being cited. As one engaged researcher said:

“I don’t have a lot of examples here, but certainly, you know, I hear anecdotes.” (Engaged researcher)

Another said, “I have heard stories on Twitter” (Engaged researcher). It was apparent from many of the interviews that there is a need for further empirical work in this area in order to provide an evidence base for practice.

Infrastructure and sustainability

**Infrastructure.** The view was commonly expressed by participants who commented on infrastructure that many technologies were already available to support use of preprints. Those mentioned by participants include repository and publishing solutions (open source, such as OSF Preprints, Eprints or
DSpace and proprietary technology, such as Figshare), but also the broader scholarly communication infrastructure (e.g. indexing via Crossref). However, several infrastructural issues were emphasised as being important by participants, including technology considerations (such as interoperability, search and discovery), as well as process considerations (including licensing, versioning management, and digital preservation). Functionality and usability considerations for individual services were also mentioned, such as search and annotation of preprints.

Interoperability was regarded as a priority by many participants, with standards often being seen as key. Use of digital object identifiers (DOIs) for preprints (which could then be linked to later versions of the paper) was seen as a particular priority, but other standards such as ORCID, and service providers, such as Crossref, were also mentioned as being important.

Use of standards was seen as an important enabler of effective search and discovery of preprints. Discovery was seen as being achievable largely through network-level discovery services, such as Google Scholar, but interestingly, even greater emphasis was placed on social media, particularly Twitter. It was common for researchers to report finding out about preprints of interest from Twitter, either by following particular individuals or signing up to Twitter feeds set up and managed by preprint servers (including automated Twitter posting when preprints are released). Twitter was reported to be the main way several participants found out about preprints in their field. Several researchers also mentioned gaining feedback on their own preprints as a result of posting links to them on Twitter, with comments sometimes being posted on Twitter itself or other social media rather than the preprint server. The importance of Twitter was also emphasised by service providers, with one representative of a preprint server stating:

“I would say that the momentum behind [name of the preprint server] owes a great deal to Twitter, and to Facebook, a bit less so. But nevertheless the effect is there. These are means of amplifying work once it has been posted.”
(Preprint server provider)

This is a significant finding. The fact that part of the infrastructure upon which preprint services are currently reliant includes generic discovery services, such as Google Scholar, and social media, particularly Twitter, needs to be taken into account in considering future developments of preprint services. Usage (or at least availability) of Twitter was assumed by participants to be widespread – a reasonable assumption in most Western countries, but not in others, particularly China.

Open licensing was also seen as an important infrastructural issue, which enabled interoperability and discovery. Preprint servers typically offer depositors Creative Commons (CC) licence options and there was some discussion in our interviews of the best one among these. Some authors were clearly confused about the options available to them when depositing their work. Some were aware of various requirements (including from funders) but were also wary of the possible consequences of signing particular licenses on a preprint when it comes to formal publication at a later stage.

Management of versioning was mentioned by a number of researchers although experience of this was mixed, with many authors not taking advantage of the facilities offered by preprint servers in terms of tracking versions. Preprint service providers were divided about whether these services allowed for withdrawal of items, but this was regarded as important by a number of participants in the case of misleading results or disputes between co-authors, for example.

Digital preservation, the final infrastructural issue discussed by participants at any length, was regarded as important in principle but often de-prioritised in practice because of its costly nature.

Policy. Participants identified key policy issues at all levels of the policy stack: publisher, funder and institutional. A cross-cutting issue applying to all policy levels discussed by many participants was, however, the value of preprints as reflected in and recognised by policy. Preprints were recognised to be valuable in providing early open access to research but their value in the scholarly record was commonly qualified by participants. Put simply, they were not considered as valuable as the author accepted manuscript (AAM) or the version of record (VoR), a view reinforced in the minds of many by the fact that many funder and government OA mandates did not have any requirements relating to preprints, but focused rather on AAMs and VoRs. This affected the extent to which different actors regarded preprints as a policy or operational priority. The concept of “standing” (Neylon et al., 2017) is relevant here, with, as an example, standing relating to whether preprints are seen as an appropriate object for evaluation in exercises such as the UK’s Research Excellence Framework, about which there remains ambiguity. One representative of a university stated:

“they’re not acceptable for REF, so they’re not even part of the equation. So it’s the author’s accepted manuscript the currency we deal with.”
(Research performing organisation)

The value of preprints was seen rather in the extent to which they are a component of a more open research system. Whilst they were seen by some as a useful counterbalance to expensive OA publishing funded by APCs (article processing charges), mostly they were seen as important in terms of their contribution to the overall open science agenda:

“...in terms of preprints, I’m more interested in a different problem, which is the problem of the opening up the whole research endeavour throughout the whole research process through the collection of data to the analysis of data to the curation of data through the writing research protocols. And then analysing the...the results and writing the software and all that stuff. Each of those things themselves should be considered a research output. The preprint is towards the end of that.”
(Research performing organisation)
This was important at all levels of the policy stack. Institutional-level policy was seen by many participants in this light. Preprints might be encouraged in a general way by institutional policy, but deposit of preprints would not normally be mandated. In fact, some participants reported that institutional policy was silent on the matter, and in practice preprints were not accepted as deposits to the IR.

Funders, similarly, often encouraged use of preprints and allowed them to be cited in grant applications, but they were not included as an acceptable form of an output which should be made available OA as part of a funder OA mandate. One preprint provider suggested funders might in future mandate use of preprints but pointed to only one small US funder who was currently doing so. There were other possible funders policies that could encourage change:

“The other thing that funders really could do is to make very public the fact that they will allow a grantee to cite preprints in her progress reports and then her grant renewal application.” (Preprint server provider)

Publishers’ policies were criticized by many participants as preventing authors from depositing preprints either before submission (the Ingelfinger rule) or after acceptance (contractual exclusions or embargoes). The environment was often seen as a confusing one for authors, something which was itself discouraging.

