# Building trust in preprints: recommendations for servers and other stakeholders

Jeffrey Beck[1], Christine A Ferguson[2], Kathryn Funk[1], Brooks Hanson[3], Melissa Harrison[4], Michele Ide-Smith[2], Rachael Lammey[5], Maria Levchenko[2], Alex Mendonça[6], Michael Parkin[2], Naomi C Penfold[7], Nici Pfeiffer[8], Jessica K Polka[7]*, Iratxe Puebla[7], Oya Y Rieger[9], Martyn Rittman[5], Richard Sever[10], Sowmya Swaminathan[11]

1. National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, USA
2. EMBL-EBI, Hinxton, UK
3. AGU/ESSOAr, Washington DC, USA
4. eLife, Cambridge, UK
5. Crossref, Lynnfield, USA
6. SciELO, São Paulo, Brazil
7. ASAPbio, San Francisco, USA
8. Center for Open Science, Charlottesville, USA
9. Ithaka S+R, New York, USA
10. Cold Spring Harbor Laboratory, Cold Spring Harbor, USA
11. Nature Research, Springer Nature, London, UK

* For correspondence: jessica.polka@asapbio.org

## Abstract

On January 20 and 21, 2020, ASAPbio, in collaboration with EMBL-EBI and Ithaka S+R, convened over 30 representatives from academia, preprint servers, publishers, funders, and standards, indexing and metadata infrastructure organisations at EMBL-EBI (Hinxton, UK) to develop a series of recommendations for best practices for posting and linking of preprints in the life sciences and ideally the broader research community. We hope that these recommendations offer guidance for new preprint platforms and projects looking to enact best practices and ultimately serve to improve the experience of using preprints for all.

# Table of Contents

# Executive summary: best practice recommendations

*The following is a summary of major recommendations stemming from the #bioPreprints2020 meeting held on January 20 and 21, 2020 at EMBL-EBI (Hinxton, UK).*

## Preprint servers

- Create a PID for each version; link all versions together.
- Allow authors to indicate data availability (and provide links to data stored elsewhere) when submitting a preprint, make this information discoverable by humans and machines, and transfer this information to journals.
- Adopt standardized definitions of withdrawal (original preprint remains) and removal (original preprint removed). Clearly display withdrawal and removal notices and notify journals if a preprint has been withdrawn or removed after journal publication.
- Create recommended metadata (see table 2 for details).

Table 1. Summary of recommended metadata

| Essential | Desired | Optional |
|---|---|---|
| Article identifier<br>Posted date<br>Repository name<br>Article title<br>Author names<br>Withdrawal status<br>File modification status<br>IsVersionOf<br>License | Repository owner<br>Abstract<br>Author affiliation(s)<br>Version ID | Author roles<br>Submitter<br>Affiliation ID(s)<br>Author ID (s)<br>Funder ID<br>Grant ID<br>Journal version<br>Withdrawal/removed reason<br>References<br>Article type<br>Language<br>Keywords and terms<br>Changelog<br>Available resource statement<br>Available resource type<br>Available resource links<br>COI statement<br>Quality control information<br>Discipline |

## Peer review services

- Create and manage metadata for review events; make this available to preprint servers and others.

## Journals

- When transferring a submission to a server, collect data availability statements to be transferred to the server and work toward interoperability of data availability statements.
- Notify preprint servers in the case of article withdrawal/removals, for example via Crossref Crossmark.

# Introduction

Preprints allow researchers to disseminate their work when they are ready for it to be shared with their community. Supporting the role of preprints in the research process requires a consideration of cultural elements related to awareness, perceived value, and potential reservations in some communities and a technical infrastructure that facilitates the visibility of preprints and their value within the research cycle.

For example, authors need to trust that their preprint will be seen by potentially interested readers; for this, preprints must be readily discoverable and indexable by commonly used search tools. Readers need to be able to interrogate the claims of a preprint; this requires links to underlying data be available where applicable. They also need to know if the version of a preprint they are looking at has been updated, if a preprint was withdrawn or removed, or if there is relevant commentary or peer review available.

At the core of these expectations is a need for robust metadata to support discoverability and use of preprints by the community.

On January 20 and 21, 2020, ASAPbio, in collaboration with EMBL-EBI and Ithaka S+R, convened over 30 representatives from academia, preprint servers, publishers, funders, and standards, indexing and metadata infrastructure organisations at EMBL-EBI (Hinxton, UK) to develop a series of recommendations for best practices for posting and linking of preprints in the life sciences and ideally the broader research community.

The general consensus was that preprint servers and other infrastructure providers should work together to enable and empower researchers to communicate their work with greater speed and transparency and to ensure that preprints are treated as legitimate research outputs. There is an opportunity to develop practices and processes that provide a foundation to build trust in preprints, now and as they develop further.

After the meeting, attendees formed working groups to generate the recommendations below. We recognize that building trust in preprints will require both an understanding of the cultures and needs of diverse research communities and stakeholders and the technical aspects and workflows that allow discoverability and (re)use of preprints. Many of the working groups focused on areas related to technical aspects of the handling of preprints records and their metadata, so this constitutes an important part of the recommendations outlined. One of the working groups discussed engagement with stakeholders, and we also include a summary of the activities so far and those planned for the coming months.

We hope that these recommendations offer guidance for new preprint platforms and projects looking to enact best practices and ultimately serve to improve the experience of using preprints for all.

# Prioritizing metadata

Preprints will be most useful to research communities if they can be discovered by search tools and indexing services that researchers use to engage with the literature. Presently, preprint servers differ in the metadata they create, surface, and send to various services (eg Crossref); some servers currently do not send all the metadata that indexing and archiving resources, such as Europe PMC and PubMed Central (PMC), need to include preprints in their databases.

Many preprint servers operate with limited resources, and where they are not integrated with journal submission systems, authors expect a streamlined deposition process. Therefore, proposals for essential preprint metadata must be sensitive to the challenges servers experience in modifying ingestion processes or building additional infrastructure.

## Preprint vs journal article metadata

The recommendations proposed would make preprint metadata more comprehensive than metadata currently required for journal articles, enabling evolution of a scholarly communication ecosystem in which there is better linking and description of preprints.

**Table 2. Prioritized metadata fields for preprints**

| **Element** | **Description** | **Suggested Priority (essential, desired, optional)** | **Required by Europe PMC?** | **Required by PMC /PubMed?** | **Required by Crossref?** |
|---|---|---|---|---|---|
| Article identifier | A unique identifier for each article defined internally by the preprint server | Essential | Yes | Yes | Yes |
| Posted date | Date of first public posting on preprint server | Essential | Yes | Yes | Yes |
| Submission date | Date submitted to preprint server | Optional | No | No | No |
| Repository name | Name of preprint server (e.g. bioRxiv) | Essential | Yes | Yes | Yes |
| Repository owner | Name of organisation(s) or group(s) that handles preprints posting and sets policies for screening, withdrawal, etc. (e.g. Cold Spring Harbor Laboratory) | Desired | Yes | Yes | Yes |
| Article title | Title of the preprint as displayed in the posted version | Essential | Yes | Yes | Yes |
| Abstract | Abstract (or summary) of the preprint as displayed in the posted version | Desired | Yes | Yes | No |
| Author names | Names of all authors, as displayed in the posted preprint. Authorship should follow conventions used for journal articles. | Essential | Yes | Yes | Yes |
| Author roles | A description of the contribution of each author, using CRediT taxonomy (https://casrai.org/credit/) | Optional | No | No | No |
| Submitter | Author or other party that has logged in to submit the preprint | Optional | No | No | No |

