



This is a repository copy of *Accelerating scholarly communication : the transformative role of preprints.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/151295/>

Version: Published Version

Monograph:

Chiarelli, A., Johnson, R., Richens, E. et al. (1 more author) (2019) *Accelerating scholarly communication : the transformative role of preprints.* Report. Knowledge Exchange , London.

<https://doi.org/10.5281/ZENODO.3357727>

Reuse

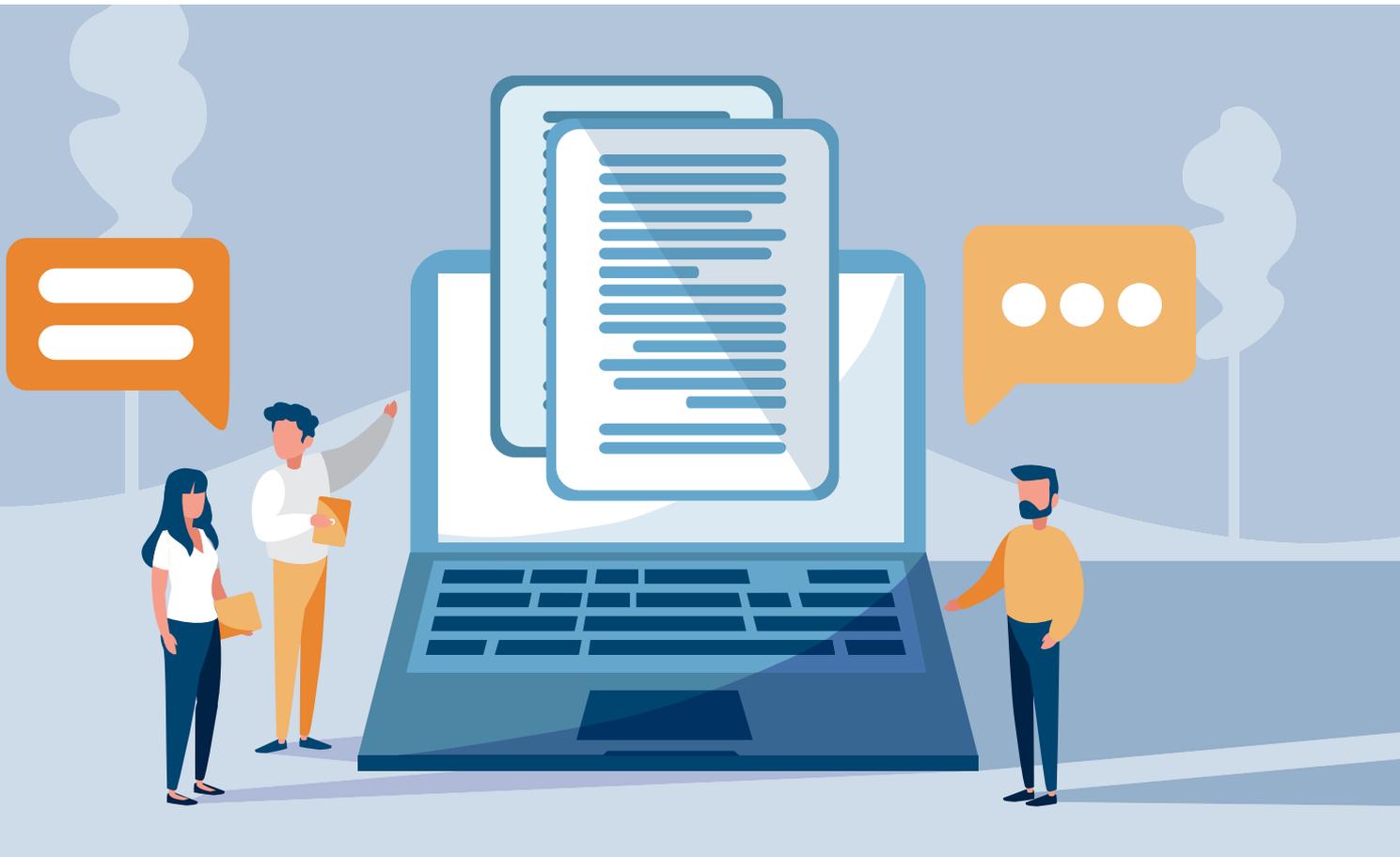
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



Accelerating scholarly communication

The transformative role of preprints

Published September 2019

© Knowledge Exchange 2019

Title: Accelerating scholarly communication:
The transformative role of preprints

Authored by: Andrea Chiarelli, Rob Johnson,
Emma Richens (Research Consulting), Stephen
Pinfield (University of Sheffield)

Email: andrea.chiarelli@research-consulting.com

DOI: [10.5281/zenodo.3357727](https://doi.org/10.5281/zenodo.3357727)

All content published can be shared (CC BY 4.0)
creativecommons.org/licenses/by/4.0



Acknowledgements

This study was underpinned by the work of the KE Task and Finish Group on preprints, led by Karin van Grieken (SURF), Juliane Kant (DFG) and Serge Bauin (CNRS), and including a range of other engaged members:

- ▶ Andy Turner (University of Leeds)
- ▶ Angela Holzer (DFG)
- ▶ Bas Cordewener (Knowledge Exchange)
- ▶ Birgit Schmidt (Göttingen State and University Library)
- ▶ Frank Manista (Jisc)
- ▶ Gernot Deinzer (Regensburg University)
- ▶ Jeroen Sondervan (Utrecht University Library)
- ▶ John Doove (SURF)
- ▶ Jon Tennant (IGDORE; Center for Research and Interdisciplinarity, University of Paris)
- ▶ Neil Jacobs (Jisc)
- ▶ Olivier Le Gall (INRA)
- ▶ Sarah James (Knowledge Exchange)
- ▶ Saskia Woutersen-Windhouwer (Leiden University/Netherlands Institute of Ecology (NIOO))

Our thanks also go to the 38 international stakeholders who agreed to be interviewed as part of this project and provided their views on preprints and the open sharing of research. A full list of project contributors is available in Appendix A.

Executive summary

Introduction

Background and rationale

The traditional academic publishing process is widely recognised as time-consuming for authors and reviewers and, in many cases, is slow to disseminate new knowledge. Over the past few years, the sharing of preprints, or versions of research outputs, ahead (or even instead) of formal publication has become more widespread in a number of academic disciplines. This study aims to advance Knowledge Exchange's (KE) previous work in the area of preprints, which consists of a **2018 review on this evolving landscape** (knowledge-exchange.info/event/preprints).

Objectives

The overall objective of this study was to explore the place of preprints in the research lifecycle from the points of view of researchers, research performing organisations, research funding organisations and preprint servers/service providers. Our investigation covered:

- ▶ Core benefits and usage in the case of researchers, including incentives and disincentives
- ▶ Attitudes of research performing organisations (RPOs) and research funders
- ▶ Values, strategies and aims of service providers

Methodology

This study was based on a comprehensive literature review and a set of 38 interviews that were transcribed and qualitatively coded for the purposes of thematic analysis. We mainly focused on research areas where preprint posting is growing (e.g. biology, chemistry and psychology, which were the focus of our interviews) but also considered disciplines where preprint posting is common (e.g. physics, mathematics, computer science, economics) or relatively less widespread (e.g. humanities).

The second wave of preprint servers

Preprints have become increasingly popular

Explosive growth has characterised the preprints landscape over the last few years. The first wave of preprint servers started with the high energy physics and economics communities in the 1990s, but, since 2010, the movement has been growing in popularity in other disciplines. Increasingly available and standardised technical solutions have enabled the launch of a range of disciplinary preprint servers focusing on the broad and early dissemination of research.

Preprints can support open scholarship

Preprints can support open scholarship by enabling free online access and potentially increasing the pace of research. They have a potentially transformative role in the scholarly communication landscape. Nevertheless, there remains significant uncertainty as to whether recent growth in interest in and take up of preprints services will be sustained, and how broadly preprints will be adopted across disciplinary communities.

The researcher's perspective

Disciplinary communities treat preprints differently

We found that ambiguity on the definition of a preprint is present across all disciplines and stakeholder groups: this means that any discussion of preprints is inherently complex and must be sensitive to context. The most common interpretations are that a preprint is either:

- ▶ A version of a paper ready to be submitted; or
- ▶ An early version of a paper shared to receive feedback before submission

However, a number of other views arose in this study. For example, the idea that preprints might be research outputs that are not intended as papers for peer review or that might not make it to the published stage (e.g. null results) was advanced.

Early and fast dissemination is a key motive for preprints posting

Early and fast dissemination appears to be the main motive behind preprint posting. In addition, increased opportunities for feedback seem to be highly valued, even though comments are not often added directly on preprint servers. Advantages for early career researchers are also often mentioned, as preprints can be added to CVs to increase the chances of being hired or promoted.

The lack of peer review and the fear of rejection by journals are barriers to uptake

Interviewees reported that their main concern when it comes to reading and reusing preprints is the fact that they haven't been peer-reviewed. This means that, potentially, incorrect findings could be shared broadly or reported on by the media. However, there is also an expectation that researchers and journalists will behave ethically and professionally, which should minimise the risk of the above.

Rejection by academic journals is another barrier to uptake, as some researchers fear that depositing a preprint might lead to editors not accepting their submissions on the grounds of the 'Ingelfinger rule'. This, however, appears to be only a perceived barrier, as many publishers now explicitly accept preprint posting.

A wide range of preprint servers are available today

In the course of this project, we identified over 60 platforms that can be used to store, share and, in some cases, comment on preprints. Today, the availability of a server that is fit for the purpose of any given researcher is almost guaranteed. It should be noted that preprint servers are often started from the bottom up and maintained by disciplinary communities, which indicates that they are likely to address any technical requirements or customs existing in a research field.

Twitter has been playing a key enabling role

Researchers and preprint servers often rely on Twitter for preprint discovery and sharing purposes. Researchers can follow Twitter bots posting preprints as set up by individual preprint servers but also share their own preprints. Twitter is, in practice, how many researchers appear to encounter preprints for the first time and is one of the key pathways for making and receiving comments.

Mapping the preprints landscape

A wide range of stakeholders are involved in and affected by preprints

The preprints landscape is currently characterised by some degree of fragmentation, which suggests that future developments are likely to benefit from closer collaboration between the stakeholder groups involved. These include researchers, research performing organisations, research funders, service providers and publishers. We note that, at present, there is significant experimentation in terms of approaches and technologies, and that the extent to which stakeholders are collaborating is unclear in some cases.

Different practical approaches to preprint servers are being taken

A range of technical solutions are available to implement preprint servers in practice, including the popular Open Science Framework and digital repository solutions (e.g. EPrints, figshare, DSpace, Invenio, Drupal); ad-hoc and proprietary infrastructure is also widespread. The choice of solution has little impact on the openness of preprints deposited but does affect user experience, the level of control that the owners and managers of preprint servers can exert over their platforms, and the effort required to do so.

Preprints are poorly integrated into publication workflows

Current technologies seem largely suitable to support the uptake of preprints. For instance, digital object identifiers or permalinks can be assigned to preprints, withdrawals are possible on preprint servers and open licensing options are offered. However, versioning features are not used by many authors and the automatic tracking of a manuscript through the publication process is difficult. In most cases, preprint posting is disconnected from traditional publication workflows: this means that researchers would typically post a preprint independently ahead of publication and then add new versions after making revisions.

'Information overlap' and digital preservation are growing concerns

In some cases, preprints are posted as the author's accepted manuscript on both a preprint server and an online repository (e.g. an institutional repository). We call this phenomenon 'information overlap' and raise the challenge of covering the costs of technical infrastructure in cases where this duplicates efforts that are already otherwise funded. Currently, there is a lack of consistency in terms of approaches to the long-term preservation of preprints, and this is not seen as a priority due to limited budgets and the quick pace of change within the preprints landscape. However, long-term preservation is recognised as a growing concern that should be addressed in the future, including in terms of what preprints should or should not be within the scope of digital preservation activities.

Are preprints riding the hype wave?

Preprints and preprint servers have been growing in popularity very quickly over the past few years. Based on a hype cycle interpretation, the visibility of preprints can be expected to decrease from the current "peak of inflated expectations", and we note that some players in this landscape might merge or disappear in time. After a "trough of disillusionment", preprints and preprint servers might once again grow in visibility and reach the level of

mainstream adoption that is currently seen in the communities served by arXiv and RePEc (physics, mathematics, computer science, economics, among others).

The future of preprints

It is not clear who will take the lead in preprint posting

We investigated the question of whether preprint posting will evolve as a researcher- or publisher-centric phenomenon. The answer is not clear at present, but we note that researchers are mostly responsible for posting preprints today. A shift to a publisher-centric model could potentially improve the tracking of preprints throughout and after publication, but there are growing concerns of market consolidation in the scholarly communication landscape. Clearly, the choice between a researcher- or publisher-centric approach will affect funding, too: in the former case, grants or pooled funds would likely form the bulk of funding for preprint servers, while in the latter these could be supported by publishers, provided they perceive sufficient potential for a return on the investment required.

Do traditional journals need to evolve?

In a shifting landscape that could be transformed by the increased use of preprints, the role of and costs attached to traditional academic journals is liable to be questioned. Furthermore, overlay journals reviewing and sharing content posted to preprint servers are already being used in some disciplinary communities, and this could also affect the extent to which traditional journals might have to reframe their value proposition.

Licensing options should be carefully considered

Licensing is recognised as a challenge when it comes to preprints, as the promise of broader reuse of research outputs is underpinned by permissive licence terms. Several study participants were not able to fully justify their choice of licence for their own preprints, which suggests that this should be a key area of focus for preprint servers in the future.

Preprints can support fairer research(er) evaluation

A growing number of research funders are starting to acknowledge and accept preprints as suitable for inclusion in grant applications, and we recognise the role that preprints can have in researcher evaluation. There is an increasing push to focus on individual outputs rather than on publication venues such as high-impact journals.

Preprint servers should aim to address perceived pain points

Preprint servers today are being started by enthusiastic proponents of open scholarship but may not always meet a perceived need in their research communities. The increasing focus on open scholarship in the research landscape is certainly contributing to some extent of behavioural change, but preprint servers might have to focus more on addressing researchers' pain points if they are to lead to lasting change.

Conclusions

Three future scenarios in preprint posting

We see three possible scenarios for the future of preprints:

- ▶ **Scenario 1 – Turn of the tide:** the second wave of preprint servers fades, and preprints remain a major component of scholarly communication only in the fields where they are already firmly established, e.g. those served by arXiv and RePEC
- ▶ **Scenario 2 – Variable adoption:** preprints grow in some additional fields such as those within the scope of ChemRxiv and bioRxiv, but not all
- ▶ **Scenario 3 – Preprints by default:** preprints grow in all fields (at different paces) and are accepted by the research community at large

Scenario 1 is expected to materialise if current efforts to promote preprints fail. Scenario 2 is likely to be the case in the short-to-medium term, but it might be a transition between the other two scenarios, or alternatively, an endpoint if further developments fail to materialise. Scenario 3 can only happen if all stakeholders involved cooperate to turn the promise of preprints into reality and is likely to be an option only in the long term. Even then, it may be that certain disciplinary areas, such as the Humanities, do not adopt preprints at any scale.

Five areas should be considered to ensure a sustainable future for preprints

This study led to the identification of five areas that require further investigation:

1. Responsibilities and business models
2. Involvement of commercial players vs community ownership
3. Evidence on the advantages and disadvantages of preprint posting
4. Pathways to awareness raising
5. Approaches to training and support

We note that active engagement is needed to build a sustainable future for this growing scholarly communication practice: the higher the level of stakeholder coordination, the more positive any outcomes will be for the research community.

Five take-away messages



Early and fast dissemination, increased opportunities for feedback and openness are seen as the main benefits of preprints.



The main concerns over preprints are the lack of quality assurance, media potentially reporting inaccurate research and journals rejecting articles if a preprint has been posted.



Twitter has been playing a key enabling role in the current second wave of preprints and preprint servers. It also appears to be the main way researchers are exposed to preprints in the first place.



It is not clear who will be responsible for posting preprints in the long-term – researchers or publishers? This will partly be affected by the availability of sustainable business models.



Traditional academic journals might have to reframe their value proposition should preprints grow significantly in popularity in the future.