Financial sustainability and business models. Four models for delivering preprint services emerged from the data, each of which has different implications for sustainability and funding, an issue which was emphasised as being concern by many participants:

1. Standalone preprint servers using in-house technologies e.g. bioRxiv, arXiv;
2. Standalone preprint servers using third-party technologies e.g. ChemRxiv using Figshare infrastructure;
3. Publisher-supported preprints e.g. PeerJ, F1000; and
4. Publisher posting preprints to a preprint server e.g. PLOS partnership with bioRxiv.

Models 1 and 2 involve an ‘author-driven’ mode of posting preprints; Models 3 and 4 are a new ‘publisher driven’ mode.

Model 1 is the ‘classic’ model of preprints, apparently assumed by most participants in their interviews. Model 2 is a version of 1 in which some infrastructure is outsourced to a third party, something which might help to enable sustainability. Models 1 and 2 also include publisher-operated services such as Preprints.org and SSRN, which are owned by MDPI and Elsevier, respectively. These are not classified as publisher-supported (Models 3 and 4) but as standalone servers, because they still follow a paradigm of individual authors being responsible for posting their own preprints.

In this respect, Models 3 and 4 are different, since they involve the publisher rather than then author driving the preprint submission process, either through the publisher itself providing a preprints service or by depositing on the author’s behalf. Models 3 and 4 move away from the traditional ‘author-driven’ preprints practices – where the author voluntarily deposits a paper as part of their own workflow, separately from submission to a journal – to a ‘publisher-driven’ preprints model, where the publisher to whom a manuscript has been submitted makes it available as a preprint. This is a fundamental shift which has major implications for the way preprints are conceived and the way preprints services are configured.

Financial sustainability was a concern for many of the participants with one describing the financial sustainability of a number of preprint services as “fragile” (Research performing organisation), often being based on short-term project funding. One preprint service provider described the work they did in partnership with the Center for Open Science as based largely on the goodwill of “volunteers”. One participant commented that independent preprint servers have no business model associated with them other than grant funding. Although this was sometimes seen as a mark of immaturity, with experimentation in funding and sustainability models ongoing, some participants did point to problems in the funding of arXiv, a well-established preprint server, which had been through several periods of uncertainty with regard to its funding during its history. These kinds of concerns, raised by many of the participants, relate to models 1 and 2 particularly.

Despite this, most participants seemed to favour a not-for-profit approach to preprint servers, some for reasons of sustainability, others emphasising the need for independence. Developments associated with models 3 and 4, and even 2, were, therefore, viewed with suspicion or even hostility by some participants. A number expressed concerns about the consolidation of services associated with academic workflows into the hands of a few commercial companies, including Elsevier ownership of SSRN, and Digital Science’s ownership of Figshare. This was seen as potentially jeopardising the independence of the relevant preprint services, even if they were currently operating as standalone. The language used by some participants to describe this development was in some cases strong, with one describing the sale of SSRN to Elsevier as “a huge betrayal of trust” (Engaged researcher) and another observing that this and similar developments have led to “power concentration” (Research performing organisation) in a small number of commercial providers. On the other hand, one representative from the preprint provider observed, “I don’t think that your average researcher thinks about that” (Preprint server provider); and it was indeed evident that many researchers amongst our participants did not show an awareness of such issues.

The future of preprints

There was considerable uncertainty amongst participants about the future role of preprint servers. When questioned whether preprint servers could form a significant part of a system of scholarly communication which would be an alternative
to or replace peer-reviewed journals, most were sceptical. There was, however, some discussion of the potential value of ‘overlay journals’ – where virtual journals are created from content held in preprint servers, having been peer-reviewed and selected after their circulation. Some suggested that automated filtering rather than human-based peer review might have a role to play in creating overlay services. There was some awareness of experimental work in the overlay area, but few were able to identify working examples.

It was clear from comments of participants that if preprints were to play a more significant role in scholarly communication, major improvements to the preprints infrastructure would be needed. This would include incorporation of preprints into scholarly and publisher workflows, provision for production of preprints in standards-based formats (e.g. XML) and greater consideration of preservation services. All of this would require major investment. However, even in a system where preprints did not replace existing channels of communication, such as journals, many of these developments were still considered necessary in order to make the preprint infrastructure more robust.

Whilst uptake of preprints was seen by many to be increasing, the role of policy in this area was uncertain. It was seen as particularly problematical for use of preprints to be mandated, as opposed to encouraged, by funders. There was a clear perception that preprints should be adopted by researchers who see the benefits themselves, rather than in response to a mandate:

“There needs to be an intrinsic interest of the research community to communicate via preprints. I don’t think preprint posting can be enforced top-down or from anyone other than the research community and specifically the disciplinary communities themselves.” (Research funder)

It is noteworthy that this view was expressed by a policy maker.

**Discussion and conclusions**

The findings of this research clearly relate to the innovation adoption decision process identified in IDT, beginning with developing knowledge of innovation to confirmation of its adoption (Table 7).

It was evident that our participants were at different stages in the adoption process and this is a reflection of their peers’ and subject communities’ practices. Knowledge of preprints is rising but many are still not beyond the persuasion stage. Community norms remain crucial and have not shifted in many cases, therefore constraining individuals’ decisions. There was, nevertheless, some willingness to experiment, particularly amongst general OA supporters. There was some awareness of potential benefits becoming evident in practice but still at low levels. The rate of adoption is influenced in our data by a number of factors highlighted in IDT (Table 8).