| | | | | | |
|---|---|---|---|---|---|
| Author affiliation(s) | Names and addresses of institutions to which the authors are affiliated, as displayed in the posted preprint. May include “No affiliation” if the author asserts the work was not conducted in connection with an institution.. | Desired | No | No | No |
| Affiliation ID(s) | A unique identifier that identifies the authors' affiliations, e.g. ROR (https://ror.org) | Optional | No | No | No |
| Author ID (s) | A unique identifier for each author, e.g. ORCID (https://orcid.org) | Optional | No | No | No |
| Funder ID | Unique identifier for any funding of the work in the preprint, e.g., Open Funder Registry (https://www.crossref.org/services/funder-registry/) | Optional | No | No | No |
| Grant ID | Unique identifier for funding grants used to support the work in the preprint | Optional | No | No | No |
| Version ID | A number or identifier that differentiates the preprint instance from other instances (or versions) of the same document posted on the same platform; these may or may not have different unique identifiers | Desired | Yes (often inferred from article identifier) | Yes | No |
| Journal version | DOI of journal published article (isPreprintOf) | Optional | No | No | Yes (if published) |
| Withdrawal status | Options include live and withdrawn | Essential | No (but would be) | No | No |
| File modification status | Used to indicate that the original files posted on the server have been changed from the original (e.g. in the context of a removal or partial removal). Not to be used when a new version has been posted, but original files from earlier versions remain. | Essential | No (but would be) | No | No |
| Withdrawal /removed reason | Comment describing reason for withdrawal/removal | Optional | No | No | No |
| IsVersionOf | Unique identifiers of other preprint versions of the same document (but not journal article versions) | Essential | No | No | Yes (if published) |
| References | Open, machine-readable references | Optional | No | No | No |

| | | | | | |
|---|---|---|---|---|---|
| Article type | Specify whether preprint is an article, review, poster, short report, etc; use of standard ontologies, such as those defined by JATS (https://jats.nlm.nih.gov) are recommended | Optional | No (but would be for repositories with multiple types) | Yes | No |
| Language | The language in which the preprint is written | Optional | No | Yes | No |
| Keywords and terms | Keywords related to the theme of the preprint | Optional | No | No | No |
| License | Name or brief description of the license under which the preprint is released, e.g. 'public domain', 'Creative Commons CC BY 4.0' or 'All rights reserved', including a link to the full license details | Essential | No | Yes | No |
| Changelog | Brief description of what has changed since the previous version | Optional | No | No | No |
| Available resource statement | Describes if a resource is publicly available. Options are "Yes," "No," or "N/A" | Optional | No | No | No |
| Available resource type | Options for data, code/software, preregistration, methods, analysis, other | Optional | No | No | No |
| Available resource links | A link to any resource used by the preprint and deposited in a repository | Optional | No | No | No |
| Competing interest statement | Describes competing interests held by the authors related to the preprint | Optional | No | No | No |
| Ethics statements | Declarations of compliance with norms, legislation, or regulatory bodies; statements on data handling,or patient consent | Optional | No | No | No |
| Quality control information | Descriptions of screening checks carried out on this preprint by the server | Optional | No (inferred from repository) | No | No |
| Discipline | Needed for discoverability across services | Optional | No | No | No |

## Proposed metadata priorities

Rather than an exhaustive discussion, we highlight fields that generated avid discussion, whether included in the final recommendations or not.

### Repository name/owner

For some preprint servers, there are several different organisations involved, with governance and ownership separated or shared. Examples include SSRN, where preprints are submitted to First Look platforms controlled by individual journals, and OSF, which is a platform that hosts a number of preprint servers. Discussion in the working group looked at how to differentiate and identify owners, governance, and hosting platforms; however input from preprint server representatives indicated that these differences were not very important to them and that identification of the preprint server name was most critical. At the same time, some organisations require both a repository name and owner to be provided in metadata, such as PMC. It was decided that the name and owner would be included as separate fields, and that at least one of these should be completed. However, for completeness and to ensure indexing we recommend that both are added by preprint servers to metadata.

### Article type

Some discussion centred around the article type and the possibility of including outputs such as posters, short reports, abstracts, videos, and so on. It was decided that the focus of these recommendations should remain on preprints as potential research articles, i.e., predominantly articles and reviews. Other types require additional flexibility and non-specificity in the schema that complicate the goal of providing high quality metadata about preprints and are thus out of scope of these recommendations (for example, meeting/conference title, date of presentation, etc).

Future schema could look at versions of the other types of output. The final recommendation was to include article type as an optional field, with the intention that the types used would be relatively narrow in scope, although without using a predefined list.

### Object state

Some feedback requested adding to the schema an object state with respect to peer review, for example, whether it was submitted for peer review, a revised version, or accepted for publication. The argument for this is that these data are collected by a number of preprint servers (eg ESSOAr) and could be useful to readers. The argument against is that it is difficult for the preprint server to verify the state without contacting the publisher, and they would not be able to add a value in cases where the authors didn’t report it. The status could also change at any time and may be out of date when someone reads the metadata. Further, it would be a non-trivial task for preprint servers to keep the status current. Given the difficulties, it was decided not to include this field in the metadata recommendations; instead, the data could be captured as event data.

### Withdrawal and removal

A further discussion about withdrawal and removal of preprints is included later in the report, and we recognise that this is not an area with settled practices. From the metadata point of view, the fields recommended to be included were an essential field stating whether the preprint is live or withdrawn(see definitions in the section below entitled “Defining preprint withdrawal & removal”); a field indicating that original files have been taken down or modified (in the case of a removal or partial removal); and a further optional field of free text stating the reason for either withdrawl or removal. These options were included to reflect the recommendations of the working group that look more closely at removal of preprints. Given that servers employ a variety of approaches for handling versions, there may be challenges in associating these tags with the appropriate preprint version, and further discussion on their application in those cases is warranted.

### Versions

A full discussion on versioning is available in the next section, however for the metadata recommendations it is strongly recommended to give a unique ID to each version in order that they can be easily differentiated. A further optional field allows for a changelog to give a brief description of changes between versions. The isVersionOf field should be used to provide identifiers of different versions of the same preprint. Peer-reviewed journal article versions of the preprint should be linked using a separate field (journal version). Various methods of versioning are in use by preprint servers and the recommended field should be flexible enough to cope with the vast majority of cases. In discussions, it was felt important to differentiate peer-reviewed journal articles as there are many use cases where this differentiation is important.

### Availability statements

In the discussions, it was felt that there would be great value in including fields that allow links to related resources, such as code, data, or research objects. While we accept that these are not currently widely reported by authors or sought by preprint servers, there is a growing interest in including them for reasons of transparency and reproducibility. By adding recommendations for these fields, we aim to build capacity and standard ways of reporting to enable a transparent and comprehensive record and tracking of research outputs. The optional fields ‘available resource type’ and ‘links to available repositories’ allow a resource type to be defined and linked via the preprint metadata. A list of resource types has not been included and was felt to be beyond the scope of this group; however, we note that the Center for Open Science is working on a list that will be forthcoming soon.

### Quality control information

A field unique to preprints and distinct from research articles includes quality control information. There is currently no standardization about control checks for preprints posted online and no common language to describe what checks are carried out. However, it was felt that it is important to report the checks that the preprint has undergone and that this would be useful to readers, especially where preprints are being relied on for further research, key results are cited, or ethical concerns are raised.

## Approaches to versioning

Versioning (the ability to upload a revised copy of an article) is a key feature of preprint servers; it allows researchers to correct, expand, and improve their articles over time. This is important because preprints are used to share early-stage research.

Since preprints can change significantly between versions, it’s important that researchers are able to accurately refer to the work as it was at the time of citation. This requires that preprint servers not only preserve all versions, but also maintain metadata that accurately and unambiguously describes the relationships between them. This metadata is useful for helping indexing servers and reference managers deduplicate records and point readers to new versions when available.