Contents

Executive summary	3
Introduction	3
The second wave of preprint servers	3
The researcher's perspective	3
Mapping the preprints landscape	4
The future of preprints	5
Conclusions	6
Five take-away messages	7

1. Introduction	10
Background and rationale	10
Objectives	10
Methodology	11
Structure of the report	11
Limitations	11

2. The second wave of preprint servers	12
Recent growth	12
A brief history of preprints	13
Preprints and open scholarship	14

3. The researcher's perspective	16
The challenge of defining preprints	16
Benefits for the research community	17
Early adopters and sceptics	19
Posting preprints in practice	21
The enabling role of Twitter	21

4. Mapping the preprints landscape	23
Key stakeholders	23
Technology and operations	24
Size and trends in the preprints landscape	25
Infrastructural challenges	26
Information overload and information overlap	26
Riding the hype wave?	27

5. The future of preprints	30
Taking the lead in preprint posting	30
Cultural resistance	31
The evolving roles of academic journals, preprint servers and overlay services	31
Licensing and preprints	32
Preprints, responsible metrics and evaluation	34
Addressing perceived pain points	35
Supporting the preprints movement	35
6. Conclusions	37
Summary and conclusions: Three future scenarios	37
Making progress	40
7. References	42
Knowledge Exchange study outputs	42
Other references	42
Appendix A – Acknowledgements and participants	48
Appendix B – Interview questions	51
Appendix C – Overview of preprint servers	54

1. Introduction

Knowledge Exchange, a group of national organisations from six European countries, commissioned and co-designed this study as part of their work on digital infrastructures to enable open scholarship. This report investigates the preprints landscape: it highlights current thinking in this dynamic area and makes recommendations for future work.

Background and rationale

This study arose from the need to better understand the preprints landscape. Preprints are versions of research papers, typically prior to peer review and publication in a journal. The practice of sharing these research outputs online has increased rapidly in popularity over the past few years, partly in response to the slow pace of traditional academic workflows, from article submission to publication.¹ Knowledge Exchange (KE) has been working on the topic of preprints since 2018² and commissioned this study to investigate the current state and broader implications of this evolving area.

This work started in September 2018 and led to the development of:

- ▶ A slide deck summarising initial findings³
- ▶ A publication under open peer-review available on F1000Research⁴
- ▶ The present report

Objectives

The overall objective of this study was to explore the place of preprints in the current research lifecycle from the points of view of researchers, research performing organisations, research funding organisations and preprint servers/service providers. Particularly, we set out to investigate:

- ▶ Core benefits and usage in the case of researchers, including incentives and disincentives
- ▶ Attitudes of research performing organisations (RPOs) and research funders
- ▶ Values, strategies and aims of service providers

Footnotes

- 1 AMS Secretary. (2018). Backlog of Mathematics Research Journals. ams.org/journals/notices/201810/rmoti-p1289.pdf
 - 2 Tennant, J., Bauin, S., James, S., & Kant, J. (2018). The evolving preprint landscape: Introductory report for the Knowledge Exchange working group on preprints. <https://doi.org/10.31222/osf.io/796tu>
 - 3 Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Practices, drivers and impediments in the use of preprints (Phase 1 report). [https://repository.jisc.ac.uk/7381/1/Practices,_drivers_and_impediments_in_the_use_of_preprints_\(Phase_1_report\).pdf](https://repository.jisc.ac.uk/7381/1/Practices,_drivers_and_impediments_in_the_use_of_preprints_(Phase_1_report).pdf)
 - 4 Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Preprints and Scholarly Communication: Adoption, Practices, Drivers and Barriers. <https://doi.org/10.12688/f1000research.19619.1>
-

The present report builds on our initial findings shared in slide form and aims to highlight key messages and areas for future work for all stakeholders identified as having a role in the preprint landscape.

Methodology

This study was carried out by reviewing relevant literature on the topic of preprints and by interviewing a range of 38 international stakeholders in the preprints landscape. Interviews were recorded, transcribed and qualitatively coded for the purposes of analysis and reporting. Quotes in this report are included using an intelligent verbatim approach to transcription (i.e. any fillers and repetitions in the text have been removed for improved legibility).

Our research focused on disciplines where the use of preprints is increasing quickly: these included biology, chemistry and psychology, with the corresponding preprint servers bioRxiv, ChemRxiv and PsyArXiv. We did, however, broaden the scope of our analysis for the purposes of this report, including other preprint servers and service providers. This study used Innovation Diffusion Theory⁵ as an evaluation framework.

Structure of the report

This report includes both findings from a literature review (LR) and original empirical research (OER). After this introduction, it is structured as follows:

- ▶ The second wave of preprint servers (LR)
- ▶ The researcher's perspective (OER)
- ▶ Mapping the preprints landscape (LR)
- ▶ The future of preprints (OER)
- ▶ Conclusions (OER)

In some parts, literature findings and original research might be presented side-by-side, but the above split describes the main focus of each section.

We note that researchers are the only stakeholder group to whom we have dedicated an entire section. This is because preprint posting will likely struggle to play a role in scholarly communication unless it is closely aligned with researchers' motivations to share, read and cite scholarly content.

Limitations

Study participants were gathered via convenience sampling, that is, we interviewed stakeholders who were both available and willing to participate. Moreover, interviews were limited to individuals based in Europe and North America. Therefore, it may not be appropriate to generalise the findings of this study, and outlying results may be over-represented. Furthermore, we note that:

- ▶ We chose not to interview traditional academic publishers, as the publishing community is already discussing preprints in a structured way, for example via the Committee on Publication Ethics (COPE).⁶ Their role and importance, however, are clearly acknowledged and discussed throughout the report
- ▶ Our analysis is underpinned by qualitative coding, and we note that this relies on analytical judgement and interpretation

Footnotes

- 5 Rogers, E. M. (2003). Diffusion of innovations.
https://books.google.co.uk/books?id=9U1K5LjUOwEC&redir_esc=y
 - 6 COPE. (2018). COPE Discussion Document – Preprints.
https://publicationethics.org/files/u7140/COPE_Preprints_Mar18.pdf
-

2. The second wave of preprint servers

Preprint servers have been available since the early 1990s for the physics, mathematics and economics communities but have started growing more widely only over the past few years. Preprints originated in online form as a practical solution to the issue of sharing and reading hardcopy research prior to formal publication. Preprints form part of the open scholarship landscape and exist in parallel to traditional journal articles – the key difference is that preprints may not be submitted for peer-review.

Recent growth

The growth of preprint servers over the last few years has been nothing short of explosive. **Figure 1** (p.13) builds on previous work to map the preprints landscape⁷ and shows that preprint servers started appearing in the 1990s (see **Appendix C**, p.54). The movement slowed down to some extent between the late 1990s and 2010 but has seen a resurgence over the last ten years and particularly the last five.

Preprint servers have been created to share preprints, that is, versions of research outputs typically prior to peer review and publication.² As outlined in **section 3** (p.16), defining preprints is not simple, as disciplinary communities and norms play a significant role in determining what a preprint is and what it is worth to researchers; however, for the purposes of this overview, the above definition will suffice.

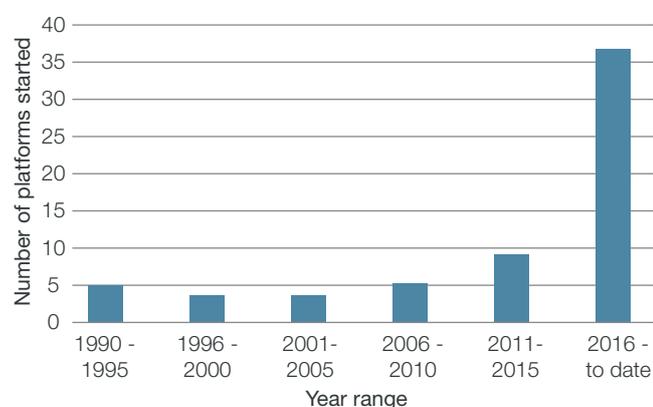
In terms of preprints making it to peer-reviewed form, we note that as many as 59% of preprints posted to ChemRxiv (chemistry) and 67% of those posted to bioRxiv (biology) are eventually peer-reviewed and published formally;⁸ in the case of preprints posted to bioRxiv, recent research shows that “the majority of published preprints appeared in a journal less than six months after being posted”.⁹ In the communities served by arXiv, the number of preprints making it to peer-reviewed form has been estimated at over 90%,¹⁰ which

suggests that there is further potential for growth in other disciplines. However, numbers of preprints remain relatively low when compared with traditional academic publishing. For example, in the field of biology, the number of preprints posted in 2019 relative to new publications in PubMed stands at just 2.3%.¹¹

Footnotes

- 2 Tennant, J., Bauin, S., James, S., & Kant, J. (2018). The evolving preprint landscape: Introductory report for the Knowledge Exchange working group on preprints. <https://doi.org/10.31222/osf.io/796tu>
 - 7 Rittman, M. (2017). Research Preprints – Preprint servers. <https://researchpreprints.com/preprintlist>
 - 8 Nguyen, T.M. (2019). Chemistry preprints pick up steam. <https://cen.acs.org/acs-news/publishing/Chemistry-preprints-pick-steam/97/i3>
 - 9 Abdill, R.J. & Blekhman, R. (2019). Meta-Research: Tracking the popularity and outcomes of all bioRxiv preprints. <https://doi.org/10.7554/eLife.45133>
 - 10 Gentil-Beccot, A., Mele, S. & Brooks, T. (2009). Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories. <https://doi.org/10.1007/s11192-009-0111-1>
 - 11 Penfold, N. (2019). Twitter (29 May 2019). <https://twitter.com/npscience/status/1133734784939769856>
-

Figure 1 – Growth of platforms allowing the sharing of preprints in time



Note: Figure 1 was created based on information available online and only includes currently active platforms that could be identified within the timeframe and scope of this project. While the information is considered to be correct at the date of publication, we cannot guarantee its accuracy.

A brief history of preprints

Some disciplinary communities started seeking the open sharing of pre-refereed research long before the advent of today's digital tools and the open scholarship movement.

- ▶ The high energy physics (HEP) community started sharing hardcopy literature prior to publication by post in the 1960s, once the process to duplicate articles had become economic¹²
- ▶ The same applied to the economics community, which was sharing working papers – the designation for pre-refereed work in economics – in the 1950s
- ▶ The US National Institutes for Health launched the Information Exchange Groups (IEG) in the 1960s, aiming to share any biology “preprint, comment, discussion” by post¹³

Early preprint sharing took place via personal correspondence but the role of physical repositories (e.g. libraries) became more important in time. One of the first issues that sharing hardcopy preprints led to was information overload.¹¹ The increasing number of hardcopy preprints was making research more difficult, as sifting through thousands of articles was impractical for any individual researcher (we note that similar issues would likely have applied to peer-reviewed hardcopy work, too).

Digital systems to manage bibliographic records were an initial fix to the unmanageable number of preprints available. However, it wasn't until the advent of digital typesetting systems such as **TeX** (<https://en.wikipedia.org/wiki/TeX>) that things really changed. Digital typesetting systems allowed authors to write research articles in electronic form using plain text, to be rendered directly on the reader's device. The next roadblock the preprints movement hit was that mailboxes were getting full too quickly, even though plain text articles were small in terms of their file size.

Footnotes

- 12 O'Connell, H. (2000). Physicists Thriving with Paperless Publishing. <https://arxiv.org/abs/physics/0007040v2>
- 13 Cobb, M. (2017). The prehistory of biology preprints: A forgotten experiment from the 1960s. <https://doi.org/10.1371/journal.pbio.2003995>

The physics and economics communities developed different yet complementary solutions to share electronic preprints:

- ▶ In 1991, physicist Paul Ginsparg created a central repository at Los Alamos National Laboratory, from which preprints could be obtained directly. The creation of this repository marked the birth of e-prints – electronic preprints – as opposed to their hardcopy versions. In 2001, Ginsparg moved to Cornell University: this year marked the change of the name of the above central repository to today’s arXiv¹⁴
- ▶ In 1993, the Working Papers in Economics project (WoPEc) was started to enable electronic dissemination of economics working papers. WoPEc grew into “an interconnected network of over 60 archives holding over 13,000 downloadable papers and over 50,000 descriptions of offline papers from close to 1,000 series, as well as data about over 4,000 academic Economics departments and research institutes”. This is now called Research Papers in Economics, or RePEc¹⁵

Looking back at these disciplinary communities, it is easy to see that the desire to exchange research openly and prior to formal publication (a process often spanning several months)¹⁶ was the key motivation behind the creation of preprint servers. Since preprint sharing originated in hardcopy form, digital tools offered a chance to rationalise, simplify and broaden access to a system that was (to some extent) already in place. Today’s preprint servers (see **Section 4, p.23**) appear to be following a similar policy. However, while the desire for sharing work in preprint form has been historically high in the physics and economics communities, this may not be the case for all disciplines.

Preprints and open scholarship

The first preprint servers were created to facilitate the open sharing of research prior to formal publication. However, the idea of openness today is immediately associated with the concepts of open scholarship and open access. While a detailed analysis of the topic is beyond the scope of this report, we highlight the following aspects of the relationship between preprints and open scholarship:

- ▶ Preprints can support open scholarship by enabling free online access and potentially increasing the pace of research^{10, 17}
- ▶ While the above benefits with respect to open scholarship are recognised, preprints are typically shared in pre-refereed form (in the first place) and this might affect the extent to which researchers are willing to consider and use them as they would traditional publications¹⁸

Footnotes

- 10 Gentil-Beccot, A., Mele, S. & Brooks, T. (2009). Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories.
<https://doi.org/10.1007/s11192-009-0111-1>
 - 14 Butler, D. (2001). Los Alamos loses physics archive as preprint pioneer heads east.
<https://doi.org/10.1038/2F35083708>
 - 15 Karlsson, S. & Krichel, T. (1999). RePEc and S-WoPEc: Internet access to electronic preprints in Economics.
<http://openlib.org/home/krichel/papers/lindi.html>
 - 16 Huisman, J. & Smits, J. (2017). Duration and quality of the peer review process: the author’s perspective.
<https://doi.org/10.1007/s11192-017-2310-5>
 - 17 Fraser, N., Momeni, F., Mayr, P. & Peters, I. (2019). The effect of bioRxiv preprints on citations and altmetrics.
<https://doi.org/10.1101/673665>
 - 18 Neylon, C., Pattinson, D., Bilder, G. & Lin J. (2017). On the origin of nonequivalent states: How we can talk about preprints.
<https://doi.org/10.12688/f1000research.11408.1>
-

Furthermore, the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, although dated, recognises the role of the internet as an emerging medium for knowledge dissemination and that this will “significantly modify the nature of scientific publishing as well as the existing system of quality assurance”.¹⁹ This reinforces the view that preprints and preprint servers can play a transformative role in scholarly communication workflows.

“The preprint agenda is a reaction against the very expensive Gold open access that is required by some funders. It may appeal to those who lack the funding for Gold open access.”

Researcher

The debate on the transition to open access has increased in intensity significantly in the last year, following the announcement of 'Plan S' by a group of European Funders (September 2018). This initiative stipulates that scientific publications that result from research funded by public grants must be published in compliant Open Access journals or platforms, including repositories. The implications of Plan S (which is rapidly evolving) have been widely discussed and debated, but we note here that the implementation guidance states that the “early sharing of research results through preprints is [...] strongly encouraged”.²⁰ Preprint posting, however, is not seen as meeting the proposed open access requirements, which apply to peer reviewed scholarly articles. This, to some extent, further confirms the distinction made in the bullets above.

The co-founders of bioRxiv and the Editor in Chief of eLife have jointly proposed 'Plan U',²¹ which recommends preprint posting should be a funder requirement to achieve free access to research. Plan U is based on the expectation that most preprints would subsequently be peer reviewed and puts much of the

technical burden on preprint servers, such as permanence of deposition (e.g. articles can be withdrawn but a record would remain), indexable and standardised metadata, linking with relevant publishing infrastructure and long-term preservation.

Other approaches combining open scholarship, open access and the use of preprints have been hypothesised. As an example, Green has recently proposed a transformation of scholarly publishing involving preprints: in this scenario, articles would first be posted as preprints and invited to peer review only if they are attracting sufficient attention. Therefore, peer review and publication would only be carried out in select cases.²² This ties back to the idea that the internet could reshape scholarly communication and systems of quality assurance advanced in the Berlin Declaration on Open Access.

Footnotes

- ¹⁹ Max Planck Society. (2003). Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. <https://openaccess.mpg.de/Berlin-Declaration>
- ²⁰ cOAlition S. (2019). Principles and Implementation. coalition-s.org/principles-and-implementation
- ²¹ Sever, R., Eisein, M. & Inglis, J. (2019). Plan U: Universal access to scientific and medical research via funder preprint mandates. planu.org
- ²² Green, T. (2019). Is open access affordable? Why current models do not work and why we need internet-era transformation of scholarly communications. <https://doi.org/10.1002/leap.1219>

3. The researcher's perspective

Defining a “preprint” is a challenge, as disciplinary communities assign different values to research outputs in preprint form. Furthermore, the research community is split between early adopters, who see many advantages in preprint posting, and sceptics, who struggle to see benefits and highlight concerns such as the lack of peer review. The role of Twitter is important, as researchers are often exposed to preprints via social media and discover new ones by following peers and the accounts of prominent preprint servers.