It is worth mentioning that, in some cases, it is not clear whether researchers do not post preprints because there is no

<table>
<thead>
<tr>
<th>Table 7. Findings in relation to innovation adoption decision making factors from innovation diffusion theory.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Decision making step</strong></td>
</tr>
<tr>
<td>1. Knowledge</td>
</tr>
<tr>
<td>2. Persuasion</td>
</tr>
<tr>
<td>3. Decision</td>
</tr>
<tr>
<td>4. Implementation</td>
</tr>
<tr>
<td>5. Confirmation</td>
</tr>
</tbody>
</table>

It is noteworthy that this view was expressed by a policy maker.
Table 8. Findings in relation to the rate of innovation adoption from innovation diffusion theory.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Components</th>
<th>Summary of current findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived attributes</td>
<td>Relative advantage</td>
<td>The potential value of preprints is clear to most although the existence of a relative advantage for their use is not.</td>
</tr>
<tr>
<td></td>
<td>Compatibility</td>
<td>The compatibility of preprints with existing practices and systems is perceived as variable and often unclear.</td>
</tr>
<tr>
<td></td>
<td>Complexity</td>
<td>Perceived complexity of publisher policies and additional workload associated with deposit can be a disincentive.</td>
</tr>
<tr>
<td></td>
<td>Trialability</td>
<td>Preprint servers can be a strong basis for experimentation.</td>
</tr>
<tr>
<td></td>
<td>Observability</td>
<td>Some participants have observed the benefits of preprints based on colleagues’ experiences.</td>
</tr>
<tr>
<td>Communication channels</td>
<td>Mass media</td>
<td>The widespread use of Twitter is a noticeable characteristic which enables discoverability of preprints and growing awareness of preprint servers. It should be noted, however, that use of Twitter is not universal e.g. in China.</td>
</tr>
<tr>
<td></td>
<td>Interpersonal</td>
<td>Personal recommendations are crucial in encouraging uptake.</td>
</tr>
<tr>
<td>Nature of the social system</td>
<td>Norms</td>
<td>Preprints are dealt with differently based on whether people are early adopters of open science practices. In most other cases, preprints are considered as an important development, but scepticism still has to be overcome (e.g. with respect to practical advantages, funding streams and long-term preservation). Disciplines also play a significant role, as preprints are widespread in some research areas but only emerging in many others.</td>
</tr>
<tr>
<td></td>
<td>Interconnectedness</td>
<td>Communities are important in shaping perceptions of the value of preprints. We found that two factors related to interconnectedness affect attitudes towards preprint posting: (i) the adjacency to disciplines where preprint posting is common (e.g. disciplines close to others within the scope of arXiv); and (ii) pressures and attitudes of close peers, particularly co-authors.</td>
</tr>
<tr>
<td>Extent of change agents’</td>
<td>Promotion efforts</td>
<td>Promotion efforts are limited, and it is unclear whose role this should be. Stakeholders in the open science arena are promoting preprints within their circles and online. Some funders (e.g. EC, Wellcome) are making explicit efforts to promote preprints, but more significant and broader support (including from publishers) is likely to be required for higher uptake.</td>
</tr>
<tr>
<td>promotion efforts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of innovation decision</td>
<td>Optional, collective or authority</td>
<td>Likely to continue to be optional for individuals influenced by their disciplinary community. Mandated uptake is seen as unlikely although encouragement from policy makers is likely to be strengthened.</td>
</tr>
</tbody>
</table>

Discipline-appropriate server for them, or there is no server because researchers in the field do not (want to) post them. The data shows that some researchers are happy to be early adopters often because of sympathy with general open science and open access goals, as well as being motivated by the potential benefits they see in their own use of preprints. Others remain sceptical, fearing that journals may reject their submissions, which is still the case in some areas (Mallapaty, 2019), and questioning the value of circulation of pre-peer-reviewed outputs. The current environment for many disciplinary communities is therefore characterised currently by some experimentation, but also by uncertainty and fragmentation.

Another key issue is trust. Trust is an essential feature of scholarly communication and was a recurring theme in the data, along with responsible posting and use of preprints. Nicholas et al. (2014) have shown that “researchers play down difficulties of establishing trustworthiness, not because there are none, but because they have well-developed methods of establishing trust.” However, preprints cut across those methods and create new ambiguities and uncertainties. Many of our participants were conscious that new norms were needed (COPE Council, 2018) in this space and it was often the apparent absence of these that limited enthusiasm for preprints.

The preprints landscape is rapidly changing, and disciplinary communities are at different stages in the innovation diffusion process. The current high level of experimentation means that a one-size-fits-all approach to preprints is neither feasible or appropriate at present (if it ever will be). Something that will clearly play a role in the success (or lack thereof) of preprints is cooperation between the range of stakeholders involved: though some of the issues we highlighted might appear independent of one another, we note that the majority of these affect multiple stakeholders at once.

Our findings have given rise to a number of key questions that we believe need to be addressed so that preprints can be supported sustainably in the future. Active engagement with these questions should lead to improved clarity and provide solid foundations for policy development and implementation. The questions and their relationships to the different actors involved are illustrated in Figure 1. These include some key questions that need to be addressed by particular stakeholder groups, such as funders or preprint server providers. We have also added publishers as a separate group to this figure, taking account of issues that have arisen from the groups considered in this current study. However, there are also a large number of questions that need to be addressed through dialogue.
between different stakeholder groups. These are illustrated in the figure. One of the key challenges is to find channels for such dialogue to take place in order to develop solutions which are widely accepted.