If we could invent a system for maintaining versioning information from scratch, we'd like to see a permanent identifier (a PID) used to identify a work *and* all of its versions. Each version of that work would then be assigned its own version identifier (VID). Individual versions of that work could be identified (and resolved) with a combination of the PID and VID separated by a control character that unambiguously signals the boundary between the PID and the VID.

Most preprint servers use DOIs as PIDs, but DOIs are just strings, and publishers can and do include any character. There are no reserved characters that could be defined to be the control character that separates the PID and the VID unambiguously. Thus, in practice, preprint server providers have "hacked" versioning functionality in one of three ways, outlined below and in table 3.

### Approach 1: Single PID

The preprint server or platform registers one PID and updates the metadata to reflect the most recent version of the preprint, sometimes using metadata fields to preserve information about previous versions. For example, bioRxiv has used "posted" and "accepted" dates to store information about the time of posting of different versions.

- Pros: This approach makes it easier for indexing servers and other users of the data to avoid creating duplicate, potentially unconnected, records of the same preprint, one for each version. This also minimizes fragmentation of information about a paper across multiple metadata records and reduces the likelihood that citation counts are split.
- Cons: Metadata may be overwritten at the indexing service. While it is possible to encode multiple dates in the metadata, some information, such as author list and abstract, may be completely overwritten in some indexing service records.

### Approach 2: One PID for each version

The preprint server or platform registers a single PID for each version which may be linked to one another (for example, Cambridge Open Engage uses the "is-version-of" relationship type. F1000, which does not use preprint DOIs, refers to the previous version with an "update-to" field). In practice, many preprint servers define these series of DOIs in a way that makes it easy for readers to page through the versions by ending the DOI string with indicators such as ".1", ".v1", "-v1", or "/v1". However, since many servers employ different approaches, metadata users seeking to parse this version information automatically must adapt to each naming convention individually.

- Pros: This approach is functional for F1000 and other platforms.
- Cons: Multiple DOIs can create duplication problems for indexers if links between versions are unclear. Citations to the overall work may not be aggregated.

### Approach 3: Concept PID

The preprint server or platform registers one PID for each version, plus an extra (called a Concept DOI) that is continually updated to point to the most recent one. Zenodo concept and version DOIs are not related to one another in any way that is discernible from their DOI string. However, a ChemRxiv concept DOI is the same as the version DOIs minus the terminal characters that define the version number. Despite these differences, both approaches are equivalent in terms of:

- Pros: The concept PID introduces one calling point for the article that always points to its most recent version.
- Cons: As explained by the Zenodo DOI FAQ, the Version and Concept DOIs are not distinguished from one another in any structured way. Readers may be confused about whether to cite the concept or a specific version. Unless versions are clearly linked, this approach may create duplication problems for indexers and citations to the overall work may not be aggregated.

**Table 3: Preprint server and platform versioning approaches**

| Preprint Server/repository | PID | PID structure | example | notes |
|---|---|---|---|---|
| All F1000Research platforms | Crossref DOI | DOI for each for version | https://doi.org/10.12688/f1000research.17927.1<br>https://doi.org/10.12688/f1000research.17927.2<br>https://doi.org/10.12688/f1000research.17927.3<br>https://doi.org/10.12688/f1000research.17927.4 | All versions get individual registered dois. |
| Zenodo | Datacite DOI | “Concept” DOI duplicating latest version | https://doi.org/10.5281/zenodo.3665724 (master)<br>https://doi.org/10.5281/zenodo.3665725 (v1)<br>https://doi.org/10.5281/zenodo.3666256 (v2) | |
| bioRxiv | Crossref DOI | Single DOI, specific URL for each version | e.g. DOI: https://doi.org/10.1101/022368<br>Versions:<br>https://www.biorxiv.org/content/10.1101/022368v1<br>https://www.biorxiv.org/content/10.1101/022368v2 | Single DOI that defaults to the most recent version. Note that bioRxiv has recently changed the DOI format (https://doi.org/10.1101/2020.01.30.927871) where the date is the submission approval date for the first version. |
| arXiv | arXiv ID | Single arXiv ID prefix, suffix for each version | https://arxiv.org/abs/2001.05557 (resolves to latest)<br>https://arxiv.org/abs/2001.05557v1<br>https://arxiv.org/abs/2001.05557v2<br>https://arxiv.org/abs/2001.05557v3 | |
| All OSF preprint servers | Crossref DOI | Single DOI, past versions available by download | https://doi.org/10.31235/osf.io/md7ts | Example v2 download link: https://osf.io/download/5c55173ee16f550019872a13/?version=2&displayName=Peterson%20-%20Media%20Decline%20-%20Jan%2 |

| | | | | |
|---|---|---|---|---|
| | | | | 024%202019-2019-12-12T14%3A29%3A02.818Z.pdf |
| Research Square | Crossref DOI | DOI for each for version | https://doi.org/10.21203/rs.2.17683/v1<br>https://doi.org/10.21203/rs.2.17683/v2<br>https://doi.org/10.21203/rs.2.17683/v3 | |
| Authorea | Crossref DOI | DOI for each for version | https://doi.org/10.22541/au.159170748.81320866<br>https://doi.org/10.22541/au.159170748.81320866/v2 | All versions get individual DOI. Second version and beyond are composed of original DOI + appended “/vX” where X is 2 and beyond |
| ChemRxiv | Crossref DOI | “Concept” DOI duplicating latest version | https://doi.org/10.26434/chemrxiv.6820229<br>https://doi.org/10.26434/chemrxiv.6820229.v1<br>https://doi.org/10.26434/chemrxiv.6820229.v2<br>https://doi.org/10.26434/chemrxiv.6820229.v3 | DOI with no suffix is the latest version (e.g. version 3 in this example) |
| ESSOAr | Crossref DOI | DOI for each for version | https://doi.org/10.1002/essoar.10501118.2 | versions are noted w/ a .# after main DOI |
| MedRxiv | Crossref DOI | Single DOI, specific URL for each version | | |
| Preprints.org | Crossref DOI | DOI for each for version | https://doi.org/10.20944/preprints202003.0078.v1 | All versions registered with individual DOI |
| SciELO Preprints | Crossref DOI | single DOI retained for all versions | | |
| MitoFit Preprint Archives | Crossref DOI | DOI for each for version | | |
| PeerJ Preprints | Crossref DOI | DOI for each for version | | |
| Preprints with The Lancet | Crossref DOI | single DOI retained for all versions | http://dx.doi.org/10.2139/ssrn.3544826 | |
| SSRN | Crossref DOI | single DOI retained for all versions | http://dx.doi.org/10.2139/ssrn.3575559 | Single DOI that defaults to the most recent version. |
| Cell Sneak Peek | Crossref DOI | single DOI retained for all versions | http://dx.doi.org/10.2139/ssrn.3460240 | |
| Cambridge Open Engage | Crossref DOI | DOI for each for version | https://doi.org/10.33774/coe-2020-fsnb3<br>https://doi.org/10.33774/coe-2020-fsnb3-v2 | |

### Recommendations

It is likely that different service providers will adopt different practices, so recommendations should be flexible enough to accommodate different approaches. That said, we recommend two practices that will help prevent loss of metadata and duplication of records downstream.

- Preprint servers can prevent the loss of metadata by creating a PID for each version, with or without a master PID.
- In order to properly track versions, preprint DOI records should be updated to contain the PIDs of all other versions of the preprint, for example through the Crossref “is-version-of” relationship type. Note that published versions of the paper should be linked with the “is-preprint-of” relationship type, which is distinct.