The challenge of defining preprints

Recent work on the future of scholarly communication shows that a mix of researchers and funders are keen to see new output formats and feedback loops, increased transparency and reproducibility, more pathways to research impact and, importantly, faster research.²³ At least in principle, what is broadly understood to be a preprint would fit the bill. However, our research shows that there is no agreement across research communities on precisely what a preprint is:

- ▶ Most see a preprint as a version of a paper ready to be submitted or as an early draft uploaded online (“posted”, in preprint jargon) to receive comments from the community
- ▶ Some see preprints as research outputs that haven't been completed as papers for peer review (e.g. work that is not meant for formal publication in the first place) or that might not make it to the published stage (e.g. null results)
- ▶ A minority see preprints as the author's accepted manuscript (AAM) posted on a preprint server, possibly to comply with national/funder policies or for personal preference, but not yet formatted into the version of record (VoR) published by the journal

The first of these three definitions is likely the most canonical view of preprints with respect to traditional publishing workflows.²⁴ However, we note that different disciplinary communities have slightly different interpretations of the term “preprint” and that a unified view may not be possible (nor is it clear at this stage whether this would be desirable).²⁵

The standing of preprints, i.e. their value or reputation, is also defined by disciplinary communities. This means that, in addition to the difficulty of understanding what a preprint is in different disciplines, the weight it will carry for them will also vary in practice. For example, preprints are highly regarded by the communities served by arXiv (e.g. physics, mathematics and computer science), while they are considered mostly as works in progress by those using SSRN (e.g. social sciences and humanities, which constitute the bulk of the content on this platform).²⁴

Footnotes

- 23** Elsevier, Ipsos MORI. (2019). Research futures: Drivers and scenarios for the next decade. elsevier.com/connect/elsevier-research-futures-report
- 24** Neylon, C., Pattinson, D., Bilder, G. & Lin J. (2017). On the origin of nonequivalent states: How we can talk about preprints. <https://doi.org/10.12688/f1000research.11408.1>
- 25** European Commission. (2019). Future of Scholarly Publishing and Scholarly Communication. <http://doi.org/10.2777/836532>
-

The standing of preprints can also be discussed with respect to national assessment exercises, but we note that this is an evolving area and should be considered in the broader context of research evaluation and metrics. As an example, the UK's 2021 Research Excellence Framework does consider preprints as valid research outputs but not as equivalent to articles.²⁶ On the other hand, the OA policy used by the Excellence in Research for Australia programme clearly states that any versions of articles that have not been refereed (including preprints) are not acceptable.²⁷

“The term “preprint” itself includes the idea that you’re building it towards something. That it’s only the preprint and then something will come later from it.”

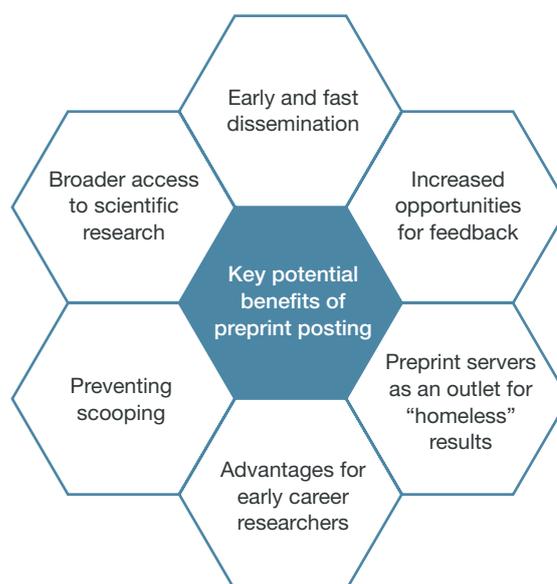
Research funder

Finally, we note the importance of distinguishing preprints from preprint servers. The former term describes a type of research output, which could be in hardcopy or electronic, early-stage or ready for submission. On the other hand, preprint servers are the technical infrastructure underpinning the use of preprints in electronic form. This indicates that preprints could exist even without dedicated preprint servers. For example, the community could use generalist repositories (e.g. Zenodo, figshare, Open Science Framework) or platforms such as ResearchGate to deposit pre-refereed research. Today's research, including the present report, tends to deal with both preprints and dedicated preprint servers, but we note that this close relationship might not be the case forever.

Benefits for the research community

Our research showed that the key perceived benefits of preprint posting for researchers are early and fast dissemination and increased opportunities for feedback (Figure 2). Broader access to scientific research is also significant and related to the above, as preprints are normally expected to be openly accessible online.

Figure 2 – Potential benefits arising from preprint posting



Footnotes

²⁶ Hill, S. (2018). Twitter (30 May 2018).

<https://twitter.com/stevenhill/status/1001897100567891971>

²⁷ ARC. (2017). Open Access Policy.

arc.gov.au/policies-strategies/policy/arc-open-access-policy-version-20171

Some of the broadly advertised benefits of preprint posting are sometimes perceived by researchers as potential risks, and finding the right balance is key.²⁸ Our interviewees highlighted the following:

- ▶ The idea that preprints prevent scooping by establishing priority is shared by many, but some researchers are concerned about the opposite, i.e. research being scooped because it has been made available early on in a potentially unfinished form
- ▶ The possible advantages for early career researchers of quickly building up a track record of publication through preprints appear to be broadly supported by our interviews and the literature. However, the extent to which these advantages will materialise largely depends on whether research funders and RPOs value preprints in practice
- ▶ Some have advanced the idea of using preprint servers as outlets for “homeless” results, i.e. outputs that currently do not have dedicated publication venues (e.g. **null results** https://en.wikipedia.org/wiki/Null_result). Using preprint servers for this purpose might raise some concerns in terms of their scope and role in scholarly communication, i.e. what are preprint servers really for? There doesn't seem to be agreement on this, and there is a risk of creating ambiguity in terms of workflows, DOI creation, indexing, and licensing if so-called “homeless” research is shared in parallel with regular articles in preprint form

“The primary purpose of preprints is to communicate scientific knowledge as early as possible to as wide an audience as possible.”

Researcher

It is currently difficult to quantify the advantages of preprint posting in fields where the preprints culture is just growing, and these remain largely anecdotal. While most advantages are broadly mentioned in the international literature^{4,10,29,30} and by those aware of preprints as a phenomenon, evidence is scarce. We note however that, in the case of economics and RePEc, when a working paper and a journal version are both available, “the working paper is downloaded many times more than the article”.³¹ This could have implications in terms of researcher evaluation, as the preprint server may include important metrics to complement those attached to the journal version, but also means that academic journals might have to think about their value proposition in light of the success of preprints in some disciplines.

Footnotes

- 4 Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Preprints and Scholarly Communication: Adoption, Practices, Drivers and Barriers. <https://doi.org/10.12688/f1000research.19619.1>
- 10 Gentil-Beccot, A., Mele, S. & Brooks, T. (2009). Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories. <https://doi.org/10.1007/s11192-009-0111-1>
- 28 Sarabipour, S., Debat, H.J., Emmott, E., Burgess, S., Schwessinger, B. & Hensel, Z. (2018). On the value of preprints: an early career researcher perspective. <https://doi.org/10.7287/peerj.preprints.27400v1>
- 29 Bourne, P.E., Polka, J.K., Vale, R.D. & Kiley R. (2017). Ten simple rules to consider regarding preprint submission. <https://doi.org/10.1371/journal.pcbi.1005473>
- 30 Alliance nationale de recherche pour l'environnement. (2017). Preprints are a valid form of scientific communication. allenvi.fr/actualites/2017/preprints-communication-scientifique-recevable
- 31 Zimmerman, C. (2019). 5000 working paper series on RePEc: working papers are still central to economics. <https://blog.repec.org/2019/05/31/5000-working-paper-series-on-repec-working-papers-are-still-central-to-economics>

The rationale for preprint posting may be clear in a minority of research fields. However, in most others, the various stakeholders involved would need a nudge to take up a new practice and fit it within their already busy schedules. The extent of cultural change that would be required is noted in the literature.³² The need for a 'carrot' element in the preprints equation mostly refers to researchers as the primary writers of research outputs. However, funders (who may have to assess preprints), RPOs (who might need to provide support), publishers (who would need to take a position on whether they accept the practice) and more will need to contribute to ensure preprints find their place in the scholarly communication landscape.

"I don't have a lot of examples [of the benefits of preprints] here, but certainly, you know, I hear anecdotes."

Researcher

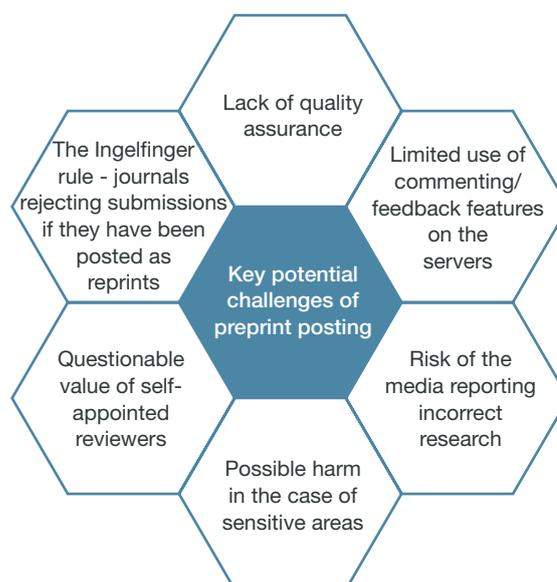
Early adopters and sceptics

Our interviews highlighted the presence of two distinct attitudes towards the practice of posting preprints:

- ▶ Some researchers are supporters of preprint posting and tend to be early adopters and follow emerging community trends. They are committed to open scholarship and embrace new practices believing that their benefits (**Figure 2**, p.17) will materialise
- ▶ Others are more sceptical and feel that preprints might be just 'yet another job'. In their view, the challenges of posting preprints (**Figure 3**) outweigh the benefits

Naturally, every innovation will see some contrast between early adopters and sceptics. We have discussed the benefits perceived by the supporters of preprints in the previous section, so we will now focus on the point of view of sceptics. Similarly to the case of the benefits listed above, which tend to be difficult to evidence, sceptics often express hypothetical objections.

Figure 3 – Potential challenges arising from preprint posting.



Footnotes

- ³² Weingart, P. & Taubert, N. (2017). The Future of Scholarly Publishing: Open Access and the Economics of Digitisation. <http://doi.org/10.5281/zenodo.1003185>

We note that many challenges mentioned by study participants and in the literature with respect to preprints and preprint servers would not materialise unless authors or media reporters engaged in unprofessional behaviours:

- ▶ The fact that preprints are unrefereed may be an issue. However, if they are treated as pre-review outputs and critically assessed by readers and re-users, no particular issues should arise. In addition, since reputation plays an important role in academia, the likelihood of authors risking posting poor or falsified results may reasonably assumed to be low
- ▶ Media may pick up research in preprint form and report on it. However, the consensus is that the onus is on journalists and researchers to behave responsibly and professionally, acting where appropriate to corroborate the findings of a preprint before this is shared via mass media or to report it with caveats³³
- ▶ Harm in sensitive areas, for instance, those related to human health or protected animal species, is not likely to materialise if researchers behave ethically and in line with professional standards in their fields. Preprint servers dealing with these topics, such as medRxiv, include screening processes to ensure that research with a potentially harmful impact is not posted³⁴

“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.”

Researcher

The possibility of rejection by academic journals when a preprint has been posted appears to be one of the root causes of scepticism. We note that the above-mentioned Information Exchange Groups (IEG) started in the 1960s by the NIH fell due to pressures from journals and learned societies: these “considered the organised

circulation of preprints in both biology and physics to be a threat to their financial interests and to their perceived status as guardians of scientific integrity”.¹³

Although the scholarly communication landscape has since changed, we highlight that scholarly publishers and learned societies do retain the ability to affect norms and behaviours in disciplinary communities. For instance, the so-called “Ingelfinger rule”,³⁵ which arose in 1969 and after the fall of the IEG, stipulates that a journal would not publish findings already shared elsewhere. This concept is widely understood by researchers today, but fears that it might apply widely to preprints may not be in line with reality: article rejection on the basis that a preprint has been posted seems unlikely, as a range of journals now have policies compatible with **preprint posting** (https://en.wikipedia.org/wiki/List_of_academic_journals_by_preprint_policy).³⁶ In particular, we note that Elsevier, Springer Nature and Wiley, which together published over 30% of the global article output in 2016,³⁷ currently accept preprint posting prior to submission to their journals. Furthermore, in some fields, the sharing of preprints is already accepted broadly: in palaeontology,

Footnotes

- 13 O'Connell, H. (2000). Physicists Thriving with Paperless Publishing. <https://arxiv.org/abs/physics/0007040v2>
- 33 Tennant, J., Gatto, L. & Logan, C. (2018). Preprints help journalism, not hinder it. <http://doi.org/10.1038/d41586-018-06055-3>
- 34 medRxiv. (2019). Coming soon: medRxiv. <https://connect.medrxiv.org>
- 35 Altman, L.K. (1996). The Ingelfinger rule, embargoes, and journal peer review - part 1. [https://doi.org/10.1016/S0140-6736\(96\)91016-8](https://doi.org/10.1016/S0140-6736(96)91016-8)
- 36 Teixeira da Silva, J.A. & Dobránszki, J. (2019). Preprint policies among 14 academic publishers. <https://doi.org/10.1016/j.acalib.2019.02.009>
- 37 Efficiency and Standards for Article Charges – ESAC. (n.d.). Market watch. <https://esac-initiative.org/about/apcmarket>

for example, this is the case for around 60% of journals.³⁸ A possible way to address uncertainty around journal policies is the use of sources such as **SHERPA RoMEO** (<http://sherpa.ac.uk/romeo/index.php>) or the recently-created **Transpose database** (<https://transpose-publishing.github.io/#/about>): an international collaboration, the website lists a large number of journals and includes information on their policies, including with respect to preprints.

The academic literature also mentions some additional issues that weren't discussed in our interviews due to our focus on certain preprint servers and on the pre-review stage. For example, authors who have published on F1000Research (an open access publishing platform) have reported that open peer review of preprints on the platform could lead to poorer-quality reviews lacking criticism.^{39,40} The opposite has also been argued as well – that introducing transparency in peer review ensures reviewers produce higher quality reports. The literature does mention that reviewers might be unwilling to get involved in the first place or to be too critical if their comments are going to be published publicly,⁴¹ but the reduced amount of criticism could also reflect the above-mentioned fact that researchers tend to submit preprints of a high standard when they know these will be immediately visible online.

The underlying theme when it comes to challenges in the preprints landscape is trust. The possible challenges and risks noted in this report could, in time, be overcome under the assumption that authors, media and the other stakeholders in this area work professionally and ethically.

Posting preprints in practice

In the course of this project, we sought to develop an overview of existing preprint servers (see **section 4**, p.23 and **Appendix C**, p.54 for details). We identified over 60 platforms that can be used to store and share preprints, though a handful are online repositories with a wider scope that also accept preprint posting (e.g. Zenodo, figshare, Open Science Framework).

Thanks to the fast development of new preprint servers over the last ten years, the availability of a server that is fit for the purpose of any given researcher is almost guaranteed. Preprint servers ranging from physics to humanities, medicine to agriculture, geosciences to mind and contemplative practices are now available, as well as generalist repositories and servers with a national/geographic/language focus. While the sheer variety risks creating confusion in some cases, we note that the vast majority of these new preprint servers have been started from the bottom up and are being maintained by tight disciplinary communities.

The enabling role of Twitter

A surprising finding of this study was that Twitter plays an enormous role in supporting the uptake of preprints. Many participants mentioned that Twitter is the way they were first exposed to preprints. Typical experiences included a peer sharing a preprint or a member of an interviewee's network commenting on one.

Footnotes

- 37** Efficiency and Standards for Article Charges – ESAC. (n.d.). Market watch. <https://esac-initiative.org/about/apcmarket>
- 38** Tennant, J. & Lomax, D. (2019). An overview of Open Access publishing in palaeontology. <https://doi.org/10.26879/968>
- 39** For more information on open peer review, see: Johnson, R., Watkinson, A. & Mabe, M. (2018). The STM Report. An overview of scientific and scholarly publishing. stm-assoc.org/2018_10_04_STM_Report_2018.pdf
- 40** Kirkham, J. & Moher, D. (2018). Who and why do researchers opt to publish in post-publication peer review platforms? - findings from a review and survey of F1000Research. <https://doi.org/10.12688/f1000research.15436.1>
- 41** Cosgrove, A. & Cheifet, B. (2018). Transparent peer review trial: the results. <https://doi.org/10.1186/s13059-018-1584-0>
-

“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.”