This is in many respects an agenda for further research, discussion and policy design. We expect, following a rapid rise in preprints since about 2013, that many of these questions will come to be seen as increasingly important over the next five years. The urgency with which they are addressed, and the ways in which they are answered by the different stakeholders, will shape the role that preprints play in scholarly communications in the future.

**Data availability**

**Underlying data**

The authors confirm that, for approved reasons, some access restrictions apply to the data underlying the findings. Data underlying this study cannot be made publicly available in order to safeguard participant anonymity and that of their organisations. Ethical approval for the project was granted on the basis that only extracts of interviews would be shared (with appropriate anonymisation) as part of publications and other research outputs. In order to share data with other researchers, the participants must be contacted and consent to this data being released. In order to request data release, other researchers should contact the corresponding author or Chair of the University of Sheffield Information School Research Ethics Committee (ischool_ethics@sheffield.ac.uk).

**Extended data**


Extended data.csv contains our interview questions, split by stakeholder group and mapped to innovation diffusion theory.
Extended data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

Grant information
This study has been funded by Knowledge Exchange (KE), a group of national organisations from six European countries supporting research infrastructure and services to enable the use of digital technologies to improve higher education and research: CSC in Finland, CNRS in France, DEFF in Denmark, DFG in Germany, Jisc in the UK and SURF in the Netherlands.

Acknowledgements
This article draws on the findings of an initial phase of work on the preprints landscape (Chiarelli et al., 2019b) supported by the KE Task and Finish Group on preprints.

References


ASAfpbo: University policies and statements on hiring, promotion, and journal license negotiation. (n.d.). Reference Source


Publisher Full Text


Khera R: Sharing science at today’s pace: an experience with preprints. BMJ. 2018. Reference Source


Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Reference Source
Pinfield S, Middleton C: Researchers’ adoption of an institutional central fund for open-access article-processing charges: A case study using Innovation Diffusion Theory. SAGE Open. 2016; 6(1).

Publisher Full Text

Reference Source
Pittman M: Preprints as a Hub for Early-Stage Research Outputs, 2018.

Publisher Full Text

Reference Source

Publisher Full Text

PubMed Abstract | Publisher Full Text

PubMed Abstract | Publisher Full Text

Reference Source

Reference Source

Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

PubMed Abstract | Publisher Full Text | Free Full Text
Velterop J: Of subscriptions and article processing charges. 2018.

Reference Source

Publisher Full Text

PubMed Abstract | Publisher Full Text

Publisher Full Text
Open Peer Review

Current Peer Review Status:  ?  ?

Version 1

Reviewer Report 05 August 2019

https://doi.org/10.5256/f1000research.21512.r50950

© 2019 Polka J et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Jessica K. Polka  
ASAPbio, San Francisco, CA, USA

Naomi C. Penfold  
ASAPbio, San Francisco, CA, USA

In “Preprints and Scholarly Communication: Adoption, Practices, Drivers and Barriers,” Chiarelli and colleagues present a literature review, commentary, and a summary of attitudes and perceptions towards preprints found by interviewing 38 stakeholders. The interviews convey many commonly-held perceptions about the benefits and disadvantages of preprints, and the authors suggest questions for further exploration and testing. With revisions (clarifying the origin of claims, contextualizing the responses from study subjects, and correcting factual information in the text) this paper will become a very useful addition to the literature.

Major comments

- Design and Subjects:
  - We encourage authors to rely on a more robust review of published literature (including the grey literature) to complement their interviews, as the published literature contains useful context for comments which are obfuscated by the process of anonymous reporting (For example, the sentence, “One preprint service provider observed that only 10% of preprints received comments.” would be more useful to readers if accompanied by a reference with attribution: https://asapbio.org/biorxiv)
  - Since this work may be of interest to a wide range of readers, we recommend including more detail regarding the chosen experimental design.
    - Please provide justification for the experimental design, for example, why were interviews used instead of a survey?
    - What are the methodological considerations that would be helpful for readers outside of qualitative research to understand: limitations, constraints, advantages over more quantitative methods?
    - Please define semi-structured interviews — how were researchers encouraged to expand on thoughts?
• Please define “purposive heterogeneous sampling approach.”
• Please clarify why the sample size is 38.

• Please provide more information about the subjects to help readers assess relevance:
  • What is their geographical context?
  • How knowledgeable were subjects, and do they bring bias for or against open science into their statements?
  • How did interviewees arrive at their opinions? How much close experience do they have with preprints? Were participants asked to answer only within their own expertise?
  • There are quoted statements that reflect inaccurate understandings and are not always qualified by the author: for example, “relating to whether preprints are seen as an appropriate object for evaluation in exercises such as the UK’s Research Excellence Framework, about which there remains ambiguity. One representative of a university stated: “they’re not acceptable for REF, so they’re not even part of the equation. So it’s the author’s accepted manuscript is the currency we deal with.” (Research performing organisation)” This is inaccurate and misleading (see https://www.ref.ac.uk/publications/guidance-on-submissions-201901/) but there is no further commentary from the authors. Please add clarification that the interviewee responses are not necessarily factual. (It would be interesting to compare interviewee opinions with known facts to understand which topics are well-understood.)

• Clarity of presentation:
  • Please provide more quantitative context to the opinions expressed. For example, for phrases such as “Some researchers” it would be useful to be more specific. Please also be more quantitative in assessment of the themes — it is useful to know if a sentiment was expressed by a majority or minority, as shown clearly in tables of benefits and challenges. Please provide aggregate data of thematic analyses, not just quotes, and use specific numbers and proportionality (Two of Five researchers mentioned X as a benefit).
  • The paper combines elements of a literature review, white paper/policy piece, and a qualitative study. It would be helpful to clarify the origin of ideas by separating the work clearly into Results and Discussion sections. (As one example, under the heading “Financial sustainability and business models,” it is unclear whether the proposed taxonomy was expressed by interviewees or generated by the authors. Similarly, it is unclear whether use of the term “seminal” comes from interviewees or authors.)