## Surfacing review events

Preprints have sometimes been described as versions of articles that have not (yet) been peer reviewed. However, with journals and other services posting reviews alongside preprints, such a definition is no longer accurate. There nevertheless remains a distinction between review processes that are considered final and result in a Version of Record (or VoR; as defined by NISO: “A fixed version of a journal article that has been made available by any organization that acts as a publisher by formally and exclusively declaring the article “published”) and those that do not (e.g. portable peer review initiatives that do not themselves register a new DOI for the revised article). bioRxiv recently changed its disclaimer to read “This article is a preprint and has not been certified by peer review” to address these points.

The availability of peer reviews from a variety of non-traditional sources and other online discussions has the potential to enrich understanding of preprints. However, these benefits will only accrue if such events are tracked and made easily discoverable by readers. It is therefore important to define appropriate standards for identifying, enabling and surfacing third-party reviews, commentary and related activities around preprints.

### Discussion points

The following were discussed during the meeting and within the working group:

- Threshold for what is considered peer review
- Definition of certification and publication ‘state’
- Hosting and archiving practices and user experience
- Indexing and citation practices
- Control of what is displayed
- Post-VoR review activity

### Recommendations

There is a spectrum of review/commentary events that spans everything from brief comments to formal peer review, including annotations, tweets, blog posts, commentary, peer review reports on third-party sites, and portable peer-review initiatives (coordinated by journals or by journal-independent platforms). It was decided that the current technical distinction should be maintained between a) formal certification processes that result in a VoR and preclude subsequent formal publication and indexing (e.g. journal publication with registration of a new DOI) and b) other events that associate reviews and commentary with the preprint DOI but do not preclude subsequent formal publication (e.g. portable peer review services that do not generate a VoR). The former result in a change in “state” of the article (from preprint to formal publication); the latter do not. Review/commentary events that are ‘non-state changing’ can occur around preprints or VoRs (journal articles, F1000R, etc.)

### Peer review services

Peer review services and other sources of commentary are responsible for creation and preservation of their own metadata. They can ensure their outputs are tracked and surfaceable by preprint servers by:

- Registering a Crossref peer review DOI,
- Choosing to be a contributing source to Crossref Event Data, which does not require DOI registration and is free of charge.
- Creating a new, non-preprint DOI for the article if the process results in final certification of the article and generation of a VoR – i.e. change of state.

Other approaches are under development, for example a system proposed by COAR for using Linked Data Notifications to connect review to repositories.

### Preprint servers

Preprint servers can use Crossref Event Data and other tools to alert readers to commentary and third-party review events around preprints. This is to be encouraged but is optional and may be done on a case-by-case basis. The only circumstances in which a preprint server should update their own metadata with information about peer review is when an article undergoes a change of state through certification by publication in a journal and registration of a new DOI. Preprint servers are currently asked to link their content to the corresponding journal article if/when they become aware that one has been published; the Crossref preprint schema includes “IsPreprintOf”, which accepts the journal article DOI.

### Indexing services

Indexing services, preprint servers and other tools can pull information on events related to the preprint from Crossref Event Data or other sources for display alongside preprint records returned in search results, etc. The indexing service, preprint server or tool can choose which sources of information they pull from these sources and how they display the information.

### Authors

Authors wishing to cite preprints in the context of peer reviews or other commentary can cite both the preprint and review object, ideally using DOI for each.

## Limitations to this approach

Event Data does not currently provide much structure for information that might be useful to describe peer review. For example, Event Data from Twitter contains title, date, and author fields, but no obvious way to encode information such as the role of those authors, ORCiDs, or author affiliations. Notably, some peer review services deposit reviews in Zenodo. These events would be picked up in DataCite Event Data, but not in a way that is distinguishable as a review as opposed to a more general reference to a paper. Finally, Event Data is restricted to information about articles that have a DOI, which would preclude its use on servers such as arXiv and HAL that do not issue DOIs.

Crossref Peer Review DOIs contain metadata specifically designed to describe reviews, and they can refer to objects that do not themselves have DOIs through the general typed relations schema. However, an entity wishing to register these DOIs must be a Crossref member and pay the membership fee as well as a Content Registration Fee for each DOI registered.

Note that Crossref relationships, such as IsPreprintOf, can point to multiple preprints or articles, meaning that the schema can link a preprint to multiple published articles. This feature may be relevant for overlay journals that wish to convey a change of article “state” through certification.

Another potential concern with separating metadata about preprints and reviews is the need to cite an entire constellation of objects rather than a single reference for a preprint. This could conflict with journal limits on the length of reference lists; in the era of online publication this will hopefully become less of a stumbling block.

Finally, providing metadata to Crossref either to register DOIs (which is currently associated with a fee) or to participate in Event Data may be onerous for peer review platforms and services, particularly those operating in nontraditional settings or with few technological resources.

## Encouraging data availability at the point of preprint

Data, code and supporting materials enable the research community to assess the validity of a claim made in an article and build on the work reported. Researchers in some disciplines are currently expected to provide these materials if requested before or during the peer review process, as well as for wider consumption at the point of journal publication. Papers for which data are shared have a citation advantage, and there is evidence that the general public (Pew Research Survey, 2019) and researchers (COS Survey) view publically available data as important for making trust and credibility judgements of preprints.

Preprint servers have a role to play in encouraging data sharing, and thereby building trust and credibility, early in the research process. While journals might ultimately have different data-sharing requirements, encouraging data availability at the point of preprinting may improve the amount of data available at the downstream journal publication.

However, we realize that there are barriers to increasing data sharing at the point of preprinting that need to be seriously considered and that data and code sharing still present tension with current researcher workflows. Thus, the desire to encourage data sharing must be balanced against the practical realities of preprinting—too many barriers to posting a preprint may discourage researchers from doing it at all. Note that some journals do not yet implement data statements; at topfactor.org, a number of journals have level 1 data transparency policies (which require authors to state whether data is available or not) but few have a level 3 data transparency policy, indicating that data are required and checked by the journal. Furthermore, different communities have different norms when it comes to data sharing, and there may be ethical or legal restrictions to the sharing of some datasets. Education may be required to familiarize authors with what is meant by the term “data availability” and how much information can or should be shared, especially when data could identify patients or share the location of an endangered species.

Data sharing and any associated challenges span beyond preprints, so we acknowledge it is not something for preprint servers alone to solve. However, we as a community can work to implement nudges that actively encourage better data sharing practices and increase data visibility. Here we propose actions to support a productive direction.

## Current status

34/46 preprint servers surveyed allow authors to upload at least some supplemental materials/information along with their preprint; however, few specifically encourage data sharing or clearly cue the existence of available data to researchers or through APIs. In this survey, only medRxiv, F1000 and affiliated Open Research platforms require a data availability statement. Several preprint platforms do encourage data availability statements (MitoFit Archives, PeerJ Preprints (no longer accepting new submissions) and preprints.org), however, their data-sharing policies (and those of many journals) enable authors to make assertions that do not actually amount to public data sharing (e.g., stating that ‘data available upon request’ or ‘data is available in the paper/supplement’). Research has shown that data requests result in low rates of data sharing, and ‘data in the paper’ is often aggregate data rather than raw. Because of these limitations, we propose that preprint authors be encouraged to post data publicly rather than make it available upon request.

Of an initial sample of COVID-19-related preprints made available in PMC under a CC license (n=307), roughly 70% had some sort of associated data content (i.e., supplementary material or DAS). However, we currently do not know how many authors who could be sharing data, code, or materials with their preprint are actually doing so, and we have no means of analysing the impact of data sharing at this early stage. In a 2018 survey only one preprint server had a policy about data (specifically, a recommendation to make data available) and only two servers had more than 50% of preprints with some data available, most servers reported much lower levels of data availability. A pilot by arXiv between 2011 and 2013 allowed researchers to submit data files. Analysis of preprints and research data deposited revealed the metadata was incomplete, and deposits often lacked readme files, rendering the data of limited use. This suggests that researchers may either be unwilling to invest in offering more documentation or structure or require more support to increase the usability of the data provided. This support may be infeasible for preprint servers to provide.