Researcher

Twitter is widely known as a social network, and its use by academics for professional purposes, though still limited, is growing.⁴² An open and publicly available medium by nature, Twitter is increasingly being used by scientific communities to:

- ▶ Follow Twitter bots posting preprints as set up by individual preprint servers
- ▶ Share their own preprints
- ▶ Discuss preprints via comments (“replies”, in Twitter jargon)
- ▶ Contact publishers of high-impact journals if a preprint has received significant attention

The point on making and receiving comments is particularly significant. This is possible on some preprint servers and comments can be made either on a full preprint or on specific portions of text (depending on technological solutions). However, when readers comment on preprints via social media they make their feedback less discoverable and, thus, not as accessible to all interested parties. In some cases, altmetrics algorithms might be able to track discussions on Twitter, but this is often possible only if the preprint’s DOI continues to be mentioned. Over the course of this study, which started in September 2018, the follower counts on Twitter of preprint servers under observation (bioRxiv, ChemRxiv and PsyArXiv) increased by between a few hundred and a few thousand individuals/organisations. This supports our claim for the importance of Twitter in enabling the development of preprint servers and signals continued

interest from these disciplinary communities. However, we note that the follower counts of preprint servers accounts remain relatively low in social media terms – bioRxiv is the most popular Twitter account among the three with almost 34k followers, followed by ChemRxiv and PsyArXiv with 7.5k and 4.7k, respectively.

Finally, we note that the EarthArXiv preprint server was launched via a targeted social media campaign that led to a rapid development of the platform:⁴³ this shows yet another way Twitter can be leveraged to promote preprint posting.

Nevertheless, it is important to note that use of Twitter is limited in many parts of the world, most notably China where it is currently blocked. Within an academic context, there is also evidence that social scientists and computer and information scientists are over-represented on Twitter, whereas mathematical, life, and physical scientists are under-represented.⁴⁴

Preprint servers and their authors will therefore need to make greater use of other communication and discovery channels in the future if they are to assume a central position in the scholarly discourse. These may include, among other solutions, the setup of custom alerts, a function already implemented by arXiv.

Footnotes

- ⁴² Mohammadi, E., Thelwall, M., Kwasny, M. & Holmes, K.L. (2018). Academic information on Twitter: A user survey. <https://doi.org/10.1371/journal.pone.0197265>
- ⁴³ Narock, T. W., Goldstein, E., Jackson, C. A., Bubeck, A., Enright, A., Farquharson, J. I., ... Ampuero, J. (2018). Earth Science is Ready for Preprints: The First Year of EarthArXiv. <https://doi.org/10.1029/2019EO121347>
- ⁴⁴ Ke Q., Ahn Y-Y., Sugimoto CR (2017). A systematic identification and analysis of scientists on Twitter. PLoS ONE 12(4): e0175368. <https://doi.org/10.1371/journal.pone.0175368>

4. Mapping the preprints landscape

The recent growth of preprint servers means that the landscape now includes over 60 solutions for a range of disciplines, languages and countries. Many stakeholders are involved when it comes to preprints, such as researchers as both authors and readers, research performing organisations, funders, publishers, service providers and more. The technology underpinning preprints is widely available, but some infrastructural challenges exist. The ‘hype’ in areas where the preprints culture is currently growing might be temporary, but it appears likely that at least some of the existing servers are here to stay.

Key stakeholders

This study started by considering researchers, RPOs, research funders and service providers. As discussed in our Methodology section, we did not interview academic publishers, but they are here discussed to reflect their significant role in scholarly communication (**Table 1**, p.29).

The preprints landscape is evolving fast and, in some cases, in a fragmented manner:

- ▶ New preprint servers are being regularly started up
- ▶ There is significant experimentation in terms of approaches and technologies
- ▶ Little collaboration is in place between existing players

While these are not issues per se, we highlight that broader acceptability of preprint posting could benefit from a reduction in the current extent of variability and uncertainty in the landscape.

A possible way forward is cooperation between the stakeholders listed in **Figure 4** (p.24) and **Table 1** (p.29), but this will depend on whether this is seen as desirable by all. Most likely, some extent of coordination across scholarly communication stakeholders will be

needed in the future, as lasting cultural change is significantly easier to achieve when mandates, expectations, practices and infrastructure are aligned. As a starting point, we note that important sector stakeholders have taken notice of the preprints movement.^{19, 20, 22}

The fact that influential players such as **Crossref** (<https://github.com/CrossRef/rest-api-doc>), **Europe PMC** (<https://europepmc.org/downloads/preprints>) and **Google Scholar** (<https://scholar.google.com/intl/en/scholar/publishers.html>) are now formally recognising preprints is a significant step towards recognising the inclusion of preprints in mainstream academic workflows. However, more efforts are required to ensure the stakeholder groups involved have a shared understanding and some agreement on what comes next in this fast-moving area.

Footnotes

- ¹⁹ cOAlition S. (2019). Principles and Implementation. coalition-s.org/principles-and-implementation
 - ²⁰ Sever, R., Eisein, M. & Inglis, J. (2019). Plan U: Universal access to scientific and medical research via funder preprint mandates. planu.org
 - ²² Elsevier, Ipsos MORI. (2019). Research futures: Drivers and scenarios for the next decade. elsevier.com/connect/elsevier-research-futures-report
-

Technology and operations

We identified 60+ platforms allowing the sharing of preprints and highlight different approaches to technical implementation:

- ▶ 25 preprint servers in our sample are based on the Open Science Framework, which is open source and gaining popularity
- ▶ 18 servers employ proprietary/ad-hoc solutions (note that this has no impact on the openness of the preprints they host)
- ▶ Six are based on the EPrints digital repository solution
- ▶ Other solutions, used by three or fewer preprint servers, include figshare, DSpace, Invenio, F1000 and Drupal

The choice between the above options by the preprint server mainly has consequences in terms of user

experience and web design: these considerations apply to readers accessing preprints (website interface) and to authors sharing them (posting workflows). Furthermore, some commercial solutions may be in a better financial position to invest in developing the relationships and interfaces needed to maximise content discoverability.

The choice of technology is somewhat related to the way platforms operate in practice and what kind of control their owners or managers wish to exert. **Figure 4** shows that either authors or publishers are typically responsible for sharing pre-refereed content, and that the platforms that enable this can be either standalone or owned/managed by publishers.

Footnotes

45 Narock, T. & Goldstein, E.B. (2019). Quantifying the Growth of Preprint Services Hosted by the Center for Open Science. <https://doi.org/10.3390/publications7020044>

Figure 4 – Models for the sharing of research prior to peer review and examples of platforms and publishers⁴⁵

Publisher owned/ managed platform	SSRN Preprints.org		F1000Research
	Proprietary technology arXiv	Third-party technology bioRxiv ChemRxiv Preprint servers based on OSF	bioRxiv/PLOS
Standalone platform	Author posting		Publisher posting

One of the main differences between author and publisher posting is that publishers would tend to post pre-refereed content as part of a holistic publishing process. However, publishers do not necessarily need their own platforms to include preprints within their workflows: for example, PLOS has partnered with bioRxiv to achieve this, effectively creating a publisher-triggered workflow that uses a standalone community resource.

Finally, we note that the use of proprietary technology allows the highest level of control and customisation to preprint server managers and publishers such as F1000. With solutions developed in-house, any functionality could, at least potentially, be implemented. This comes with the need to employ staff to carry out web development and technical maintenance, however. The use of third-party solutions reduces the technical burden on the preprint server's administration, but, for example, in the case of ChemRxiv means that the cost of a commercial solution must be covered.

Size and trends in the preprints landscape

The number of preprints hosted by a single server can vary significantly: our research shows figures ranging between ~20 preprints in Medieval Studies and the over 1.5 million hosted on arXiv. This comparison highlights the impact of disciplinary culture on the posting of preprints: some disciplines are just starting to experiment, while the physics, mathematics and computer science communities lead the way. However, the size of disciplinary communities and the average pace of research in different areas will also affect these figures.

Another factor to keep in mind when looking at preprint counts is that researchers in some areas (chiefly the humanities) may value monographs more than they do articles. Therefore, the extent to which preprints (meant as pre-publication versions of articles) will matter to them might be limited. The idea of sharing preprint versions of monographs has emerged more recently,^{46, 47} but neither

the literature nor our interviews suggest any consensus is emerging on the way forward. We note, however, that the very first experiments with electronic book publishing, although unrelated to preprints and closer to open/free access, date as far back as 1971, when Project Gutenberg was started.⁴⁸

Finally, we highlight the global reach of preprints. There are some preprint servers with a geographic focus, for instance, for Africa, China, India and Indonesia,⁴⁹ or with a language focus, for instance, Arabic or French. At present, it is not possible to forecast with any certainty whether this approach will co-exist alongside the use of preprint servers by disciplinary communities, and what the respective merits of each model might be. However, studies of OA publishing practices indicate that authors' disciplinary affiliations tend to carry greater weight than national loyalties or their country of residence.⁵⁰

Footnotes

- 46 Geltner, G & Willinsky, J. (2018). Preprint to Monograph: A Path to Travel By. guygeltner.net/blog/652018preprint-to-monograph-a-path-to-travel-by
 - 47 Springer Nature, Pyne, R., Emery, C., Lucraft, M. & Pinck, A.S. (2019). The future of open access books: Findings from a global survey of academic book authors. <https://doi.org/10.6084/m9.figshare.8166599.v1>
 - 48 Moore, S. (2019). Revisiting 'the 1990s debutante': scholar-led publishing and the pre-history of the open access movement. <http://dx.doi.org/10.17613/gty2-w177>
 - 49 Mallapaty, S. (2019). Indian scientists launch preprint repository to boost research quality. <http://doi.org/10.1038/d41586-019-01082-0>
 - 50 Eger, T., and Scheufen, M. (2018). The Economics of Open Access: On the Future of Academic Publishing. jipitec.eu/issues/jipitec-9-3-2018/4812
-

Infrastructural challenges

Based on our landscape review, it appears clear that the technology to support the uptake of preprints is available. For example, DOIs or unique identifiers/permalinks can be assigned to preprints (this has been happening since the early 1990s on arXiv),¹¹ withdrawals are possible on preprint servers and open licensing options are currently available. However, some challenges remain:

- ▶ It is currently difficult to automatically track manuscripts through the publication process, as preprint servers and academic publishers do not have shared workflows
- ▶ Digital preservation is a concern due to its cost and is not considered an immediate priority due to the extent of experimentation in the landscape
- ▶ It is sometimes difficult to identify that a given research output is a preprint based solely on its metadata
- ▶ Preprints servers accept submissions with no requirements in terms of layout, which may lead to articles that appear poorly formatted compared to their journal equivalents

In some cases, we note that the above challenges are not as significant: as an example, F1000Research currently includes the sharing of articles prior to and under open peer review within their publication workflows. Therefore, matters related to tracking and metadata are more easily resolved as a single platform can manage both the preprint and the published version.

We also highlight the issue of permanence of deposition: completely removing documents from preprint servers in cases where authors wish to withdraw their work is not seen as good practice. In such cases, it is advisable to include a withdrawal statement on the server, but the submission should remain available unless it has to be removed for legal reasons: the rationale for this is that

the DOI system aims to make submitted works citable and part of the scientific record (as an example, see the policy by **bioRxiv** - [biorxiv.org/about/FAQ](https://www.biorxiv.org/about/FAQ)). Preprints.org and SSRN currently offer an option for authors to remove their work posted in preprint form; we note that Preprints.org has compiled guidance on what this entails and clearly states that DOIs will not be created in cases where authors wish to retain this level of control.⁵¹

Information overload and information overlap

We investigated whether the posting of preprints might be perceived as a contributor to information overload. This can be described as the phenomenon where a researcher feels like they are “barely keeping [their] head above the flood of information” due to the increasing amount of scholarly material available online.⁵² Aside from the fact that new tools are growing to help researchers make sense of all the literature that is now available (e.g. **Iris (Iris.ai)**, **Open Knowledge Maps (https://openknowledgemaps.org)**, **ScienceOpen (scienceopen.com)**), preprints are not seen as significantly worsening the issue, which is, on the other hand, seen as an opportunity by some.⁵³ The number of articles published yearly worldwide only keeps increasing, which suggests that preprints, at least at this stage, are but another drop in the ocean of content.³⁹

Footnotes

- ¹¹ Penfold, N. (2019). Twitter (29 May 2019). <https://twitter.com/npscience/status/1133734784939769856>
 - ³⁹ Johnson, R., Watkinson, A. & Mabe, M. (2018). The STM Report. An overview of scientific and scholarly publishing. [stm-assoc.org/2018_10_04_STM_Report_2018.pdf](https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf)
 - ⁵¹ Preprints Editorial Office – Preprints.org. (2019). Update of Preprint DOI registration. preprints.org/announcement/show/33
 - ⁵² Landhuis, E. (2016). Scientific literature: Information overload. [http://doi.org/10.1038/nj7612-457a](https://doi.org/10.1038/nj7612-457a)
 - ⁵³ Neylon, C. (2011). It's not filter failure, it's a discovery deficit. [http://doi.org/10.1629/2421](https://doi.org/10.1629/2421)
-

“I think there is a lot of information out there, but I think there’s also the potential to find technical solutions that will avoid the information overload.”

Preprint server provider

However, our interviews and literature review unearthed an issue related to information overload, which we call “information overlap”. Going back to the way people interpreted the definition of a preprint (see section 3, p.16) it is significant that some considered this as any version of a paper prior to publication. This, for example, includes author accepted manuscripts (AAMs), which in some countries may also be deposited in (sometimes multiple) institutional or national repositories (e.g. the former is the case in the UK and the latter in France using the HAL platform). The natural question is therefore whether this overlap between the scope of preprint servers and other repository solutions is desirable. Although an overlap doesn’t seem particularly harmful in itself, we highlight the following questions:

- ▶ What is the rationale for preserving an AAM in both an online repository and a preprint server, particularly where deposit in a repository might be mandated nationally? (e.g. this is the case in the UK)
- ▶ Who should be responsible for covering the costs of technical infrastructure where this duplicates efforts that are already otherwise funded?

“I systematically put all my preprints on arXiv and I will put them on HAL, too.”

Research funder and researcher

These questions are also related to open access to research articles: if an output has been published via open access and is therefore already available online as the publisher’s version of record, is there a real need to preserve its preprint form? Some might argue that a preprint shows the “history” of the article, as authors might have uploaded different versions as their work went through peer commenting and peer review; this, however, doesn’t seem to be a strong argument at present, as the uptake of versioning features on preprint servers appears limited. Preserving historical copies of a research article may, indeed, be useful in some cases – for example where the preprint includes additional content compared to the final peer-reviewed version (e.g. if the article has to be shortened based on journal guidelines).

The point of the above critical questions is largely to ensure the future financial sustainability of preprint servers (see section 5, p.30): the more outputs have to be hosted and preserved in the long term, the higher maintenance and server costs will tend to be. However, we also note that the archival of scholarly content, even when this is available in open access form, should not be outsourced carelessly: at present, most scholarly content is accessed on publishers’ servers and libraries themselves do not have copies they can preserve.³⁹ Some initiatives are operating to address this issue, such as **LOCKSS** (lockss.org) and **CLOCKSS** (<https://clockss.org>), but preprints do not appear to be included within their scope.

Footnotes

- ³⁹ Johnson, R., Watkinson, A. & Mabe, M. (2018). The STM Report. An overview of scientific and scholarly publishing. stm-assoc.org/2018_10_04_STM_Report_2018.pdf

Riding the hype wave?

Preprint servers and related services have been growing particularly fast over the past few years. However, the enabling factor of this growth dates back to the 1970s, when the digital revolution started. The ensuing advances in computer science and database systems led to the first preprint servers in the 1990s, and this can be seen as the so-called “technology trigger” that empowered the research community to use preprints broadly.

The idea of preprints growing following an initial enabling event is in line with the concept of the hype cycle, which can be used to qualitatively examine trends in innovation.⁵⁴ The hype cycle includes five phases (see **Figure 5**) through which innovation often goes and represents the visibility of a given phenomenon in time.

In the case of preprints, most disciplines are now experiencing growth. Based on a hype cycle interpretation, the hype around the new wave of preprints and preprint servers we are arguably seeing at the moment may be expected to peak and be followed by a period of realism or even disillusionment. However, after this, it may be

expected that at least some servers then move on to the “slope of enlightenment” and the “plateau of productivity”. Of course, we wouldn’t expect this to happen in all cases. Some disciplines such as those served by arXiv or RePEc have already reached a stage of maturity (the “plateau of productivity”) and are unlikely to go through the cycle again.

The hype cycle interpretation is not intended to dampen enthusiasm towards preprints and preprint servers; it simply aims to highlight that the expected level of uptake may not be fully met and some players might merge or disappear in time. This particularly refers to technological solutions and platforms, rather than to the uptake of preprint posting itself.