• Readability:
  • Please use active voice and concise language to improve readability.
  • Bring interview questions and list of participants into this report — it is hard to interpret findings while having to continuously refer back to other sources, e.g. the 2019b interview structure paper.

• Claims:
  • The abstract sets the reader up to expect a review of preprint servers that have emerged since 2013. However, the methods section discloses that coverage of the study is limited to
biology, chemistry, and psychology. It would be helpful to readers to state this up front. It is also unclear how this scope was applied in the selection of study subjects. It is difficult to determine if the scope influenced the selection of participants as listed in https://zenodo.org/record/2654832#XTeUB5NKjs0 – the questions listed at https://zenodo.org/record/3240426 do not seem specific to these three disciplines and the quoted answers in this report suggest interviewees’ responses were not limited to these three disciplines. Please clarify how the scope of the interviews was constrained to the stated subjects, or if not, please revise the stated scope to more accurately reflect the scope employed.

- The abstract does not reflect the content. The interview responses are summarised in general terms, not related to defined scope or the newer servers. A substantial proportion of this article is review material: either in the explicit literature review section, or as author-contributed content interspersed with summaries of interviewee responses. For example, conclusions listed in the abstract appear to be drawn from the literature review.

- “Our study is the first using empirical data to understand the new wave of preprint servers” — this is not supported by the study content: no results are specifically related to the newer servers identified in the introduction; quoted excerpts indicate that interviewees responded in general about all servers and disciplines. Furthermore, prior surveys and bibliometric analyses have looked at a subset of these servers, either individually or in groups (eg https://osf.io/5fk6c/).

Minor comments

We recommend making the following amendments to improve the clarity and accuracy of the report.

In the Introduction and literature review:
- The summary of the longer history of preprint efforts across disciplines, would benefit from incorporating and citing “The Brief Prehistory of Preprints” (Cobb, 2017)
- Zuckerberg Foundation should be corrected to Chan Zuckerberg Initiative.
- Discussion of new servers would be improved with a note as to their current size, and with awareness that not all new ‘preprint’ servers are exclusively for preprint content (many COS servers include postprints that are not possible to easily distinguish from preprint versions).
- “This paper aims to explore...by investigating current practices, drivers and barriers to [preprint] use.” The interviews indicate perceptions and attitudes but do not reveal actual behaviours or practices. Please clarify how this study examines current practices.

Figures and tables:
- Fig 1: To increase accessibility, please provide as raw text with headings or tags to indicate which stakeholder group you believe is most relevant

Literature review:
• The authors focus on four main issues — how were these arrived at? Was the literature review systematic in any way?

• Given the emergent phase of preprint infrastructure in biology, chemistry and psychology, and as the authors note that empirical data is lacking, it may be equally valuable to include more grey literature (blogposts, webpages) in this review, given that several of the cited articles are opinion pieces, albeit as editorials or peer-reviewed review articles. Of note, we find personal blogs are a useful source of stories that detail benefits and drawbacks of preprints from personal experiences.

• “Dis-benefits:” is not a common term, perhaps use “drawbacks,” “concerns,” “disadvantages” or “potential risks”.

• When discussing the lack of quality assurance, the authors seem to make an implicit assumption here that journal-led peer review assures quality. It would be fair to contextualise this section with some literature on peer review to understand variety of peer review outcomes on quality of paper and validity of findings.

Results:

• In the discussion of preprints in the medical field, please note that the concerns raised are being addressed by medRxiv. https://www.medrxiv.org/

• In the table of business models:
  • Note that bioRxiv uses 3rd party technology, Highwire.
  • “Publisher driven” models are not new; PeerJ is older than bioRxiv, Nature had Nature Preceedings in the early 2000s, and bioRxiv has had J2B preprints since its inception.
  • How does the outsourcing of infrastructure to a third party enable sustainability? Arguably, depending on the third party, this leaves the service vulnerable to changes in service entirely outside its control.
  • Individual authors can still submit to PeerJ.
  • Most publisher-mediated preprint submissions are voluntary for authors (a notable exception being F1000).
  • Regarding the statement of subject preferences: we have conducted a survey on these preferences that may provide additional context: https://asapbio.org/asapbio-funders-workshop

The following suggestions may also help improve the paper but are not essential.

In the Introduction and literature review:
  • Open Research platforms could be discussed in context of F1000Research, since this is the model they are derived from (in collaboration with F1000).
  • Please clarify the age of PeerJ and Preprints.org in comparison with bioRxiv.
PLOS is not the only journal that deposits preprints to bioRxiv using the J2B mechanism — several do, please refer to the list of journals offering direct transfers to bioRxiv (https://www.biorxiv.org/about-biorxiv).

Regarding trust in preprints, it would be relevant to mention or cite the current effort by COS to better understand indicators of trust in preprints: https://cos.io/about/news/center-open-science-receives-grant-alfred-p-sloan-foundation-study-trust.

Figures and tables:

- Fig 1 introduces some interesting questions, however many of these are related to topics that are brought up for the first time in this figure.

- Table 6: The concern of reputational damage is indeed in the literature; please see point 5 here: https://smallpondscience.com/2017/07/24/whats-up-with-preprints/.

Literature:

- “However, Bourne et al. (2017) controversially extend the definition to include “a paper that has been peer reviewed and...was rejected, but the authors are willing to make the content public”.” — Please provide reasoning or citation as to why this is controversial.