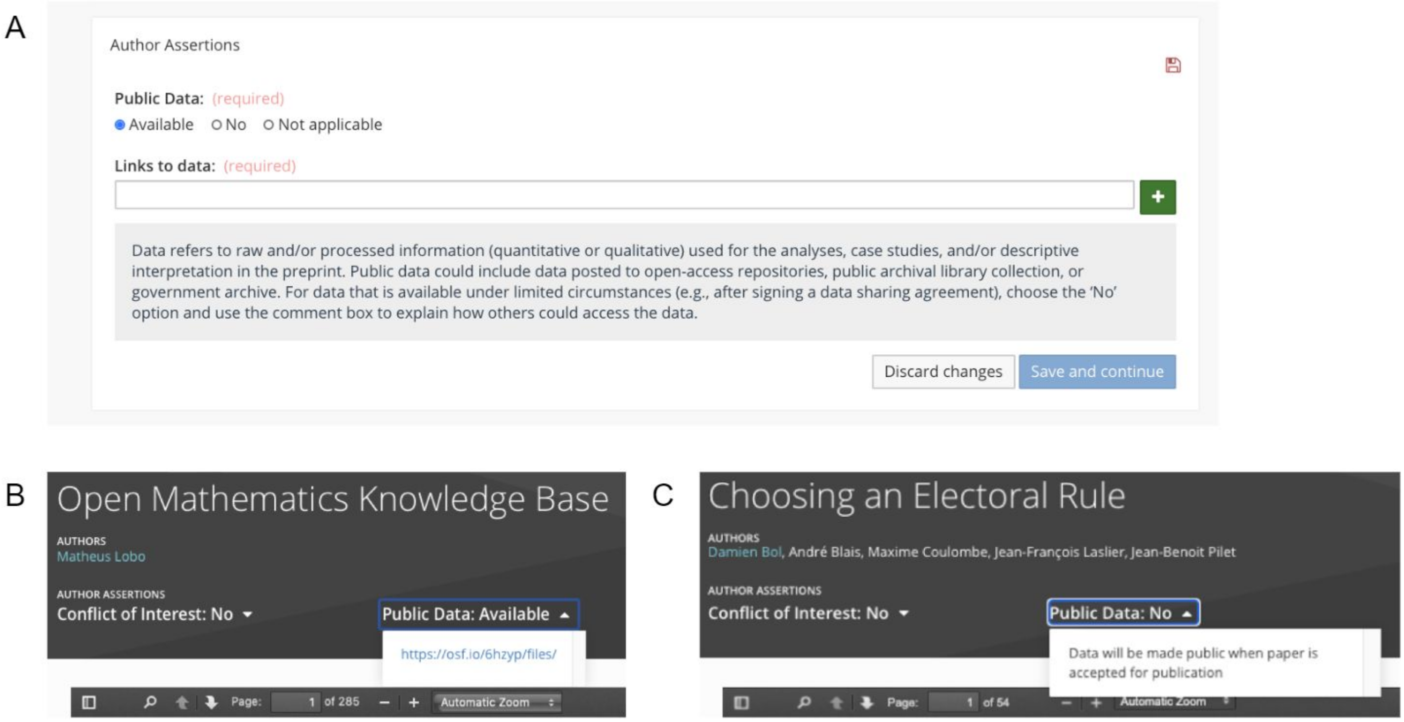
Given all of these factors, it is not possible or desirable to *require* data sharing with preprints without first understanding community readiness, as we wish to avoid deterring researchers from posting their work as a preprint.

## Recommendations

### Preprint servers

- When authors submit directly to preprint servers, the platform should require uploaders to state whether data underlying their preprint is publically available using the following categories at a minimum: ‘Yes’, ‘No’, ‘Not Applicable’. If uploaders choose ‘Yes,’ they either directly upload the data files or provide links and/or accession numbers to where the data is already publicly available. If authors choose ‘No’, then they are given the option to describe why data are not publically available (e.g. available upon publication, controlled dataset). The “Not Applicable” option is intended for papers that report no analyses. Other options, such as “available upon request” and “included in the paper” are functionally equivalent to not making data publicly available, and thus should not be included among the options. See figure 1 for the implementation of author assertion of public data on OSF Preprints infrastructure (COS). See figures Y and Z for example preprints on OSF Preprints with author assertions for data availability.

- Preprint servers should provide public information about their requirements on data sharing and/or data statements, including the fact that sequence data may be made public if reported in a preprint (see bioRxiv FAQ).
- Preprint services should make the minimal data availability information easily discoverable for both humans and machines. This means that data availability statements and their additional information (e.g. links, explanations) should be clearly displayed on preprint pages, regardless of the answer to the initial question, and should be programmatically accessible. Services should clearly indicate that these statements are author assertions and aren't checked or validated by the service, as are most statements associated or contained within a preprint submission.
- Preprint services should strongly encourage researchers to link to data and software held in external repositories and include the citation in the reference list following leading practices rather than require that data files be uploaded directly to their preprint service. Since deposition in a repository can improve data discoverability and prevent the need to upload the data elsewhere upon journal publication, deposition at a repository is preferred.
- Preprint servers should transfer collected data availability statements to journals along with manuscripts and other metadata where transfer arrangements exist.



Author Assertions

Public Data: (required)
● Available ○ No ○ Not applicable

Links to data: (required)

Data refers to raw and/or processed information (quantitative or qualitative) used for the analyses, case studies, and/or descriptive interpretation in the preprint. Public data could include data posted to open-access repositories, public archival library collection, or government archive. For data that is available under limited circumstances (e.g., after signing a data sharing agreement), choose the ‘No’ option and use the comment box to explain how others could access the data.

Open Mathematics Knowledge Base

AUTHORS
Matheus Lobo

AUTHOR ASSERTIONS
Conflict of Interest: No
Public Data: Available
https://osf.io/6hzyp/files/

Choosing an Electoral Rule

AUTHORS
Damien Bol, André Blais, Maxime Coulombe, Jean-François Laslier, Jean-Benoit Pilet

AUTHOR ASSERTIONS
Conflict of Interest: No
Public Data: No
Data will be made public when paper is accepted for publication

**Figure 1**: A) Author assertion for public data on preprint upload, links to more details about the preprint upload process can be found here, for more information, contact nici@cos.io. B) Example preprint on OSF Preprints with author asserting public data available with link. C) Example preprint on OSF Preprints with author asserting public data not available with explanation (optional).

### Authors

- Authors should add relevant DOIs and accession numbers to their manuscript (ideally in the reference list) and, where requested, to preprint metadata collected during the posting process.
- In the case of embargoed repositories, authors should understand repository policies on what (eg citation of an accession number in a preprint) will trigger data publication.

### Journals

- Publishers that transfer submissions to preprint servers should collect data availability statements (ideally as metadata, as below) to be transferred to the server and work toward interoperability of data availability statements.

### Meta-researchers

- Meta-researchers should use the data provided in the data availability statements to study and improve data sharing practice to in turn increase the rate and quality of data sharing. Qualitative analyses of provided links and reasons for not publically sharing data can be used to help develop additional instructions/resources to support researchers, as well as potentially identify common categories of reasons which could be added as structured response options in future iterations. Quantitative analyses of rates of data sharing assertions and the rates at which links point to correct and well documented data can be used to get a more thorough understanding of data-sharing rates and the extent of depth of checking that would be needed to validate author assertions.

### Generalizing this approach to code, materials, and more

Though we specifically discuss data here, the same approach can be generalized to signal the availability of code. We recommend a general metadata element that can contain availability statements and PIDs for several flexible categories such as data, code, and materials.