Footnotes

- ⁵⁴ Fenn, J. & Raskino, M. (2008). *Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time*. worldcat.org/title/mastering-the-hype-cycle-how-to-choose-the-right-innovation-at-the-right-time/oclc/213312226

Figure 5 – Preprints and the hype cycle

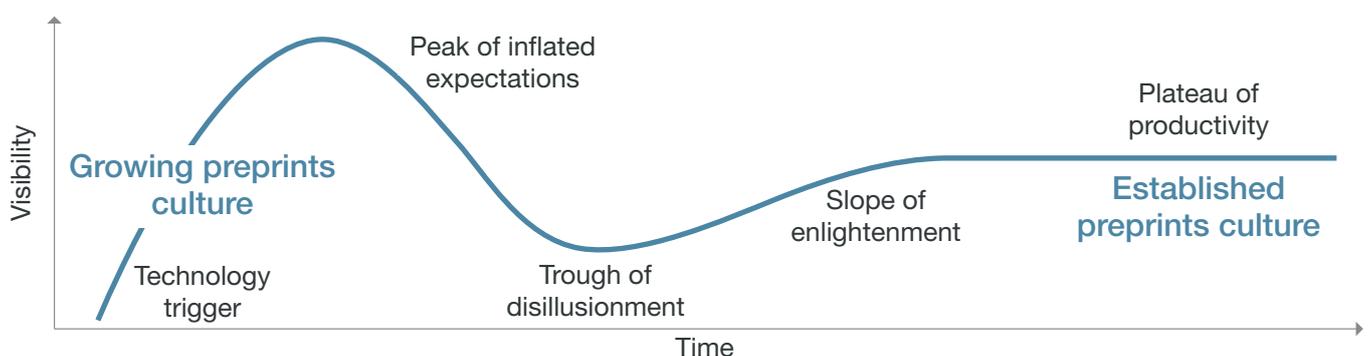


Table 1 - Key players in the preprints landscape and roles

Stakeholder group	Key current role(s)	Possible future role(s)
Researchers and disciplinary communities	<ul style="list-style-type: none"> ▶ Writing, posting, reading and reusing preprints, but to highly varying degrees by discipline 	<ul style="list-style-type: none"> ▶ Increased, or even universal, adoption of preprints as a form of scholarly communication ▶ Reviewing and commenting on preprints
Research performing organisations	<ul style="list-style-type: none"> ▶ Supporting researchers with information and help where required 	<ul style="list-style-type: none"> ▶ Promoting and advocating for the use of preprints ▶ Assessing preprints in recruitment, review, promotion and tenure processes
Research funders	<ul style="list-style-type: none"> ▶ Assessing preprints in grant proposals 	<ul style="list-style-type: none"> ▶ Promoting and advocating for the use of preprint ▶ Mandating the use of preprints (if desirable)
Preprint servers	<ul style="list-style-type: none"> ▶ Hosting preprints ▶ Promoting and advocating for preprints 	<ul style="list-style-type: none"> ▶ Promoting best practices ▶ Sharing preprints and metadata with aggregators and academic publishers
Other service providers	<ul style="list-style-type: none"> ▶ Quality assuring preprints (overlay journals) ▶ Promoting the use of preprints (preprint journal clubs) 	<ul style="list-style-type: none"> ▶ Developing new business models ▶ Enabling/supporting TDM services
Publishers	<ul style="list-style-type: none"> ▶ Quality assuring preprints submitted to them (i.e. carrying out peer review on submissions received in preprint form) ▶ Accepting or rejecting the practice of preprint posting 	<ul style="list-style-type: none"> ▶ Sharing publication status with preprint servers ▶ Cooperating with preprint servers to integrate article submission workflows ▶ Hosting preprints

5. The future of preprints

A pressing question is where the responsibility to post preprints will lie. Researchers can take care of this themselves, but this would likely lead to partial uptake. On the other hand, publishers may be able to post a preprint for virtually any article submitted to them, but there is rising concern over levels of market consolidation. The extent to which preprint posting addresses perceived pain points may be limited, but the practice does call into questions the role and proposition of traditional journals.

Taking the lead in preprint posting

In **section 4** (p.23), we discussed some of the technical platforms used by existing preprint servers and how they operate in practice.

Our study investigated the question of whether preprint posting will evolve as a researcher- or publisher-centric phenomenon, as this will affect the technologies and workflows considered. Most of the funders, librarians and researchers participating in this project highlighted that they would strongly prefer not-for-profit and publisher-neutral models (e.g. funded via consortia).

“As there is huge value in posting preprints before submitting to journals, I think this cannot be left to publishers. They can support it, but open science is publisher independent.”

Research Performing Organisation

The current landscape is characterised by widespread experimentation when it comes to preprint posting, and we note that business models do not appear to be a priority. Therefore, the above preference clearly depends on whether funding will be available in the future, and from what sources.

The choice of business model and the level of involvement of academic publishers will affect the simplicity of workflows to track the status of a preprint throughout and after publication (**Table 2**, p.36):

- ▶ If a researcher-centric model based on standalone preprint servers prevails, new workflows and automations connecting preprint servers and publishing systems will be desirable, but we note that their creation and implementation are potentially difficult
- ▶ If a publisher-centric model prevails, the tracking of preprints throughout and after publication will be simpler, as the publisher would have all the information and update its own internal and public-facing records accordingly

Furthermore, the experimentation embraced by most independent preprint servers today is typically supported by grants, time-limited funds or in-kind contributions. For preprints and preprint servers to take a more central role in the scholarly communication landscape going forward, there will be a need for either (i) a long-term funding commitment from public actors; or (ii) sustainable commercial business models. This also leads to the question of the opportunity cost of preprints and preprint servers. The availability of a wide range of solutions, all needing funding, appears to represent a risk, as multiple players might have to compete for resources in an increasingly complex landscape.

At present, we cannot say which approach will prevail (if any) due to the continuous changes in the area. The considerations listed in **Table 2** (p.36) are likely to play a significant role in future developments in the preprints landscape. We note that there are both long-running and novel initiatives to support open infrastructure, and that a range of organisations in higher education and research are making global efforts in this direction.⁵⁵ Notably, any step towards future financial sustainability will need to consider how important alignment with open research practices and independence from publishers are to the numerous stakeholder groups involved.

Cultural resistance

Early developments in the economics preprints community saw resistance to the use of a single solution to host preprints. This appeared to be related to distrust of monopolies in the economics research landscape.¹⁵ In this case, there was no opposition to preprints themselves: the issue was with the need to use a specific solution that might be perceived as being too influential.

Should academic publishers (who are already seen by some as overly powerful)⁵⁶ organically grow as the providers of preprint services in some communities, overcoming the above resistance might be an important area of focus.

The evolving roles of academic journals, preprint servers and overlay services

The role of academic journals in scholarly publishing is evolving; however, there is a general consensus that, typically, they support the registration, curation, evaluation, dissemination and archival of academic research.^{57, 26} In this study, interviewees reported that preprints posted online are normally of a high quality, as there would otherwise be a risk of reputational damage for the posting author(s). Therefore, preprint servers might host research that is suitable for formal publishing with only minor revisions, although not typeset or peer-reviewed, and only inconsistently preserved. Preprints are also increasingly present in scholarly

conversation around the future sustainability of open access.⁵⁸ As the cost of posting a preprint on arXiv is reported to be less than \$10⁵⁹ and open access article processing charges may be well beyond \$2,000, some are asking that journals make a clearer case for this difference. This is not to dismiss the important role played by academic journals in the scholarly communication landscape, but to reflect the significance of the debates sprouting from the diffusion of preprints.

Footnotes

- 15** Karlsson, S. & Krichel, T. (1999). RePEc and S-WoPEc: Internet access to electronic preprints in Economics.
<http://openlib.org/home/krichel/papers/lindi.html>
- 26** Directorate-General for Research and Innovation (European Commission). (2019). Future of scholarly publishing and scholarly communication.
<https://doi.org/10.2777/836532>
- 55** Joseph, H. (2019). Investing in Open Scholarly Infrastructure: a Community Opportunity.
<https://sparcopen.org/news/2019/investing-in-open-scholarly-infrastructure-a-community-opportunity>
- 56** Larivière, V., Haustein, S., Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era.
<https://doi.org/10.1371/journal.pone.0127502>
- 57** Wouters, P., Sugimoto, C.R., Larivière, V., McVeigh, M.E., Pulverer, B., de Rijcke, S. & Waltman, L. (2019). Rethinking impact factors: better ways to judge a journal.
<http://doi.org/10.1038/d41586-019-01643-3>
- 58** Green, T. (2019). Are we being wilfully blind about the transformation that's needed in scholarly publishing?
<https://medium.com/@TobyABGreen/are-we-being-wilfully-blind-about-the-transformation-thats-needed-in-scholarly-publishing-d0bfb61d1f05>
- 59** Ball, P. (2015). Leading mathematician launches arXiv 'overlay' journal.
<https://doi.org/10.1038/nature.2015.18351>
-

We also note the role of overlay journals, that is, platforms that source freely available content online (including preprints) and then evaluate its worth, in many cases via peer review. The **Discrete Analysis** (<https://discreteanalysisjournal.com>) overlay journal was called in a Nature article “the journal that publishes no papers”:⁶⁰ it gathers arXiv articles and shows the level of trust that is currently conferred on work in preprint form by the mathematics community. Clearly, the vast majority of disciplines are yet to reach this stage; however, if we consider arXiv as a trend-setter in this landscape, traditional journals may need to carefully consider how their proposition will evolve in time alongside the potential growth of preprints in other disciplines.

As an example, **Peer Community In** (<https://peercommunityin.org>) is currently serving evolutionary biology, ecology, palaeontology, animal science and entomology, providing recommendations for preprints (and published articles) based on peer review. Although bearing some differences from overlay journals, they recently released their economic model: the organisation is non-profit and non-commercial, and their running cost is significantly lower than that of mainstream publishers.⁶¹ The scalability of Peer Community In and similar initiatives remains unproven, and the existing reward system within academia is skewed heavily in favour of established journals. Nevertheless, with preprints continuing to grow in popularity and funders signalling their desire to disrupt existing reward mechanisms,²⁰ the possibility of disruption to the academic journals market cannot be discounted.

Finally, we note the recent launch of the preLists initiative,⁶² which allows early-career researchers to create curated lists of preprints and make these available to the community. This is yet another way preprints can be grouped and shared, although no form of review is included in this case.

Licensing and preprints

Licensing is key to enable the reuse of research outputs: it typically determines whether (i) attribution is required; (ii) derivative work can be shared; and (iii) commercial use is permitted.

Licensing is widely discussed when it comes to traditional journal publishing, and the stakeholders we interviewed mentioned the need to bring this topic to the preprint community’s attention, too.

Copyright literacy and familiarity with options such as Creative Commons licences are not particularly high at present,⁶³ which was confirmed in our interviews. Several researchers, when asked, were not able to fully justify their choice of licence for the preprint they had posted and were discussing with us. A handful even admitted not being aware of any licence currently applied to their preprint. The intent was usually to share their research freely, which they had clearly achieved, but there was little to no understanding of any restrictions on the reuse of their work.

Footnotes

- 20 cOAlition S. (2019). Principles and Implementation. coalition-s.org/principles-and-implementation
 - 60 Ball, P. (2015). The journal that publishes no papers. [nature.com/polopoly_fs/1.18444!/menu/main/topColumns/topLeftColumn/pdf/526146a.pdf](https://www.nature.com/polopoly_fs/1.18444!/menu/main/topColumns/topLeftColumn/pdf/526146a.pdf)
 - 61 Peer Community In. (2019). PCI economic model. <https://peercommunityin.org/2019/05/29/pci-economic-model>
 - 62 STM Publishing. (2019). The Company of Biologists launches preLists. stm-publishing.com/the-company-of-biologists-launches-prelists
 - 63 Secker, J. & Morrison, C. (2018). Copyright literacy in the UK: Understanding library and information professionals’ experiences of copyright. <http://openaccess.city.ac.uk/id/eprint/20082>
-

“I actually don't really know enough about licensing to be honest. The licence I put in was the licence that I found.”

Researcher

The communication of licensing options to authors needs to play a central role if the promise of preprints is to be delivered, as enhanced reuse is only possible when permissive licences are used.

This leads to considerations on text and data mining applications (TDM). This approach is often difficult to implement in practice due to the need to navigate complex licensing agreements via institutional subscriptions or ad-hoc contracts. However, if preprints are posted online with permissive licences, there would be potential for TDM to be carried out more easily (notably, a previous study found that only 17.8% of bioRxiv papers had a permissive CC BY licence).^{64, 65}

Furthermore, TDM works best when documents are carefully structured, for instance, when using the widespread XML format. At present, preprints posted by authors tend to be in Microsoft Word or pdf format, which means that TDM requires additional conversion or interpretation efforts before analysis.⁶⁶ On the other hand, F1000Research would typically share preprints using the XML format in the first place. Therefore, structured XML versions of articles are made available for download for submissions awaiting peer review.

A possible area for future work is the inclusion of templates for authors on preprint servers, as improved and standardised document structures could simplify TDM activities. At the same time, authors could present their work in a more structured and accessible way (e.g. figures not at the end of the manuscript, better looking layout), which is another desirable outcome. However, we note a possible trade-off: the additional effort required

from authors to structure their articles in a specific way might negatively affect the uptake of preprint posting.

“The problem is that preprints are usually being uploaded like in a PDF form, so machines don't have really that much access to structured data.”

Service provider

Overall, we would stress the important role of preprint servers (whatever their form or owner/manager) in ensuring authors are presented with relevant and useful information on licensing. As an example, ASAPbio have a **Preprint licensing FAQ** (<https://asapbio.org/licensing-faq>), including an infographic and textual explanations. Ideally, this type of information should always be presented to authors before they make their choice of licence, so as to ensure the possibility of reuse of their work is maximised.

Footnotes

- 64 Abdill, R.J. & Blekhman, R. (2019). Meta-Research: Tracking the popularity and outcomes of all bioRxiv preprints. <https://doi.org/10.7554/eLife.45133>
 - 65 Himmelstein, D. (2016). The licensing of bioRxiv preprints. <https://blog.dhimmel.com/biorxiv-licenses>
 - 66 Simboli, B. (2019). arXiv and the Symbiosis of Physics Preprints and Journal Review Articles. <https://arxiv.org/abs/1904.01470v2>
-

Preprints, responsible metrics and evaluation

Preprint servers typically display a varying range of online metrics including server-wide ones (e.g. total preprints posted) and preprint-specific ones (e.g. views, download count, citations received, altmetrics).^{67,68} However, the role played by preprints and their metrics when it comes to review, promotion and tenure is currently limited. Our interviewees discussed this in detail, and it appears that preprints are not being considered in RPOs for these purposes. Funders are starting to accept preprints in grant proposals (e.g. National Institutes of Health, Zuckerberg Foundation, Wellcome Trust, European Research Council, European Molecular Biology Organization), but they are seen as less valuable than peer-reviewed articles. On the other hand, the Deutsche Forschungsgemeinschaft has been accepting preprints for a long time now, and the extent to which they are valued is determined by disciplinary norms. We note that preprints are discussed in slightly different forms by funders: some mention them explicitly, while, in other cases, preprints are considered as acceptable research outputs under umbrella terms such as “other publications”.

The role of initiatives such as **DORA** (<https://sfedora.org>) was acknowledged by a number of interviewees: DORA aims to promote alternative ways to evaluate scientific research, to formally acknowledge the existence of different types of research outputs and forms of impact beyond publishing in prestigious academic journals. In this context, we note that preprints are seen as playing a role for early-career researchers,²⁸ particularly when applying for jobs or grants before having had the chance to formally publish research. These advantages are often only anecdotal, but there are cases where preprints did lead to hiring in practice.⁶⁹

Using preprints in academic hiring decisions may be seen as a positive for a number of reasons:²⁸

- ▶ Candidates could be identified without delays, i.e. there is no need to wait for formal publication
- ▶ Evaluating researchers may be simpler, as their preprints are publicly accessible
- ▶ Candidates interested in forward-looking scholarly practices could be identified, as posting preprints is a recent trend in open scholarship
- ▶ It may be possible to focus more on individuals and the research rather than on journals, which is in line with ongoing debate on the value of journal impact factors and citation-based indicators.^{70,71}

Future engagement with the stakeholders in the preprints landscape will determine whether using

Footnotes

- 28 Sarabipour, S., Debat, H.J., Emmott, E., Burgess, S., Schwessinger, B. & Hensel, Z. (2018). On the value of preprints: an early career researcher perspective. <https://doi.org/10.7287/peerj.preprints.27400v1>
- 67 Balaji, B.P. & Dhanamjaya, M. (2019). Preprints in Scholarly Communication: Re-Imagining Metrics and Infrastructures. <https://doi.org/10.3390/publications7010006>
- 68 Fraser, N., Momeni, F., Mayr, P. & Peters, I. (2019). The effect of bioRxiv preprints on citations and altmetrics. <https://doi.org/10.1101/673665>
- 69 DORA. (2018). Preprints in Academic Hiring. <https://sfedora.org/2018/08/14/preprints-in-academic-hiring>
- 70 Pudovkin, A.I. (2018). Comments on the Use of the Journal Impact Factor for Assessing the Research Contributions of Individual Authors. <https://doi.org/10.3389/frma.2018.00002>
- 71 Wang, L. & Zhan, Y. (2019). A conceptual peer review model for arXiv and other preprint databases. <https://doi.org/10.1002/leap.1229>
-

preprints in all these scenarios is perceived as appropriate; however, we note that, in fields where the use of arXiv is frequent, this is already happening.⁷²

Addressing perceived pain points

The brief history of preprints in **section 1** (p.10) shows that electronic preprints in the physics and economics communities became widespread to meet a clear need: researchers wanted to share pre-refereed research fast and more broadly and no suitable way was available other than physically posting hard copies. A practical need, therefore, led to change: arXiv and WoPEc/ RePEc were answers to perceived issues.