- Accessibility — note that preprints are not necessarily open access, where they are not provided with permissive licensing, and this is certainly the case for bioRxiv.³

- One potential benefit to ECRs that is not mentioned here is being able to demonstrate productivity to funders and hirers. This would be useful to append to the current paragraph, given it is raised in the interview questions.

- “There also appears to be little acknowledgement that the claim stands in tension to the one cited earlier that preprints often differ little from final published versions.” — There is only tension if both claims concern all preprints. It is possible for there to be preprints that improve through a lot of feedback as well as preprints that undergo no/very little change.

- Please use unshared or unpublished instead of “homeless”.

- We recommend avoiding the term “policy stack,” as there is no dependent relationship among these different groups/stakeholders.

Methods:

- This section ignores the efforts of researchers as change agents.

Results:

- Please provide evidence for this statement: “we note that adjacency between disciplines appears to be playing an important role.”

- “Many of the participants agreed that preprints servers could be a useful outlet for otherwise ‘homeless’ research outputs, even though this goes counter to the emphasis of many of preprints
as early versions of outputs later formally published elsewhere.” — this attitude may reflect a lack of publication venues for such content, not a lack of desire to publish these other outputs in a peer-reviewed destination.

References
1. Narock T, Goldstein E: Quantifying the growth of preprint services hosted by the Center for Open Science. 2019. Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
No

Are the conclusions drawn adequately supported by the results?
Partly

*Competing Interests:* We are employed by ASAPbio, a non-profit that promotes the productive use of preprints in the life sciences.

*Reviewer Expertise:* We have backgrounds in cell biology and neuroscience, and are currently working to promote the productive use of preprints.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Reviewer Report 23 July 2019

https://doi.org/10.5256/f1000research.21512.r50461

© 2019 Sarabipour S et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Sarvenaz Sarabipour
Institute for Computational Medicine, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA

Humberto Debat
Center of Agronomic Research, National Institute of Agricultural Technology (IPAVE-CIAP-INTA), Córdoba, Argentina

Summary:

In this manuscript Chiarelli et al. assess preprints from the point of view of a diverse group of stakeholders. The authors perform an analysis based on 38 interviews with representatives of research funders, research performing organizations, preprint servers and service providers; and researchers (engaged and unengaged). Additional discussion is provided on the benefits and challenges of preprint posting, along with issues such as infrastructure and financial sustainability and the definition of a ‘preprint’ in different communities, and the impact this has on further uptake. This study provides a thorough investigation on an emerging and relevant topic, giving a platform to the opinions of under-represented stakeholders on the discussion of preprints: a key driver of the transformation of the scholarly publishing landscape.

Comments:

We would like to ask the authors to address the following:

The introduction may benefit from additional literature to balance the many sections. For instance:

1. On Page 3: “However, skeptics have questioned the value of preprints and even suggested they may be dangerous – circulating versions of articles before they have been quality controlled by peer review may lead to unnecessary risk, particularly in disciplines like medicine (Sheldon, 2018).” The only negative note on preprints by Sheldon 2018 was well rebutted by at least 5 publications:
   - Maintaining confidence in the reporting of scientific outputs - Sarabipour et al. (2018)¹.
   - Preprints are good for science and good for the public - Sarabipour (2018)².
   - Preprints help journalism, not hinder it - Tennant et al. (2018)³.

Articles 1-4 above address the relationship between scientific reporting and journalism. Further to the authors remark on “unnecessary risk particularly in disciplines like medicine” - preprint servers carry a highly visible note on preprinted manuscript stating that these research products are not peer-reviewed.

Regarding biomedical and medical preprints and risk to public: This is not a compelling argument since care is delivered to patients by physicians. MedRxiv was recently established and is accepting manuscripts to accelerate medical research. The server front page notes that: “Caution: Preprints are preliminary reports of work that have not been peer-reviewed. They should not be relied on to guide clinical practice or health-related behavior and should not be reported in news media as established information.”

2. On Page 3: “the current and potential future role of preprints as a vehicle for scholarly communication” - there are and will be many role(s) for preprints in the current and future of the scholarly endeavors and they are summarized here:
3. On Page 4: “much of the literature is still to be found in editorials and opinion pieces rather than data-driven research” - the fact is that the number of data driven literature on preprints is growing rapidly and are noteworthy. Examples are:
   - Tracking the popularity and outcomes of all bioRxiv preprints - Abdill & Blekhman (2019)
   - Releasing a preprint is associated with more attention and citations - Fu & Hughey (2019)
   - The effect of bioRxiv preprints on citations and altmetrics - Fraser et al. (2019)
   - Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature - Carneiro et al. (2019)

4. On Page 4: “Accessibility (4) is crucial, with a preprint normally defined as being (or assumed to be) openly available: it “can be viewed without charge on the Web” and “Therefore, the venue for distribution of preprints is often assumed to be a freely-accessible server of some kind”. It is a fact that all preprint servers are freely accessible to any human being that has access to internet around the world: free of charge to submit by authors, free of charge to read by all anywhere on the planet.

5. On Page 5: “To say that an output is “deserving” of dissemination is, of course, a value judgement and difficult to demonstrate for each deposit as it is made, but it is one that is implicit in much of the discourse on preprints.” - this is not entirely true. The value of preprints are well known/articulated/discussed at this point as preprints accelerate research and altmetric attention to scientific work: Please see references 8-11 above. Preprints have been deposited on arXiv for nearly three decades now and are invaluable to the physical sciences community. Other forms of pre-peer review material such as computer code and protocols have been disseminated open source and their value is well noted in various fields.