## Defining preprint withdrawal & removal

In order to support the legitimacy of preprints as scholarly outputs, citation of preprints and their consideration as part of institutional review processes or grant applications, a framework that supports the permanence of posted preprint records is needed.

Many cases that would require a correction to be issued at a journal can be handled by simply posting a new version of the preprint. While the permanence of posted preprints is encouraged, situations may arise where it is necessary to notify readers about formal actions that may affect the legitimacy or accuracy of the original information. In the simplest scenario, the authors may no longer believe their findings or interpretation presented are correct. In other cases, the continued availability of the preprint could pose a danger or legal risk to the authors, the server itself, or the wider community. Preprint platforms are handling situations related to concerns on the legitimacy or accuracy of a posted preprint in a variety of ways (see table 4). A consolidated framework will provide further clarity for authors and readers as well as a guide for any platforms that need to handle such situations for the first time.

Preprint platforms may take two approaches to address the need to update the preprint record due to serious concerns about its legitimacy or accuracy: withdrawal or removal. It is recommended that preprint platforms use metadata that designates the status of the preprint as either withdrawn or removed and to communicate the status update to indexing services, as relevant.

While there may be some overlap in circumstances that result in the removal or withdrawal of preprints, there is value in having a separate framework for each of those steps, as one involves the removal of content whereas the other does not. From the perspective of transparency to readers coming to the preprint record, and consistency in the approach of handling metadata and information transfer to indexing services, the use of separate terminology would clarify expectations in terms of whether the original preprint record remains.

Platforms can withdraw or remove preprints in response to appropriate requests from authors or institutions (e.g. as a result of an institutional investigation). Typically, preprint platforms cannot be expected to respond to third-party concerns about a posted preprint by undertaking a detailed investigation. The onus is on the author(s) of the preprint to respond to third-party concerns, which may be raised within the comments or annotation section.

## Withdrawal

The purpose of the withdrawal is to alert readers about subsequent formal actions that affect the legitimacy/accuracy of the research reported in the preprint. The original preprint record remains on the platform and is accessible to readers, and a notification is added to alert readers that the preprint has been withdrawn. The withdrawal notification should be clearly visible to readers, either as a prominent alert on the preprint record or by making the withdrawal notice its own default landing page, from where the original preprint record can be accessed. The withdrawal notice should indicate who is withdrawing the preprint (the author or the preprint platform) and provide reasons for the withdrawal (e.g. methodological error, ethical issues). It is recommended that readers are directed to the corresponding authors if they require further information.

The withdrawal of a preprint is an equivalent step to the retraction of peer-reviewed publications. Withdrawal of a preprint may be requested by the authors or by a body charged with oversight of the authors (e.g. institutional leadership). The withdrawal of a preprint may be a necessary step in the following situations:

- Alert readers to serious concerns about the research in the preprint, either as a result of errors (major or minor) that cannot be addressed via updates/revisions posted as a new version, or as a result of research integrity breaches (e.g., an institutional ruling on data manipulation or plagiarism)
- The preprint reports unethical research (e.g. lack of IRB approval, lack of appropriate informed patient or research participant consent)
- There is evidence of unethical authorship practise for example, forged authorship, lack of consent from co-authors or misrepresentation of author credentials

## Removal

The purpose of removal is to remove some or all of the preprint and its metadata because the continued availability of the content represents a serious danger or legal risk to the authors, the server, research participants or a third party (e.g. the public). The removal of a preprint may be a necessary step in the following situations:

- Serious legal issues e.g. libel, copyright infringement
- The preprint includes confidential or identifying information relating to a research participant(s)
- Lack of informed consent from research participants or ethical ramifications for a vulnerable group
- The preprint reports information that represents a serious risk to public health or potential public security concerns

Removal of a preprint may be requested by the authors, the copyright owner or by a body charged with oversight of the authors (e.g., an institution or a government body). The preprint record is deleted and a removal notification is added to alert readers about the fact that a preprint record previously existed that is no longer available. If a DOI is assigned to the preprint record, the DOI should resolve to the page hosting the removal notification.

Given the interest in maintaining the continuity of the preprint records, and that circumstances that incur a major risk are expected to be rare, the removal of a preprint is only warranted in exceptional circumstances, where the risk incurred cannot be mitigated by revisions to the preprint (via versions) or a withdrawal notice. Preprint platforms may have dedicated frameworks to review the legitimacy of a removal request and consider whether it should be accepted; as part of this, they may request further information from relevant parties (authors, institutions, copyright holder etc). In exceptional circumstances if there is a major breach to privacy or intellectual property, or a major risk to public security, it may be necessary to remove or redact part of the preprint metadata.

It should be noted that expunging the record for published articles is an extremely rare occurrence, and that even if the individual preprint record is removed, it is likely that traces related to the document will remain online e.g. via third-party indexers, the removal of the individual preprint does not guarantee that all possible online traces of the document would cease to exist. Withdrawal is therefore recommended wherever possible. Note also that withdrawal protects innocent parties by providing a transparent record of events, minimizing the possibility for harmful speculation as to their involvement.

**Table 4: Examples of preprint withdrawal and removals and how they relate to the proposed definitions**

| **Stated reasons for withdrawal/removal** | **Notice** | **File status and metadata representation of withdrawal/removal** | **Categorization under proposed definitions** |
|---|---|---|---|
| Withdrawn temporarily by authors | https://www.biorxiv.org/content/10.1101/765610v2 | Full text of version 1 remains. Crossref metadata includes “withdrawn” in the abstract. | Withdrawn |
| Withdrawn by server due to false affiliation | https://www.biorxiv.org/content/10.1101/497875v2 | Full text of version 1 remains. Crossref metadata includes "withdrawn" in the abstract. | Withdrawn |
| Withdrawn by author as no longer valid | https://arxiv.org/abs/2002.01248 | Full text of the previous version remains. arXiv metadata includes “withdrawn” in the comments. | Withdrawn |
| Withdrawn by server due to overlap with other sources | https://arxiv.org/abs/2002.00746 | Full text of the previous version remains. arXiv metadata includes “withdrawn” in the comments. | Withdrawn |

| | | | |
|---|---|---|---|
| Withdrawn at institutional request due to submission without co-author consent | https://peerj.com/preprints/2910/ | Full text of version 1 remains. Crossref metadata includes “withdrawal” in title and abstract. | Withdrawn |
| Withdrawn owing to erroneous inclusion of confidential information relating to a third party | https://www.biorxiv.org/content/10.1101/455451v1 | Files removed. Crossref metadata contains “withdrawn” in abstract. | Removed |
| Withdrawn due to author disagreement | https://www.researchsquare.com/article/rs-15022/v2 | Full-text html and pdf versions available for v1. Abstract replaced with withdrawal notice and pdf redacted for v2. Crossref metadata show full abstract for v1 doi and withdrawal notice for v2 doi. Title, authors and affiliation retained in metadata. | Withdrawn |

## Situations where the preprint has an associated journal article

In situations where the research included in a preprint has subsequently been published in a journal, if the preprint server executes a withdrawal or a removal they should alert the journal to the concerns identified. Similarly if a journal decides to retract or completely remove an article that is also available as a preprint, the journal editors should notify the preprint platform so that the latter can consider whether a withdrawal notification or a link to/notification of the retraction (or another step such as a new version, or removal) is needed. While a mechanism (such as Crossref’s Crossmark) that allows notifications to be automated would provide an ideal solution to such communications between preprint servers and journals, it is ultimately the author’s responsibility to ensure the preprint and published record are appropriately updated. In the absence of automated processes, the author should notify the preprint server or journal of actions taken on the other record(s) of the paper so that they can consider if necessary steps are needed on the preprinted paper or the journal article, as applicable.