In many cases, today's preprint servers are started by enthusiastic proponents of open scholarship, in order to pursue free sharing, transparency and increased research impact. The extent to which this might meet a practical need in their communities is not always clear. Even if a growing number of researchers are supportive of open scholarship principles, this may not be strong enough to drive behavioural change in the short term. One of the most significant pain points that posting preprints addresses is the slow pace of the academic publishing process. This, however, is not the case in all disciplines. Therefore, the "sense of urgency"⁷³ that might drive some researchers to post preprints may not be shared by the academic community as a whole.

"In my case, I would tend to prefer to just try to publish in open access journals [rather than posting a preprint] - in general, the review times in my field are not as horrible as in other fields."

Researcher

The vision for open scholarship is evolving and is being discussed worldwide, but the role of preprints and preprint servers is only one of its facets. If open access, which is perhaps the most visible side of open

scholarship at present, is yet to become anchored in scholarly culture, it is understandable that the preprints movement may be lagging somewhat behind. Notably, however, a sense of urgency has been introduced via various governmental and funder mandates in the case of open access policy, while nothing of the sort has happened with respect to preprints to date (in the first place, because in many disciplines preprints have become globally significant only recently, and secondly due to the difficulty of implementing and monitoring such mandates).

Supporting the preprints movement

Our interviews highlighted that, at present, it is not clear who should be responsible for promoting preprints and any advocacy efforts. This role is currently taken by preprint servers themselves (see **Table 1**, p.29), and our interviews highlighted that RPOs do not see preprints as a priority, mainly due to the low level of maturity of this practice and the uncertainty around their weight in researcher evaluation and funding applications.

Future roles in the preprints landscape will likely be shaped by disciplinary communities, but there is scope for RPOs and funders to provide more support to authors. Preprint servers themselves can advise authors to some extent, but researchers would have to consult with RPOs and funders when it comes to their own policies and accepted practices. In the meantime, continued uncertainty around journal policies, and whether posting preprints might affect a researcher's career or performance evaluations, is liable to act as a brake on wider uptake of the practice.

Footnotes

⁷² Vale, R.D. (2015). Accelerating scientific publication in biology. <https://doi.org/10.1073/pnas.1511912112>

⁷³ Kotter, J.P. (2007). Winning at change. providersedge.com/ehdocs/transformation_articles/WINNING_AT_CHANGE.pdf

Table 2 - Differences between a researcher- and a publisher-centric approach

Model	Expected benefits for the research community	Expected drawbacks for the research community
Researcher-centric	<ul style="list-style-type: none"> ▶ High alignment with open scholarship principles ▶ Community ownership ▶ Enhanced commitment by individual researchers to transparency and reproducibility ▶ Higher potential for experimentation and inclusion of emerging practices ▶ Lower risk of market consolidation and ensuing need to win over researchers and other stakeholders who might fear this 	<ul style="list-style-type: none"> ▶ New workflows and automations, which may be difficult to implement, would be desirable ▶ Need for funding from the research community (e.g. via consortia) ▶ Higher need for cooperation between publishers and preprint servers ▶ Responsibility to post preprints on researchers, with the risk of low uptake ▶ Higher effort required to carry out TDM due to the format of submissions and the lack of structured XML versions
Publisher-centric	<ul style="list-style-type: none"> ▶ Reduced need for public funding, as publishers could integrate preprint sharing within their workflows ▶ Reduced need to create new workflows and automations ▶ Responsibility to post preprints on publishers, with potential for automation and higher uptake ▶ Lower effort required to carry out TDM, as publishers may post preprints in XML form (this is already happening in some cases, e.g. F1000Research) 	<ul style="list-style-type: none"> ▶ Increased risk of market consolidation and ensuing need to win over researchers and other stakeholders who might fear this ▶ Lack of community ownership ▶ Possible risk of unilateral decisions in terms of infrastructure and features ▶ Possible limitations based on the copyright and licensing options offered

6. Conclusions

Today, the growth of the preprints movement is undeniable, but we note that the practice remains small compared to the size of the academic publishing landscape. We see three possible scenarios for future developments, ranging from 'turn of the tide', where the second wave of preprint servers fades, to 'preprints by default', where growth continues in all fields and preprints reach widespread acceptance by the research community.

Summary and conclusions: Three future scenarios

This study set out to investigate the preprints landscape and, in particular, to better understand the researcher perspective, attitudes of research performing organisations and funders and the values, aims and strategies of service providers. The results arising from our interviews can be broadly summarised as follows:

- ▶ Early and fast dissemination, increased opportunities for feedback and openness are seen as the main benefits of preprints
- ▶ The main concerns over preprints are the lack of quality assurance, media potentially reporting inaccurate research and journals rejecting articles if a preprint has been posted
- ▶ Twitter has been playing a key enabling role in the current second wave of preprints and preprint servers. It also appears to be the main way researchers are exposed to preprints in the first place
- ▶ It is not clear who will be responsible for posting preprints in the long-term – researchers or publishers? This will partly be affected by the availability of sustainable business models

- ▶ Traditional academic journals might have to reframe their value proposition should preprints grow significantly in popularity in the future

Furthermore, this project uncovered and explored a number of workflows, relationships and dependencies in the preprints landscape, which we have summarised in **Figure 6 (p.39)** in relation to the current academic publishing process. Building on our key findings, and assuming that the academic publishing process won't vary significantly in the short-to-medium term, we believe that the following scenarios might describe the future of preprints:

- ▶ **Scenario 1 – Turn of the tide:** the second wave of preprint servers fades, and preprints remain a major component of scholarly communication only in the fields where they already are, i.e. those served by arXiv and RePEC
- ▶ **Scenario 2 – Variable adoption:** preprints grow in some additional fields such as those within the scope of bioRxiv, PsyArXiv and ChemRxiv but not all
- ▶ **Scenario 3 – Preprints by default:** preprints grow in all fields (at different paces) and are accepted by the research community at large

The principles of innovation diffusion theory⁵ can help us understand how the research community might steer the evolution of the preprints landscape. Innovation diffusion theory argues that the adoption of innovation (in this case, preprints) can be supported by clarity and positivity in the community, the choice of effective communication channels and some extent of promotion efforts by one or more change agents. There also need to be norms in the social system which enable adoption to take place. If we consider the factors affecting the rate of adoption of preprints, our research shows that:

- ▶ The proposition of preprints is clear to most, but not everyone is convinced that the practice is appropriate. Furthermore, the effort to submit yet another research output might be a key obstacle, particularly for senior researchers
- ▶ A key communication channel in the preprint arena is Twitter. This is having a major impact and is the way many are first exposed to preprints. We note that Twitter is currently **blocked in China** (https://en.wikipedia.org/wiki/Censorship_of_Twitter#China), so this finding might be closely tied to the geographical context of the study
- ▶ Promotion efforts are currently limited, and it is unclear who, if anyone, might take on this role in future. Open scholarship enthusiasts are promoting preprints within their circles, but this is not sufficient to achieve systemic change
- ▶ The social system in which preprints operate is complex and characterised by a multitude of disciplines and players. 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)

Scenario 1 (Turn of the tide) is expected to materialise if all current efforts to promote preprints fail. This would mean, for instance, failing to reassure authors about the perceived challenges we discussed or finding unsurmountable issues in terms of funding. This scenario appears unlikely, as many stakeholders in the sector have already taken concrete steps to support preprints, including academic publishers (e.g. key publishers accepting preprint posting or F1000Research building publishing workflows including articles in preprint form).

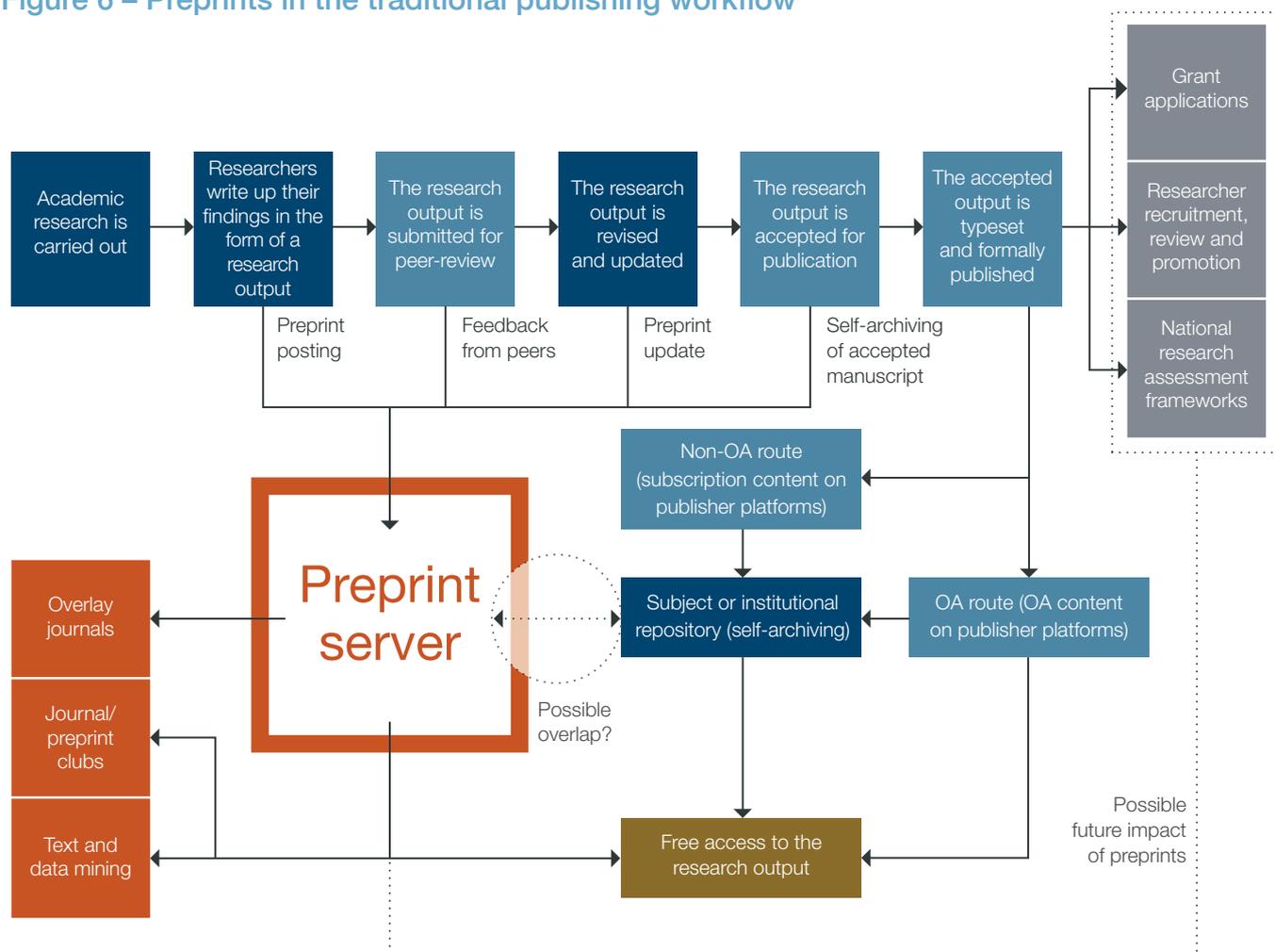
It is possible to foresee a situation where the above three scenarios occur as consecutive steps in a process, with Scenario 2 (Variable adoption) representing a transitional stage between the other two. This might be the case of the preprints landscape today as we move towards broader diffusion. However, we note that Scenario 2 could also be an endpoint. Potentially, some disciplines might simply reject the practice of posting preprints (unless, for example, widespread adoption of funder mandates make it a requirement) and this would mean that Scenario 3 is never reached.

Scenario 3 (Preprints by default) can only happen if all stakeholders involved cooperate to turn the promise of preprints into reality. This would mean carefully evaluating disciplinary approaches, business models, roles and responsibilities, technology and infrastructure, among other things. Scenario 3 is unlikely to be seen in the short or medium term; a way this scenario might be reached more quickly might be the involvement of academic publishers as the posters of preprints of all submissions received – however, this entails some risks and might be subject to heavy criticism due to the desire for community ownership of preprint-related processes and workflows.

Footnotes

- 5 Rogers, E. M. (2003). Diffusion of innovations.
https://books.google.co.uk/books?id=9U1K5LjUOwEC&redir_esc=y
-

Figure 6 – Preprints in the traditional publishing workflow⁷⁴



Legend

- Scholarly communication workflow
- Preprint-related infrastructure, services and initiatives
- Research funding, impact and assessment
- Academic publishers

Footnotes

⁷⁴ Kant, J., T&F Group on Preprints, Chiarelli, A., Johnson, R. & Richens, E. (2019). Preprints – opportunity or challenge? <https://doi.org/10.5281/zenodo.3238499>

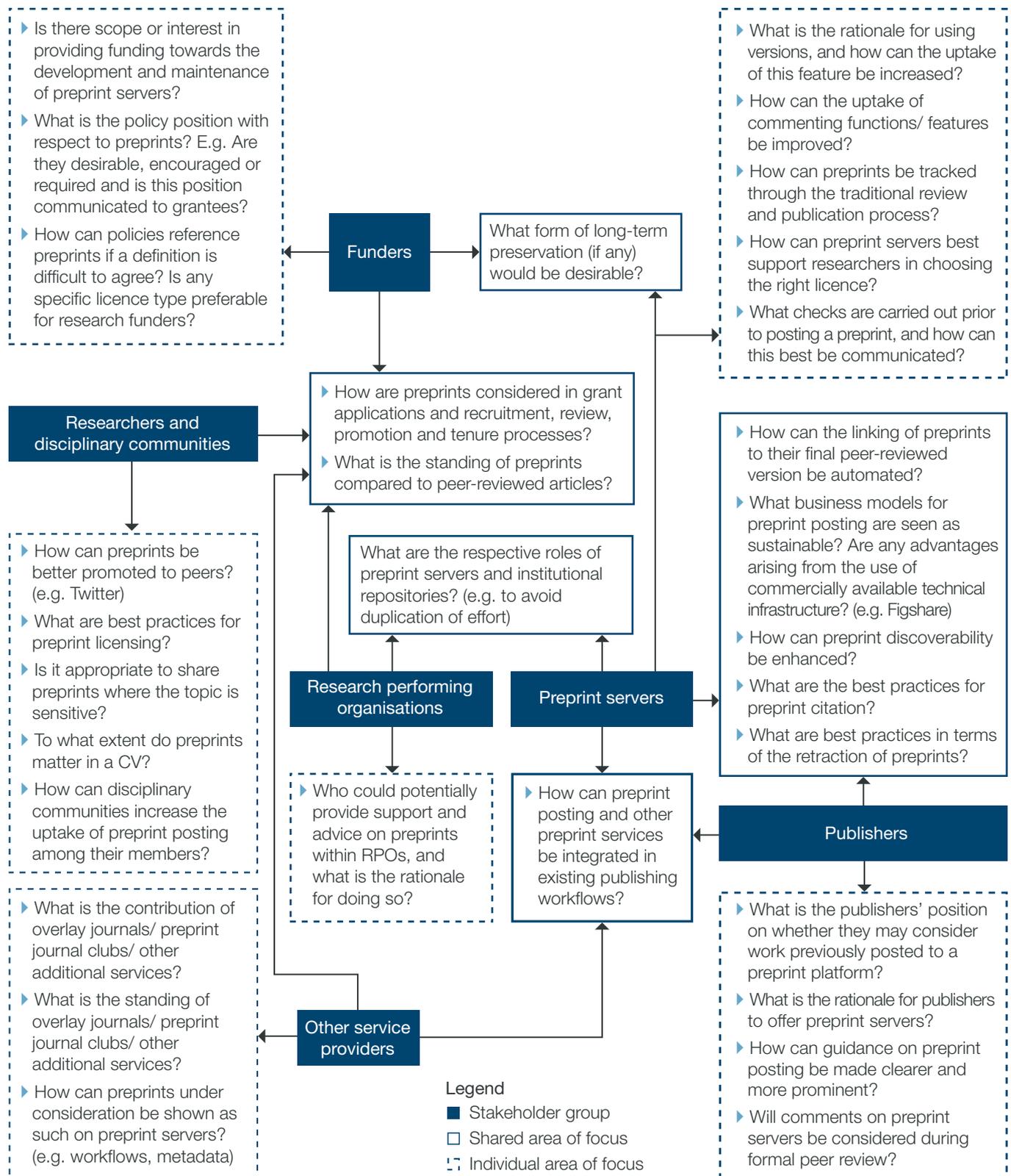
Making progress

Looking back at the findings of this study, we think there are five questions that stakeholders need to answer as a group to drive further uptake of preprints in a sustainable future landscape:

1. Is a researcher-centric model feasible in practice and would it be successful enough to drive uptake? If not would a publisher-centric model be acceptable?
2. Is control of preprints and preprint servers by commercial players (e.g. academic publishers) a deal-breaker? If so, how could national and international organisations collaborate to fund preprint servers and the cost of long-term preservation?
3. How can evidence on the potential advantages and disadvantages of preprint posting be effectively gathered?
4. What are the most suitable pathways to raise awareness and advocate for the posting of preprints?
5. What are the most effective pathways to provide researchers and other stakeholders with support to post, read and reuse preprints?