6. On Page 5: “dis-benefits of preprints” - unless the authors may provide any evidence of “dis-benefits of preprints” besides one note (Sheldon 2018), it is best to avoid using this word “dis-benefits” and consider rephrasing to “perceived dis-benefits”. Please see reference 1-5 above.

7. On Page 5: “Early dissemination can be useful to some particular members of the scholarly community, with early career researchers (ECRs)” - preprints and all other forms of scholarly pre-publication material are beneficial to ALL researchers not just ECRs. Timely release of scientific results in the form of preprints and other forms such as code and data bases have already accelerated research by encouraging sharing of ideas, resources and as discussed in previously suggested references above.

8. On Page 5: “Perhaps the most prominent criticism of preprints relates to this last issue: the lack of quality assurance through peer review (Sheldon, 2018). As well as a general concern about lowering quality standards, lack of quality control has been seen as potentially dangerous as “reports that have not undergone formal peer review [organised by a journal] could be misleading” (Lauer et al., 2015). Furthermore, uncertified science might be reported prematurely in the media and might even give rise to ‘fake news’ (Sheldon, 2018). Some insist that, at the very least, the opportunity to disseminate knowledge rapidly without peer review may encourage academics to produce low-quality outputs on fashionable topics (Teixeira da Silva, 2017).” The authors should declare that there are no evidences supporting all these bold statements. The Sheldon (2018) news article in fact explicitly states its speculative nature in the title which reads “preprints could promote confusion and distortion”. Please see references 1-5 above.

9. On Page 5: “Others have gone further and questioned the value of self-appointed reviewers, as opposed to those selected by journal editors (see the issue of “self-policing” highlighted by Harnad,
1998). This is a misleading claim. Voluntary reviews of preprints are done by researchers in that field or community and are not self-appointed by the authors. The open access nature of preprints allows the potential assessment of manuscripts by more than 2-3 (more than journal reviewers) normally and there is no evidence on these reviews having different quality than journal peer-reviews.

10. On Page 5: “Preprint posting, however, is unlikely in any case to substitute for the valuable role played by selective journals in filtering content”. This is an unfounded claim as a system of preprint servers and overlay journals may very well replace the outrageously expensive and time-consuming system of profit journal publishing in under a decade. Please see reference 12.

11. On Page 5: “Whilst this convention has come under criticism and been withdrawn by some publishers, it still exists for some journals, e.g. in medicine and chemistry (Lauer et al., 2015; Teixeira da Silva & Dobránszki, 2019)” - this sentence is misleading. In the current scenario, authors should consider changing “some” to “most” publishers.

12. On Page 6: “For example, it is not universally agreed when an output should be citable (in the literature, funding proposals or promotion cases)”. It is now becoming more and more agreed upon that preprints should be cited in articles and research grant proposals. Major funding bodies such as NIH, CZI and Wellcome Trust allow and encourage citation of preprints. Many journals allow citation of preprints and other research outputs (eLife and PLOS family of journals are examples).

13. On Page 6 and other pages: “Scooping”, and “a claim of precedence”: Scooping is a perceived concern of researchers and is unfounded. Many researchers present unpublished research at conference. Perceived preprint concerns and claim of precedence have been discussed in reference 6 above.

14. On Page 6: “or when it can be used to establish a claim of precedence” and “In disciplines where a preprint is not considered appropriate to establish precedence, it has also been suggested that making a preprint available may actually encourage research to be scooped by rival researchers who publish in a recognized journal before the preprint authors (the ‘flip side’ of the priority claim argument above) (Kaiser, 2017).” Regarding these repeated sections on scooping and priority claims:

- 2-3 papers published via preprints or journals in close temporal proximity of each other have all been projects 2-4 years in making so it is very difficult to assess who did it first.

- Who gives researchers priority? A preprint that is posted online has a date and that gives precedence to the work. In reality, as researchers we all know that priority is given to researcher by their field and their peers. Scientific research is often presented at conferences well in advance of preprints or journal publication. As we know what matters in research is scientific reputation and that is given to researchers by their peers and field during a career in scientific research.

- Multiple papers published via preprints or journals in close proximity of each other only show the reproducibility of one another are need to be celebrated not feared so repeated arguments on scooping and priority claims do not move a positive conversation on research forward specially for early career researchers and the future generations of researchers.

15. On Page 6: “Some have argued that initiatives such as Declaration on Research Assessment (DORA), with its emphasis on the quality of the output rather than venue of publication, promote use of preprints (Polka, 2018).” It is noteworthy that DORA has now become a major international initiative and is signed by over 1,400 organizations including major universities and other research institutions and 14,000 researchers.
Other comments to the MS text:

A) The Title: Authors could consider an alternative title “Preprints and Scholarly Communication: a preliminary/exploratory study on Adoption, Practices, Drivers and Barriers”

B) The Abstract:

- “Our study is the first using empirical data to understand” - please change to: Our study uses empirical data… (See comments below highlighting other empirical studies on the topic).
- “The main concerns are related to the lack of quality assurance and the ‘Ingelfinger rule’. Please change to: main perceived concerns.