## Other considerations

Situations may arise in the context of direct submissions to preprint platforms or via transfers from journal to preprint channels where authors request the withdrawal of the preprint after this has been posted, due to lack of understanding about implications of preprint posting or ramifications about submission to journals that do not consider papers posted as a preprint. As noted above, however, an approach that supports the continuity of preprint records is encouraged and thus preserving the preprint record should be the goal unless there are concerns about the preprint falling under the circumstances outlined for withdrawal or removal.

For publishers operating initiatives where authors are offered the option to post the paper as a preprint upon submission to a journal, it is particularly important for the journal to have clear information for authors that posting the preprint is irreversible and separate from the consideration of the manuscript for peer review and/or eventual publication at the journal.

Where applicable, comments on the original version(s) of a withdrawn preprint should be retained in order to provide context for readers.

There may be cases where intermediate actions (such as the removal of a single figure or supplemental files) are warranted. In this case, servers could help indexing services remain up to date by marking the old version as a removal and posting a new version without the problematic content.

Indexing services and discovery tools should use server-provided metadata indicating withdrawal/removal status to clearly label preprints appropriately when surfacing these records to users.

**Table 5: Summary of recommended framework for preprint withdrawal or removal**

| | **Withdrawal** | **Removal** |
|---|---|---|
| Does the original preprint full text remain? | Yes | No |
| Does the original preprint DOI/URL resolve to a dedicated webpage? | Yes | Yes |
| Is a new preprint version created to express the change in status? | Yes | Yes |
| Prominent notification for readers added to all versions? | Yes | Yes |
| Type of concern | Errors in research content that cannot be addressed by posting a preprint revision/new version, research ethics, unethical authorship practise | Legal issues, privacy breach, serious risk to public health or national security |
| Notification to indexing service | Yes | Yes |
| Metadata elements | 1) withdrawal status set to “withdrawn” 2) removal/withdrawal reason added (optional) | 1) withdrawal status set to “withdrawn” 2) file modification status set to “modified” 3) removal/withdrawal reason reason added (optional) |

## Beyond the scope of withdrawals and removals

Situations where concerns are noted about a preprint which can be mitigated via versioning or community feedback do not require withdrawal or removal steps. Examples of such situations include:

- Author addition or removals, or changes to the order of authors which can be addressed via a new preprint version
- Errors in analysis that can be addressed by re-analysis and/or provision of additional information via a new preprint version

- Concerns about the integrity of images or data in the preprint identified (by either a reader, a journal or an institution) but where no determination of misconduct has been reached via an institutional investigation
- Concerns about competing interests

### Reposting a withdrawn preprint

If a preprint is withdrawn due to an identified error and the author(s) seek to later post a revised version of the work as a new preprint, the new preprint record should refer to the earlier withdrawn preprint and provide context on the changes made compared with the original record.

## Cultural aspects of preprint use: engaging stakeholders to raise awareness

While the use of preprints is increasing in the biomedical sciences, we know that adoption varies per discipline (10.7554/eLife.45133) and that there is still low awareness of preprints amongst many researcher communities. To encourage further adoption, we need to raise awareness and understand the benefits and concerns that different researchers associate with preprints. Engaging with research communities (e.g. the ASAPbio Community) and librarian networks provide avenues to support researcher education and awareness.

We also know that there is a need to support better understanding of preprints among journal editors. Even though many journals have adopted policies stating that they will consider work posted as a preprint, some authors continue to worry that posting a preprint will preclude consideration by their journal of choice, and others claim to have had articles posted as preprints rejected by journals despite such policies. We should create resources to support editor education and raise awareness about the value of preprints as a complementary step compatible with journal publication.

Preprints are an important element in the research cycle; we need to communicate their value to all stakeholders, from research funders and institutions, to journals, librarians, journalists and the general public. Building an understanding of the needs of the different stakeholders will require ongoing engagement efforts.

## Future considerations

The working groups referred to above focused on elements of metadata management and stakeholder engagement. Building trust in preprints as a legitimate tool for science communication will require further work in a broader range of areas. Below we discuss items that arose as part of the conversations at the workshop that may present challenges for future preprint adoption.

### Linking preprints with published versions

Bidirectional linking between preprints and resulting peer-reviewed publications surfaces steps in the science dissemination process and provides readers with valuable information about the evolution of the work. However, it is currently difficult to confidently identify all preprints that have subsequently been published as journal articles. Moreover,subsequent updates to the journal publication, such as a retraction, may be missed. There are several approaches to making such matches: directly from journals for journal-to-preprint submissions (eg In Review); via Crossref notifications generated by fuzzy matches for title, authors and other basic metadata performed by the server or Crossref; via automated search services such as Google Scholar; or via author-generated notifications to preprint platforms once the article is published. Where individual indexing services and preprint servers have developed their own matching approaches, these tools could be made available via APIs.

Whose responsibility is it to designate and surface such relationships? Currently, Crossref asks preprint servers via email (though API access has been requested) to update the 'is-preprint-of' relation of preprint DOI records with a link to the journal version. The reciprocal field in journal DOI records, 'has-preprint,' is then updated automatically by Crossref, preserving provenance of the information. Many preprint servers display links to the journal version of an article (asapbio.org/preprint-servers), but the reciprocal links are rare on the journal side (transpose-publishing.github.io). Publishers may in some cases not be able to or prefer not to display such information.

## Sustainability (business models, archiving & long-term preservation)

A range of issues need to be taken into consideration to ensure the long-term sustainability and future development of preprint services. Although the key issue is financial stability, the development of transparent organizational models and policies should be encouraged.

Currently preprint platforms operate different business models according to their stewardship and revenue sources. The main sources of revenue at present include foundations (e.g., bioRxiv and arXiv), publishers (e.g. SSRN)  societies (e.g. ESSOAr and ChemRxiv), and libraries (e.g. arXiv, EarthArXiv). There is currently no long-term guarantee of those revenue sources, and preprint platforms so far lack a reliable and practical way to monetize services (e.g, value-added services). Preprints emerged as a "public good," and they are free to submit and free to read.

Although it might be difficult to develop a comprehensive compilation of revenue sources and expenses for preprint servers, sharing such financial information would increase understanding of the resource needs for running platforms and in turn raise awareness about the levels of investment required.

Another important element of sustainability is the implementation of long-term-archiving strategies to ensure enduring access to preprint content. Publishers and societies have a well-established tradition of working with third-party archival services to secure their digital assets. It would be beneficial for preprint platforms to explore archival service options  (for example CLOCKSS, Portico, and Internet Archive) in order to ensure the long-term accessibility of their digital content.

## A framework of expectations for preprint platform operation

The preprint platform landscape is rapidly evolving, and we have seen the launch of a number of new platforms over the last 5 years. At this stage in the development of preprints for the life sciences, it would be beneficial to have a shared framework of expectations for preprint platforms. Such a framework would serve to reinforce trust in the services the platform provides and provide a foundation for future new entrants to adhere to in order to play a role as trusted entities in this landscape. For any such framework to be successful, it will need to be driven by a core group of preprint platform operators.

Other actors in the research ecosystem have already expressed interest in some form of guidelines around preprint platform expectations. Several funders either currently provide or plan to provide their grantees with guidance about which preprint platforms meet their expectations. As an example, NIH issued guidance in 2017 on selecting interim research production repositories to facilitate the impact, measurement and the integrity of the scientific record (NOT-OD-17-050: Reporting Preprints and Other Interim Research Products) and more recently issued preprint server eligibility information as part of their pilot to include NIH-funded research in PMC and PubMed. Such guidelines are important both for grantees and for evaluating compliance with funder guidance and mandates. If preprint platforms produce a consistent framework of expectations that funders are willing to adopt and inform grantees about, this would send a strong signal about the legitimacy and value of the framework. It is therefore important that funders and servers collaborate to ensure this framework aligns with their intentions.