In addition, this research found that there are other questions that the various stakeholders involved will have to answer either individually or in collaboration with others. These are summarised in **Figure 7 (p.41)** and we note that, at present, they don't have clear answers. Active engagement with these questions is needed and is very likely to determine the scenario where the preprints movement will end – the higher the extent of stakeholder coordination, the more positive any outcomes will be.

Figure 7 – Questions for future developments



7. References

Knowledge Exchange study outputs

- ▶ Tennant, J., Bauin, S., James, S., & Kant, J. (2018). The evolving preprint landscape: Introductory report for the Knowledge Exchange working group on preprints. <https://doi.org/10.31222/osf.io/796tu>
- ▶ Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Practices, drivers and impediments in the use of preprints (Phase 1 report). <http://doi.org/10.5281/zenodo.2654832>
- ▶ Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Preprints and Scholarly Communication: Adoption, Practices, Drivers and Barriers. <https://doi.org/10.12688/f1000research.19619.1>
- ▶ Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Mapping of interview questions to areas of Innovation Diffusion Theory. <http://doi.org/10.5281/zenodo.3240426>
- ▶ Chiarelli, A., Kant, J. & Schmidt, B. (2019). Preprints: opportunity or challenge? (Research Information) researchinformation.info/analysis-opinion/preprints-opportunity-or-challenge
- ▶ Kant, J., T&F Group on Preprints, Chiarelli, A., Johnson, R. & Richens, E. (2019). Preprints – opportunity or challenge? (poster) <https://doi.org/10.5281/zenodo.3238499>

Other references

- 1 AMS Secretary. (2018). Backlog of Mathematics Research Journals. ams.org/journals/notices/201810/moti-p1289.pdf
- 2 Tennant, J., Bauin, S., James, S., & Kant, J. (2018). The evolving preprint landscape: Introductory report for the Knowledge Exchange working group on preprints. <https://doi.org/10.31222/osf.io/796tu>
- 3 Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Practices, drivers and impediments in the use of preprints (Phase 1 report). [https://repository.jisc.ac.uk/7381/1/Practices,_drivers_and_impediments_in_the_use_of_preprints_\(Phase_1_report\).pdf](https://repository.jisc.ac.uk/7381/1/Practices,_drivers_and_impediments_in_the_use_of_preprints_(Phase_1_report).pdf)
- 4 Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Preprints and Scholarly Communication: Adoption, Practices, Drivers and Barriers. <https://doi.org/10.12688/f1000research.19619.1>
- 5 Rogers, E. M. (2003). Diffusion of innovations. https://books.google.co.uk/books?id=9U1K5LjUOwEC&redir_esc=y
- 6 COPE. (2018). COPE Discussion Document – Preprints. https://publicationethics.org/files/u7140/COPE_Preprints_Mar18.pdf
- 7 Rittman, M. (2017). Research Preprints – Preprint servers. <https://researchpreprints.com/preprintlist>

- 8 Nguyen, T.M. (2019). Chemistry preprints pick up steam.
<https://cen.acs.org/acs-news/publishing/Chemistry-preprints-pick-steam/97/i3>
- 9 Abdill, R.J. & Blekhman, R. (2019). Meta-Research: Tracking the popularity and outcomes of all bioRxiv preprints.
<https://doi.org/10.7554/eLife.45133>
- 10 Gentil-Beccot, A., Mele, S. & Brooks, T. (2009). Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories.
<https://doi.org/10.1007/s11192-009-0111-1>
- 11 Penfold, N. (2019). Twitter (29 May 2019). <https://twitter.com/npscience/status/1133734784939769856>
- 12 O'Connell, H. (2000). Physicists Thriving with Paperless Publishing. <https://arxiv.org/abs/physics/0007040v2>
- 13 Cobb, M. (2017). The prehistory of biology preprints: A forgotten experiment from the 1960s.
<https://doi.org/10.1371/journal.pbio.2003995>
- 14 Butler, D. (2001). Los Alamos loses physics archive as preprint pioneer heads east. <https://doi.org/10.1038%2F35083708>
- 15 Karlsson, S. & Krichel, T. (1999). RePEc and S-WoPEc: Internet access to electronic preprints in Economics.
<http://openlib.org/home/krichel/papers/lindi.html>
- 16 Huisman, J. & Smits, J. (2017). Duration and quality of the peer review process: the author's perspective.
<https://doi.org/10.1007/s11192-017-2310-5>
- 17 Fraser, N., Momeni, F., Mayr, P. & Peters, I. (2019). The effect of bioRxiv preprints on citations and altmetrics.
<https://doi.org/10.1101/673665>
- 18 Neylon, C., Pattinson, D., Bilder, G. & Lin J. (2017). On the origin of nonequivalent states: How we can talk about preprints. <https://doi.org/10.12688/f1000research.11408.1>
- 19 Max Planck Society. (2003). Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.
<https://openaccess.mpg.de/Berlin-Declaration>
- 20 cOAlition S. (2019). Principles and Implementation. coalition-s.org/principles-and-implementation
- 21 Sever, R., Eisein, M. & Inglis, J. (2019). Plan U: Universal access to scientific and medical research via funder preprint mandates. planu.org

- 22 Green, T. (2019). Is open access affordable? Why current models do not work and why we need internet-era transformation of scholarly communications. <https://doi.org/10.1002/leap.1219>
- 23 Elsevier, Ipsos MORI. (2019). Research futures: Drivers and scenarios for the next decade. elsevier.com/connect/elsevier-research-futures-report
- 24 Neylon, C., Pattinson, D., Bilder, G. & Lin J. (2017). On the origin of nonequivalent states: How we can talk about preprints. <https://doi.org/10.12688/f1000research.11408.1>
- 25 European Commission. (2019). Future of Scholarly Publishing and Scholarly Communication. <http://doi.org/10.2777/836532>
- 26 Hill, S. (2018). Twitter (30 May 2018). <https://twitter.com/stevenhill/status/1001897100567891971>
- 27 ARC. (2017). Open Access Policy. arc.gov.au/policies-strategies/policy/arc-open-access-policy-version-20171
- 28 Sarabipour, S., Debat, H.J., Emmott, E., Burgess, S., Schwessinger, B. & Hensel, Z. (2018). On the value of preprints: an early career researcher perspective. <https://doi.org/10.7287/peerj.preprints.27400v1>
- 29 Bourne, P.E., Polka, J.K., Vale, R.D. & Kiley R. (2017). Ten simple rules to consider regarding preprint submission. <https://doi.org/10.1371/journal.pcbi.1005473>
- 30 Alliance nationale de recherche pour l'environnement. (2017). Preprints are a valid form of scientific communication. allenvi.fr/actualites/2017/preprints-communication-scientifique-recevable
- 31 Zimmerman, C. (2019). 5000 working paper series on RePEc: working papers are still central to economics. <https://blog.repec.org/2019/05/31/5000-working-paper-series-on-repec-working-papers-are-still-central-to-economics>
- 32 Weingart, P. & Taubert, N. (2017). The Future of Scholarly Publishing: Open Access and the Economics of Digitisation. <http://doi.org/10.5281/zenodo.1003185>
- 33 Tennant, J., Gatto, L. & Logan, C. (2018). Preprints help journalism, not hinder it. <http://doi.org/10.1038/d41586-018-06055-3>
- 34 medRxiv. (2019). Coming soon: medRxiv. <https://connect.medrxiv.org>
- 35 Altman, L.K. (1996). The Ingelfinger rule, embargoes, and journal peer review - part 1. [https://doi.org/10.1016/S0140-6736\(96\)91016-8](https://doi.org/10.1016/S0140-6736(96)91016-8)

- 36 Teixeira da Silva, J.A. & Dobránszki, J. (2019). Preprint policies among 14 academic publishers. <https://doi.org/10.1016/j.acalib.2019.02.009>
- 37 Efficiency and Standards for Article Charges – ESAC. (n.d.). Market watch. <https://esac-initiative.org/about/apcmarket>
- 38 Tennant, J. & Lomax, D. (2019). An overview of Open Access publishing in palaeontology. <https://doi.org/10.26879/968>
- 39 For more information on open peer review, see: Johnson, R., Watkinson, A. & Mabe, M. (2018). The STM Report. An overview of scientific and scholarly publishing. stm-assoc.org/2018_10_04_STM_Report_2018.pdf
- 40 Kirkham, J. & Moher, D. (2018). Who and why do researchers opt to publish in post-publication peer review platforms? - findings from a review and survey of F1000Research. <https://doi.org/10.12688/f1000research.15436.1>
- 41 Cosgrove, A. & Cheifet, B. (2018). Transparent peer review trial: the results. <https://doi.org/10.1186/s13059-018-1584-0>
- 42 Mohammadi, E., Thelwall, M., Kwasny, M. & Holmes, K.L. (2018). Academic information on Twitter: A user survey. <https://doi.org/10.1371/journal.pone.0197265>
- 43 Narock, T. W., Goldstein, E., Jackson, C. A., Bubeck, A., Enright, A., Farquharson, J. I., ... Ampuero, J. (2018). Earth Science is Ready for Preprints: The First Year of EarthArXiv. <https://doi.org/10.1029/2019EO121347>
- 44 Ke Q., Ahn Y-Y., Sugimoto CR (2017). A systematic identification and analysis of scientists on Twitter. PLoS ONE 12(4): e0175368. <https://doi.org/10.1371/journal.pone.0175368>
- 45 Narock, T. & Goldstein, E.B. (2019). Quantifying the Growth of Preprint Services Hosted by the Center for Open Science. <https://doi.org/10.3390/publications7020044>
- 46 Geltner, G & Willinsky, J. (2018). Preprint to Monograph: A Path to Travel By. guygeltner.net/blog/652018preprint-to-monograph-a-path-to-travel-by
- 47 Springer Nature, Pyne, R., Emery, C., Lucraft, M. & Pinck, A.S. (2019). The future of open access books: Findings from a global survey of academic book authors. <https://doi.org/10.6084/m9.figshare.8166599.v1>
- 48 Moore, S. (2019). Revisiting 'the 1990s debutante': scholar-led publishing and the pre-history of the open access movement. <http://dx.doi.org/10.17613/gty2-w177>
- 49 Mallapaty, S. (2019). Indian scientists launch preprint repository to boost research quality. <http://doi.org/10.1038/d41586-019-01082-0>

- 50 Eger, T., and Scheufen, M. (2018). The Economics of Open Access: On the Future of Academic Publishing. jipitec.eu/issues/jipitec-9-3-2018/4812
- 51 Preprints Editorial Office – Preprints.org. (2019). Update of Preprint DOI registration. preprints.org/announcement/show/33
- 52 Landhuis, E. (2016). Scientific literature: Information overload. <http://doi.org/10.1038/nj7612-457a>
- 53 Neylon, C. (2011). It's not filter failure, it's a discovery deficit. <http://doi.org/10.1629/2421>
- 54 Fenn, J. & Raskino, M. (2008). Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time. worldcat.org/title/mastering-the-hype-cycle-how-to-choose-the-right-innovation-at-the-right-time/oclc/213312226
- 55 Joseph, H. (2019). Investing in Open Scholarly Infrastructure: a Community Opportunity. <https://sparcopen.org/news/2019/investing-in-open-scholarly-infrastructure-a-community-opportunity>
- 56 Larivière, V., Haustein, S., Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era. <https://doi.org/10.1371/journal.pone.0127502>
- 57 Wouters, P., Sugimoto, C.R., Larivière, V., McVeigh, M.E., Pulverer, B., de Rijcke, S. & Waltman, L. (2019). Rethinking impact factors: better ways to judge a journal. <http://doi.org/10.1038/d41586-019-01643-3>
- 58 Green, T. (2019). Are we being wilfully blind about the transformation that's needed in scholarly publishing? <https://medium.com/@TobyABGreen/are-we-being-wilfully-blind-about-the-transformation-thats-needed-in-scholarly-publishing-d0bfb61d1f05>
- 59 Ball, P. (2015). Leading mathematician launches arXiv 'overlay' journal. <https://doi.org/10.1038/nature.2015.18351>
- 60 Ball, P. (2015). The journal that publishes no papers. nature.com/polopoly_fs/1.18444!/menu/main/topColumns/topLeftColumn/pdf/526146a.pdf
- 61 Peer Community In. (2019). PCI economic model. <https://peercommunityin.org/2019/05/29/pci-economic-model>
- 62 STM Publishing. (2019). The Company of Biologists launches preLists. stm-publishing.com/the-company-of-biologists-launches-prelists
- 63 Secker, J. & Morrison, C. (2018). Copyright literacy in the UK: Understanding library and information professionals' experiences of copyright. <http://openaccess.city.ac.uk/id/eprint/20082>

- 64 Abdill, R.J. & Blekhman, R. (2019). Meta-Research: Tracking the popularity and outcomes of all bioRxiv preprints. <https://doi.org/10.7554/eLife.45133>
- 65 Himmelstein, D. (2016). The licensing of bioRxiv preprints. <https://blog.dhimmel.com/biorxiv-licenses>
- 66 Simboli, B. (2019). arXiv and the Symbiosis of Physics Preprints and Journal Review Articles. <https://arxiv.org/abs/1904.01470v2>
- 67 Balaji, B.P. & Dhanamjaya, M. (2019). Preprints in Scholarly Communication: Re-Imagining Metrics and Infrastructures. <https://doi.org/10.3390/publications7010006>
- 68 Fraser, N., Momeni, F., Mayr, P. & Peters, I. (2019). The effect of bioRxiv preprints on citations and altmetrics. <https://doi.org/10.1101/673665>
- 69 DORA. (2018). Preprints in Academic Hiring. <https://sfdora.org/2018/08/14/preprints-in-academic-hiring>
- 70 Pudovkin, A.I. (2018). Comments on the Use of the Journal Impact Factor for Assessing the Research Contributions of Individual Authors. <https://doi.org/10.3389/frma.2018.00002>
- 71 Wang, L. & Zhan, Y. (2019). A conceptual peer review model for arXiv and other preprint databases. <https://doi.org/10.1002/leap.1229>
- 72 Vale, R.D. (2015). Accelerating scientific publication in biology. <https://doi.org/10.1073/pnas.1511912112>
- 73 Kotter, J.P. (2007). Winning at change. providersedge.com/ehdocs/transformation_articles/WINNING_AT_CHANGE.pdf
- 74 Kant, J., T&F Group on Preprints, Chiarelli, A., Johnson, R. & Richens, E. (2019). Preprints – opportunity or challenge? <https://doi.org/10.5281/zenodo.3238499>
- 75 Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Mapping of interview questions to areas of Innovation Diffusion Theory. <http://doi.org/10.5281/zenodo.3240426>
- 76 Hoyt, J. (2019). PeerJ Preprints to stop accepting new preprints Sep 30th 2019. peerj.com/blog/post/115284881747/peerj-preprints-to-stop-accepting-new-preprints-sep-30-2019