C) The Introduction:

- On Page 3: “key actors”. It is puzzling how the authors determined which the key stakeholders are in the preprints ecosystem. Authors should discuss why they did not incorporate “scientific journal publishers” as a relevant actor for this work. It is interesting that they incorporated “publishers” as a node in Figure 1 as a stakeholder group but failed to incorporate “representatives” during their interviews.
- On Page 3: “(Carà et al., 2017) .” Please delete the space.
- On Page 3: “The study is the first using empirical data”. Please see our comment on the abstract.
- On Page 3: “The study is the first using empirical data”. Please see the comments on the abstract above.
- On Page 4: “…relatively small number of peer-reviewed studies focusing on preprints – much of the literature is still to be found in editorials and opinion pieces”. The authors surely know that in general opinion pieces published in scientific journals are also peer-reviewed.
- “Defining preprints. Key components. Table 1”. A relevant aspect of preprints that is not contemplated here is that preprints are published open access free of charge. Perhaps that important feature could be incorporated in component 4. Maybe this feature, highly relevant for unprivileged research communities, was neglected given the sampling strategy employed (see comment below).
- On Page 8: “The study adopted a heterogeneous purposive sampling approach, aiming to include a wide range of perspectives from actors in the area. Participants comprised senior representatives…”. If the authors selected senior representatives of each stakeholder categories, one could question how heterogeneous the sampling approach was. Moreover, authors mention the use of snowball sampling, which is linked to homogeneous sampling. In addition, participants correspond to only eight OECD High-income economies according to World Bank (half of them only from UK and Germany), which again appear to be more reflective of a homogenous sampling approach. Authors should also highlight the intrinsic limitations of quality coding in terms of reproducibility.
On Page 8: “We undertook 38 semi-structured interviews”. Authors did not describe how sample size was determined. Authors should discuss how this small sample size might redound in uncertainty issues. The complete absence of statistical analyses of the data is striking. This affects the “empirical power” of the work, and one could question whether the described “trends” indicated in the manuscript are purely anecdotal, derived from sampling error and/or speculative.

“Like many kinds…” paragraph. This important paragraph highlighting the limitations of the study should be moved to the discussion/conclusion section.

D) The Results section in general:

The reviewers wonder if the results section as a whole, which is relatively long in comparison to the rest of the paper, should be limited to Table 5 and its description and avoid the unending narrative of anecdotal specific responses. In case the authors feel the need to present all these quoted responses, perhaps they could be moved to the supplementary section.

- “Some participants…” “Many saw a preprint as…” “Other participants saw preprints..” “there were signs this…” and elsewhere. This vague and unclear wording could be avoided by the use of absolute numbers or percentages of total responses.

- Table 5: Please consider complementing the legend to provide definitions to “systemic” and “individual”.

- Table 6: Following Table 5 title, it would be more reasonable to entitle Table 6 as “potential challenges of preprints”, or “possible” as depicted in the same table in Chiarelli et al. 2019.

- On Page 13: “One preprint service provider observed that only 10% of preprints received comments”. While this % is depicted as low in the text by the use of “only”, authors should contextualize this type of assertion. Peer reviewed articles published in journals are far less commented than preprints as reported elsewhere.

- On Page 14: “author might invite comments from people who are expected to be positive about their work”. Again, this and other speculative statements should be leveled when they are employed to contrast with traditional peer-reviewed journal articles. It is a common practice in an important number of journals to ask authors to provide contact information of potential reviewers, which could eventually lead to the same issues expressed in the preceding statement.

- On Page 14: “We note…”. This relevant paragraph, highlighting the limitations and anecdotal nature of the study, should be moved to the discussion/conclusions.

E) The Discussion/Conclusion: As expressed above, some parts of the results section could be moved here as they reflect more speculative comments based on the interview responses.

F) The References:


References

Publisher Full Text


Publisher Full Text


Publisher Full Text


Publisher Full Text


Publisher Full Text


**Is the work clearly and accurately presented and does it cite the current literature?**

Partly

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** 1) Sarvenaz Sarabipour, PhD: Systems Biology, Signal Transduction, Computational Modeling  2) Humberto Debat, PhD: Virology, Viromics
We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Comments on this article

Version 1

Reader Comment 08 Jul 2019

Donald Forsdyke, Queen's University, Kingston, Ontario, Canada

A few thoughts on a very interesting paper:

1. **Anachronistic language.** You state: “Of course, the ‘print’ part of ‘preprint’ is largely anachronistic, but like terms such as ‘paper’ and ‘manuscript’, has continued to be used even in a digital environment.” This is true. But the literal interpretation of “publish,” meaning to make publicly available, should not be considered in this category. Indeed, since some peer-reviewed published works are kept by commercial publishers behind paywalls, they are less-published (made public) than the preprints that are freely available on servers such as arXiv. From my readings of the Scholarly Kitchen Blog it seems that the publishing industry is trying to avoid the term “preprint publication.” I think authors should use the terms “preprint publication” and “formal publication,” rather than “preprints” and “publications.”

2. **Name guarantee.** Many “retired” professors (like me) already have established reputations in their fields and, as the shadows lengthen, may be quite happy to post their works on preprint servers and move on to the next project. I scan servers such as arXiv and bioRxiv daily and, as with my scans of regular journals, I used my knowledge of names in the field to help decide on the papers I will read. I also have a nose for a well-thought-out title. I am less dependent on the need for prior peer-review filtering as might be the case with new arrivals to the field.

3. **Citations.** In my field, preprint publications have DOI numbers and are citable. Preprint publications, which have never actually been peer-reviewed for formal publication, can be later cited among an author’s formal publications and then, perhaps, they find their way into Science Citation Indices. Organization such as rXivist ([https://rxivist.org/](https://rxivist.org/)) are already making it likely that a preprint publication is not a dungeon publication.

4. **Origin of Preprints.** I have added a commentary to Mathew Cobb’s article (2017) on early work in this field, which may be of interest. An interesting aspect is the idea of servers of preprint publications as scientific market-places.” I suspect I am already experiencing an early form of this as my junk-mailbox gets fuller daily with proposals from rogue publishers who may have seen my stuff on arXiv or bioRxiv. See: Cobb M. The prehistory of biology preprints: A forgotten experiment from the 1960s. PLoS Biol 15(11): e2003995.

Competing Interests: No competing interest.
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com