As a framework of expectations is developed, it will be important to bear in mind that while it should be specific enough to enable legitimacy of preprint platforms for life and biomedical sciences, it must be compatible with platforms designed for other disciplines or specific geographic communities. Any framework of expectations for preprint platforms should embrace inclusion and enable progress toward trust in preprints across all research communities.

## Building trust in preprints: an ongoing process

This report summarizes the recommendations arising from discussions by a group of stakeholders involved with preprints and the science communication process. In order to drive increased trust in preprints in the biomedical sciences, continuing steps will be needed to both support the discoverability and use of preprints and to raise awareness among different communities.

The report outlines a number of recommendations around metadata for preprints as the cornerstone to support increased discoverability and utility. Clear and consistent metadata at all stages of the preprint cycle, from posting to revisions to linking to journal publication, will ensure transparency for both authors and readers, enable clearer linking of outputs, and facilitate evidence building around preprint trends and practice. While technical developments will not bring trust on their own, they can facilitate progress toward this goal.

Building trust in preprints also requires an understanding of how different communities perceive them as a research communication tool, what has driven adoption by some, and what the challenges are for others. We will continue to work toward this. We hope that the recommendations outlined here promote greater trust in preprint use and we welcome feedback from all stakeholders in the community.

## Acknowledgements

We thank the attendees of the meeting (below) as well as Eleonora Presani (arXiv), Shirley Decker-Lucke (SSRN), Michael Magoulias (SSRN), Alberto Pepe (Authorea), Kate Perry (Wiley), Rachel Burley (Research Square), Michele Avissar-Whiting (Research Square), Javier Iglesias (MitoFit), Gunther Eysenbach (JMIR), Jason Barkeloo (Open Therapeutics), Marshall Brennan (ChemRxiv), and Allison Leung (SAGE) for their input.

# Appendices & supporting materials

## Attendees

- Jeffrey Beck, NCBI, US National Library of Medicine, NIH
- Theo Bloom, BMJ and medRxiv
- Rachel Burley, Research Square
- Tom Demeranville, ORCID
- Kevin Dolby, Medical Research Council (UK)
- Jim Entwood, Cornell University and arXiv
- Kathryn Funk, NIH National Library of Medicine (USA) and PubMed Central
- Brooks Hanson, American Geophysical Union and ESSOAr
- Melissa Harrison, eLife
- Hannah Hope, Wellcome Trust
- Michele Ide-Smith, Europe PMC
- John Inglis, Cold Spring Harbor Laboratory, bioRxiv and medRxiv
- Jamie Kirkham, University of Manchester
- Rachael Lammey, Crossref
- Maria Levchenko, EMBL-EBI and Europe PMC *(co-organiser)*
- Emily Marchant, Cambridge University Press
- Michael Markie, F1000Research
- Johanna McEntyre, EMBL-EBI and Europe PMC *(co-organiser)*
- Alice Meadows, NISO
- Alex Mendonca, Public Knowledge Project (PKP) and SciELO Preprints
- Mate Palfy, Company of Biologists
- Michael Parkin, Europe PMC
- Naomi Penfold, ASAPbio *(lead organiser)*
- Nici Pfeiffer, Center for Open Science
- Jessica Polka, ASAPbio *(co-organiser)*
- Iratxe Puebla, PLOS and representative of COPE
- Oya Rieger, Ithaka S+R and arXiv *(co-organiser)*
- Martyn Rittman, Preprints.org
- Richard Sever, Cold Spring Harbor Laboratory, bioRxiv and medRxiv
- Sowmya Swaminathan, Nature Research, Springer Nature
- Dario Taraborelli, Chan Zuckerberg Initiative
- Emily White, Focused Ultrasound Foundation and FoCUS Archive

## Agenda

Monday January 20, 2020

| | |
|---|---|
| 9:00am | Opening Remarks and Information |
| 9:15am | Updates and introductions<br>Session lead: Maria Levchenko, EMBL-EBI and Europe PMC<br>9:15am \| Preprint platform updates<br>10:15am \| Stakeholder introductions |
| 10:30am | ASAPbio Preprint Platform Directory report<br>Naomi Penfold, ASAPbio |
| 10:45am | Morning refreshment break |
| 11:00am | Session 1: Minimally useful metadata standards for preprints<br>Session lead: Jo McEntyre, EMBL-EBI and Europe PMC<br>What is/should be captured in preprint metadata?<br>How are/could these data be captured from authors or elsewhere?<br>How are/could these metadata be made openly available to third parties?<br>How do/could we leverage existing standards and infrastructure?<br>11:10am \| Introductory talks<br>Michael Parkin, Europe PMC<br>Tom Demeranville, ORCID<br>Dario Taraborelli, CZI<br>11:40am \| Breakout discussions |
| 1:00pm | Lunch |
| 2:00pm | Session 1: Minimally useful metadata standards for preprints (continued)<br>2:00pm \| Reporting back & whole group discussion |

| | |
|---|---|
| 2:50pm | Session 2: Adherence to and transparency of screening, moderation and withdrawal processes<br>Session lead: Sowmya Swaminathan, Nature Research, Springer Nature<br>Which scholarly publications practices are important to uphold at the preprinting stage?<br>Screening:<br>Which author-dependent practices should be checked before posting, and by whom?<br>How transparent could/should screening processes be?<br>How could their outcomes be communicated externally?<br>Moderation & Withdrawal:<br>What can be moderated after posting, and how?<br>In what circumstances should preprints be withdrawn or removed?<br>What would the impact of removal be on downstream aggregators?<br>How transparent could/should moderation and withdrawal processes be?<br>How could their outcomes be communicated externally?<br>3:00pm \| Introductory talks<br>Iratxe Puebla, PLOS and representing COPE<br>Theo Bloom, BMJ and MedRxiv<br>Sowmya Swaminathan, Nature Research, Springer Nature<br>3:30pm \| Breakout discussions |
| 4:00pm | Afternoon refreshment break |
| 4:15pm | Session 2: Adherence to and transparency of screening, moderation and withdrawal processes (continued)<br>4:15pm \| Breakout discussions (continued)<br>5:00pm \| Reporting back & whole group discussion |
| 5:50pm | Day one closing |
| 6:00pm | Close |
| 7:00pm | Workshop dinner |

Tuesday, January 21, 2020

| | |
|---|---|
| 9:00am | Day two opening |
| 9:15am | Session 3: Indicating preprint status<br>Session lead: Richard Sever, bioRxiv & medRxiv<br>Choose one of the following topics for breakout discussions: how to transparently and accurately convey:<br>(1) A preprint’s review status (from ‘not peer-reviewed’ to otherwise)<br>(2) Availability of supporting data and materials<br>(3) Preprint-level usage (views, downloads and citations by version) |
| 10:30am | Morning refreshment break |
| 10:45am | Session 3: Indicating preprint status (continued)<br>Reporting back & whole group discussion |
| 11:45am | Session 4: Citations, archiving, sustainability, and adoption<br>Session lead: Oya Rieger, Ithaka S+R<br>Choose one of the following topics for breakout discussions:<br>(1) Citation standards<br>(2) Archiving and sustainability of free open-access preprint platforms<br>(3) Encouraging adoption of preprints |
| 1:00pm | Lunch |
| 2:00pm | Session 4: Citations, archiving, sustainability, and adoption (continued)<br>Reporting back & whole group discussion |
| 3:00pm | Review the recommendations and roadmap<br>Session lead: Jessica Polka, ASAPbio<br>Individual work time to review the draft recommendations and roadmap |
| 3:45pm | Afternoon refreshment break |
| 4:00pm | Review the recommendations and roadmap (continued) |

| | |
|---|---|
| | Whole group discussion to review the draft roadmap and consider next actions following this workshop |
| 5:15pm | Workshop closing |
| 5:30pm | Close |
| 6:30pm | Informal pub dinner |