Appendix A – Acknowledgements and participants

Table A1: Knowledge exchange task and finish group on preprints

Name	Affiliation	Role
Andy Turner	University of Leeds	Researcher
Angela Holzer	DFG	Programme Officer
Bas Cordewener	Knowledge Exchange	Knowledge Exchange Coordinator
Birgit Schmidt	Göttingen State and University Library	Head of Knowledge Commons
Frank Manista	Jisc	European open science manager
Gernot Deinzer	Regensburg University	Specialist in Mathematics and Physics
Jeroen Sondervan	Utrecht University Library	Project Leader Open Access
John Doove	SURF	Program Manager Open Access
Jon Tennant	IGDORE; Center for Research and Interdisciplinarity, University of Paris	Researcher
Juliane Kant	DFG	Programme Officer
Karin van Grieken	SURF	Senior Project Manager
Neil Jacobs	Jisc	Head of Scholarly Communications Support
Olivier Le Gall	INRA	Researcher
Sarah James	Knowledge Exchange	Knowledge Exchange Project and Administration Officer
Saskia Woutersen-Windhouwer	Leiden University	Digital Scholarship Librarian
	Netherlands Institute of Ecology (NIOO)	Head of Library & Information Services
Serge Bauin	Centre National de la Recherche Scientifique (CNRS)	Special Adviser on Open Science

Table A2: Interviewees

Name	Affiliation	Role
Alice Lebreton	Ecole Normale Supérieure de Paris	Researcher
Angela Holzer	DFG	Research Funder
Antica Culina	Netherlands Institute of Ecology (NIOO)	Researcher
Antonio de la Vega de Leon	University of Sheffield	Researcher
Arjan Schalken	Vrije Universiteit van Amsterdam	RPO
Benjamin Brown	PsyArXiv Steering Committee	Preprint server
Birgit Schmidt	Göttingen State and University Library	RPO
Charlotte Wien	University of Southern Denmark	RPO
Daniël Lakens	TU Eindhoven	Researcher
Danny Kingsley	Cambridge University Library	RPO
Felix Schonbrodt	LMU München	Researcher
Jan Jensen	The University of Copenhagen	Researcher
Jan Philip Solovej	The University of Copenhagen	Research Funder
John Inglis	Cold Spring Harbor Laboratory Press (bioRxiv)	Preprint server
Julie Aspden	University of Leeds	Researcher
Jyrki Hakapää	Academy of Finland	Research Funder
Karolina Urbanska	Université Clermont Auvergne	Researcher
Kathleen Fitzpatrick	Humanities Commons	Other service providers
Katie Drax	University of Bristol	Researcher
Laurent Romary	Inria	Other service providers
Liz Allen	F1000Research	Other service providers
Marleen Bink	ZonMW	Research Funder
Martina Knoop	Centre National de la Recherche Scientifique (CNRS)	RPO
Martyn Rittman	Preprints.org (MDPI)	Preprint server

Table A2: Interviewees (continued)

Name	Affiliation	Role
Michael Fischer	University of Bremen	Researcher
Michael Markie	F1000Research	Other service providers
Mikael Laakso	Hanken School of Economics	RPO
Natalia Soshkinova	Institute of Molecular Biology	Researcher
Olaf Siegert	ZBW - Leibniz Information Centre for Economics	RPO
Peter Murray-Rust	ContentMine	Researcher
Richard Tunney	Aston University	Researcher
Robert Kiley	Wellcome Trust	Research Funder
Sam Smith	University of Leeds	Researcher
Serge Bauin	Centre National de la Recherche Scientifique (CNRS)	RPO
Stephanie Dawson	ScienceOpen	Preprint server
Thomas Battram	University of Bristol	Researcher
Thomas Lösch	Universität Bamberg	Researcher
Zoe Ancion	DIPF Leibniz Institute for Research and Information in Education	Research Funder

Appendix B – Interview questions

This appendix includes a list of interview questions used for the purposes of this study. Questions were asked based on stakeholder groups and time availability, meaning that not all questions were asked to all interviewees. A more detailed version of the table below is available on Zenodo.⁷⁵

Legend:

- ▶ Research Funder: RF
- ▶ Research Performing Organisation: RPO
- ▶ Preprint Server: PS
- ▶ Other Service Provider (e.g. other publishing platforms including pre-refereed content): OSP
- ▶ Researcher: Re

Interview question	Stakeholder group(s)
Does the preprint server you represent make any particular promotion efforts to encourage researchers to post preprints and other stakeholders to accept them as valid research outputs?	PS
Does your organisation consider both preprints and journal articles when evaluating research proposals?	RF
Is your organisation encouraging the posting of preprints in any way?	RF, RPO
How do you believe preprint servers should be funded?	RF, RPO, Re
In your experience, what approach do funders and national research evaluation exercises (if applicable) take to preprints? What impact, if any, do you think these approaches have on researchers' attitudes to preprints?	RPO
Is it acceptable to post a preprint when the topic is sensitive? (e.g. human health)	PS
What business model are you currently using?	PS
What is your approach to long-term preservation and how does it compare to other repository services? (e.g. Zenodo, figshare)	PS
What problems does the use of preprints cause?	Re
Is the preprint server you represent concerned about information overload, i.e. generating too much information for researchers to easily digest?	PS
To what extent is the proposition of preprints clear to you?	Re

Footnotes

⁷⁵ Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Mapping of interview questions to areas of Innovation Diffusion Theory. <http://doi.org/10.5281/zenodo.3240426>

Interview question	Stakeholder group(s)
What changes to the scholarly communication infrastructure are needed to support the uptake of preprints?	RF, RPO
What difficulties (if any) slowed down or prevented your use of preprint servers?	Re
What is the value for researchers, research organisations and funders when it comes to overlay journals?	OSP
How can preprint servers minimise the risks connected with the misuse of non-peer reviewed research?	Re
How do overlay services contribute to ensuring preprints are based on sound science?	OSP
How does the preprint server you represent support the responsible use of preprints by third parties?	PS
Is your organisation considering preprints in HR processes? (e.g. for career advancement)	RPO
Preprints have a Digital Object Identifier (a type of permanent link) and can be cited. What is your position on the practice of citing preprints?	Re
What approach do other researchers in your discipline take to the posting of preprints? How, if at all, do you think their motivations differ from yours?	Re
What are the implications of preprints for the scientific quality of research outputs?	RF, RPO
What do you feel is the level of awareness of preprints among researchers, particularly in [discipline]?	PS, OSP
What is the role of preprint servers in ensuring preprints are based on sound science?	PS
What process (if any) do you have to ensure preprints are based on sound science?	PS
Can posting preprints be used as an alternative to publishing open access (e.g. Gold or Hybrid) peer-reviewed research?	RF, RPO
What is the value of services like overlay journals and review platforms from a funder's perspective?	RF
What relationship, if any, exists between preprints and open access?	PS, OSP
Preprint servers allow readers to either write comments or email feedback to authors. In your experience, do comments and feedback improve the quality of a manuscript? How does this process compare with peer-review?	Re
To what extent do people use comments in the preprint server you represent, and what are the practical differences between peer-review and comments?	PS
What are the key benefits of preprints for the research community?	RF, RPO
What benefits do preprints have for research?	Re

Interview question	Stakeholder group(s)
What is the value of preprints for researchers, research organisations and funders?	PS, OSP
What value do preprint servers create for researchers, research organisations and funders?	PS
How familiar are you with preprints and what interactions do you have with them in your current role?	RF, RPO
How often (if at all) do you interact with preprint servers in your everyday role?	Re
Was this your first experience with preprint servers or did you encounter them in other cases?	Re
What is your experience with preprints?	PS, OSP
How do you feel research organisations and research funders are dealing with the increasing use of preprints?	Re
What is the origin of the preprint server you represent, and how has it developed over time?	PS
What key initiatives in the area of preprints are you aware of?	PS, OSP
What role can Knowledge Exchange play in addressing the gaps and issues in the area of preprints?	RF, RPO, PS, OSP

Appendix C – Overview of preprint servers

The table below builds on work by Martin Rittman⁷ and has been edited and updated in some respects. We note that there are gaps in some areas as information availability is low for some of the servers or platforms listed. When the column “Software/ Technological solution” states “unknown”, this means that the information was not available: in most cases, this would indicate the use of proprietary or ad-hoc solutions as opposed to mainstream software packages.

Note: Appendix C was created based on information available online and only includes currently active platforms that could be identified within the timeframe and scope of this project. While the information is considered to be correct at the date of publication, we cannot guarantee its accuracy.

Footnotes

- ⁷ Rittman, M. (2017). Research Preprints – Preprint servers. <https://researchpreprints.com/preprintlist>

Server or platform (linked)	Discipline (self-reported)	Managed by	Software/ Technological solution	Founding date	Notes
arXiv https://en.wikipedia.org/wiki/ArXiv	Science (Multidisciplinary, but known for mathematics and physics)	Cornell University	Unknown	1991	
Mathematical Physics Preprint Archive https://web.ma.utexas.edu/mp_arc/	Mathematical and Physics	Department of Mathematics at University of Texas at Austin	Unknown	1991	
CERN Document Server https://cds.cern.ch/collection/Preprints	Particle Physics	CERN	Invenio digital library framework	1993	
Electronic Colloquium on Computational Complexity https://eccc.weizmann.ac.il/eccc	Computer Science	Weizmann Institute of Science	Unknown	1994	
SSRN ssrn.com/index.cfm/en	Generalist	Elsevier	Unknown	1994	
Cogprints http://cogprints.org	Cognitive Sciences	University of Southampton (Permanently archived)	EPrints	1997	
RePEc http://repec.org	Economics	Volunteers	Unknown	1997	
Cryptography ePrint Archive https://eprint.iacr.org	Cryptography	International Association for Cryptologic Research	EPrints	2000	
HaL https://hal.archives-ouvertes.fr	Generalist	Centre National de la Recherche Scientifique (CNRS), INRIA, INRA, Université de Lyon	Unknown	2001	This is not a preprint server per se, but preprints can be deposited.

Server or platform (linked)	Discipline (self-reported)	Managed by	Software/ Technological solution	Founding date	Notes
PhilSci-Archive http://philsci-archive.pitt.edu	Philosophy of Science	Archive Board	EPrints	2001	
Eprints in Library and Information Science (E-LIS) http://eprints.rclis.org	Library and Information Science	E-LIS volunteers Hosted by "Federico II" University of Naples (Italy)	EPrints	2003	
Munich Personal RePEc Archive (MPRA) https://mpra.ub.uni-muenchen.de	Economics	Munich University Library	EPrints	2006	
ECONSTOR econstor.eu/dspace/	Economics and Business Studies	ZBW, Leibniz Information Centre for Economics	DSpace	2008	
ResearchGate researchgate.net	Generalist	ResearchGate	Unknown	2008	This is not a preprint server per se, but preprints can be deposited.
SSOAR gesis.org/en/ssoar/home	Social Sciences	GESIS	DSpace	2008	
viXra http://vixra.org	Generalist	Volunteers	EPrints	2009	
figshare https://figshare.com	Generalist	figshare	figshare	2011	This is not a preprint server per se, but preprints can be deposited.
Authorea authorea.com	Generalist	Authorea	Ruby on Rails and Git	2012	
bioRxiv biorxiv.org	Life Sciences	Cold Spring Harbor Laboratory (CSHL)	Drupal	2013	
F1000Research https://f1000research.com	Life sciences	Faculty of 1000	Unknown	2013	Preprints are part of the journal publishing workflow.
PeerJ Preprints https://peerj.com/preprints	Life sciences	PeerJ	Unknown	2013	PeerJ preprints is scheduled to stop accepting new preprints on Sep 30th 2019. ⁷⁶

Footnotes

⁷⁶ Hoyt, J. (2019). PeerJ Preprints to stop accepting new preprints Sep 30th 2019.

peerj.com/blog/post/115284881747/peerj-preprints-to-stop-accepting-new-preprints-sep-30-2019

Server or platform (linked)	Discipline (self-reported)	Managed by	Software/ Technological solution	Founding date	Notes
Zenodo https://zenodo.org	Generalist	Open Aire/CERN	CERN Data Centre and the Invenio digital library framework	2013	This is not a preprint server per se, but preprints can be deposited.
CORE repository https://mla.hcommons.org/core	Humanities	Modern Languages Associate (MLA) and the Center for Digital Research and Scholarship at Columbia University	Unknown	2015	
JMIR Preprints https://preprints.jmir.org	Medicine	Journal of Medical Internet Research	Unknown	2015	
Preprints.org preprints.org	Generalist	MDPI	Unknown	2015	
ChinaXiv http://chinaxiv.org	Generalist – China	Chinese Academy of Sciences	Unknown	2016	
engrXiv https://engrxiv.org	Engineering	Steering committee	Open Science Framework	2016	
OSF Preprints https://osf.io/preprints	Generalist	Open Science Framework	Open Science Framework	2016	
PsyArXiv https://psyarxiv.com	Psychology	Society for the Improvement of Psychological Science (SIPS)	Open Science Framework	2016	
SocArXiv https://osf.io/preprints/socarxiv	Social Sciences	Housed at the University of Maryland and directed by a steering committee	Open Science Framework	2016	
Wellcome Open Research https://wellcomeopenresearch.org	Medicine and health sciences	Wellcome Trust	F1000	2016	
AgriXiv https://agrixiv.org	Agriculture and Allied Sciences	OAIIndia	Open Science Framework	2017	
ChemRxiv https://chemrxiv.org	Chemistry	American Chemical Society	figshare	2017	
EarthArXiv https://eartharxiv.org	Earth Sciences	Advisory council	Open Science Framework	2017	
ESSOAr essoar.org	Geoscience	The American Geophysical Union and Atypion with support from Wiley	Literatum	2017	
Gates Open Research https://gatesopenresearch.org	Generalist	Bill & Melinda Gates Foundation	F1000	2017	

Server or platform (linked)	Discipline (self-reported)	Managed by	Software/ Technological solution	Founding date	Notes
INA-Rxiv https://osf.io/preprints/inarxiv	Generalist - Indonesia	Steering Committee	Open Science Framework	2017	
LawArXiv http://lawarxiv.info	Legal scholarship	Advisory Boards	Open Science Framework	2017	
LIS Scholarship Archive https://osf.io/preprints/lissa	Library and Information Science	LISSA Advisory board	Open Science Framework	2017	
MarXiv https://marxiv.org	Ocean and Marine-climate Sciences	Advisory board and OCTO	Open Science Framework	2017	
MetaArXiv https://osf.io/preprints/metaarxiv	Social Sciences	The Berkeley Initiative for Transparency in the Social Sciences (BITSS) and Steering Committee	Open Science Framework	2017	
MindRxiv https://mindrxiv.org	Mind and contemplative practices	Mind & Life Institute	Open Science Framework	2017	
NutriXiv https://osf.io/preprints/nutrixiv	Nutritional Sciences	Open Science Framework	Open Science Framework	2017	
PaleorXiv https://paleorxiv.org	Paleontology	Steering Committee	Open Science Framework	2017	
PhilArchive https://philarchive.org	Philosophy	PhilPapers Foundation	PhilPapers	2017	
SportRxiv http://sportrxiv.org	Sport	Society for Transparency Openness and Replication in Kinesiology (STORK)	Open Science Framework	2017	
Therapoid https://therapoid.net/en/preprint	Therapeutics	Open Therapeutics	Unknown	2017	
Thesis Commons https://thesiscommons.org	Theses and dissertations	Steering Committee	Open Science Framework	2017	
WikiJournalPreprints https://en.wikiversity.org/wiki/WikiJournal_Preprints	Generalist	WikiJournal User Group	MediaWiki	2017	
Advance https://advance.sagepub.com	Humanities and Social Sciences	SAGE	figshare	2018	
AfricArxiv https://osf.io/preprints/africarxiv	Generalist - Africa	Steering Committee	Open Science Framework	2018	

Server or platform (linked)	Discipline (self-reported)	Managed by	Software/ Technological solution	Founding date	Notes
Arabixiv https://arabixiv.org	Generalist - Arabic	Advisory Committee	Open Science Framework	2018	
ECSarXiv https://ecsarxiv.org	Electrochemistry, solid state science and technology	The Electrochemistry Society	Open Science Framework	2018	
FocUS Archive https://osf.io/preprints/focusarchive	Ultrasound	Steering Committee	Open Science Framework	2018	
FrenXiv https://frenxiv.org	Generalist - French	Steering Committee	Open Science Framework	2018	
APSA Preprints [link not available – launches August 2019]	Political Sciences	The American Political Science Association and Cambridge University Press	Unknown	2019	
BodoArXiv https://osf.io/preprints/bodoarxiv	Medieval Studies	ScholarlyHub and Committee	Open Science Framework	2019	
EcoEvoRxiv https://ecoevorxiv.org	Ecology, evolution and conservation	Steering Committee	Open Science Framework	2019	
IndiaRxiv http://indiarxiv.in	Generalist - India	Open Access India	Open Science Framework	2019	
MediArXiv https://mediarxiv.org	Media, Film and Communication Studies	Steering Committee	Open Science Framework	2019	
MedRxiv medrxiv.org	Medicine	Cold Spring Harbor Laboratory (CSHL)	Drupal	2019	
LingBuzz https://ling.auf.net	Linguistics	Michal Starke and University of Tromsø	Unknown	Unknown	
NBER Working Papers nber.org/papers	Economics	The National Bureau of Economic Research	Unknown	Unknown	

Knowledge Exchange Office
C/O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info