Check for updates

OPINION ARTICLE

On the origin of nonequivalent states: How we can talk about preprints [version 1; referees: 2 approved]

Cameron Neylon (iD) [1], Damian Pattinson[2], Geoffrey Bilder[3], Jennifer Lin[3]

[1]Centre for Culture and Technology, Curtin University, Perth, WA, 6102, Australia
[2]Research Square, Durham, NC, 27701, USA
[3]Crossref, Lynnfield, MA, 01940, USA

Abstract

Increasingly, preprints are at the center of conversations across the research ecosystem. But disagreements remain about the role they play. Do they "count" for research assessment? Is it ok to post preprints in more than one place? In this paper, we argue that these discussions often conflate two separate issues, the history of the manuscript and the status granted it by different communities. In this paper, we propose a new model that distinguishes the characteristics of the object, its "state", from the subjective "standing" granted to it by different communities. This provides a way to discuss the difference in practices between communities, which will deliver more productive conversations and facilitate negotiation, as well as sharpening our focus on the role of different stakeholders on how to collectively improve the process of scholarly communications not only for preprints, but other forms of scholarly contributions.

Open Peer Review

Referee Status: ✔ ✔

	Invited Referees	
	1	**2**
version 1 published 02 May 2017	✔ report	✔ report

1 **Kathleen Fitzpatrick** (iD) , Modern Language Association USA

2 **Todd Carpenter** (iD) , National Information Standards Organization (NISO) USA

Discuss this article

Comments (0)

Introduction

Two scientists, Jimmy Maxwell and Chuck Darwin, meet at a conference and realise that they have common research interests, though one is a physicist and the other a naturalist. So they agree to collaborate, and their work develops quickly into a theory so big it could revolutionise both their disciplines.

They write up their work and, egged on by the physicist, decide to post to a preprint server before submitting to their target journal, *The Science of Nature*. The preprint causes a sensation! It receives attention, generates heated discussion, and citations ensue from their colleagues in both disciplines. The journal submission, however, faces a rockier path, getting held up by Reviewer #3 through four rounds of revision over a sticky issue involving the techniques for measuring the forces of barnacle-rock attraction.

During the publication delay, offers start pouring into young Maxwell's inbox from universities and companies wishing to recruit the young physicist. He takes a plum job and goes on to change the course of physics forever. Chuck, on the other hand, finds offers hard to come by. His grant applications to fund a research trip to far-flung islands fail because his CV lacks the high impact articles required to make him stand out. In despair he quits the bench and opens a pet shop. Some decades later the two researchers are recognized by the award of the prestigious Prize of Nobility. Maxwell's place in the firmament is assured, while Darwin returns to his pet shop, now specialising in finches, where something about their beaks bothers him until the day he dies.

We open with this cheeky illustration to foreground one main point: *different communities* **grant the same object** *different degrees of importance*. We can complicate the story by revealing that both researchers were scooped between posting the preprint and article publication. Or funding panels in each discipline assess their applications and count the outputs as scholarly contributions in different ways. But they all illustrate the same central point. There exists no universal standard of when an output is considered as part of the formal scholarly record. Rather, it is determined by particular groups in particular contexts.

No universal definition of preprint exists (and never will)

The pace of technological change over the past two decades has far outstripped the language we use to describe the objects and processes we use to communicate. This disconnect between language and technology is at the root of the current debate around preprints. The very word "preprint" is an odd combination of retronym and synecdoche. A preprint is increasingly unlikely to ever be a precursor to anything that is physically printed onto paper. At the same time, that use of "print" takes one small part of scholarly publishing to stand in for the entire process. A preprint is different from a working paper, yet both are entirely different to an academic blog post. Additionally, all these appear in designated online repositories as digital documents that are recognizably structured as scholarly objects. Some preprints are shared with the future intent of formal publication in a journal

or monograph. But not all. The term is used to mean a host of different things, and as such, remains referentially opaque[a]. An earlier version of this article is available on the "preprint" server BioRxiv. Should we refer to that here? Should it be formally referenced? Or is that "cheating" by inflating citation counts? What do we call the version of this article on *F1000Research* after posting, but prior to the indexing that follows approval by peer review?

Wikipedia is a good source for identifying common usage. At the time of writing, it defines a preprint as "a draft of a scientific paper that has not yet been published in a peer-reviewed scientific journal." This definition encompasses everything from early, private drafts of a paper that the authors have never shared with anyone, all the way to drafts of accepted manuscripts that have yet to go through a publisher's production process. Interpreted liberally, the Wikipedia page itself might even be included[1]. The definition also conflates science and scholarship in a way that is both common and unhelpful. For many readers it would exclude work from the social sciences and humanities, as well as book chapters and other drafts destined for venues beyond "a peer-reviewed scientific journal".

Other organizations have constructed their own meanings and terms to fit the agenda of their constituencies. SHERPA, a UK organisation dedicated to studying scholarly communication, has a more precise definition for preprints: "the version of the paper before peer review"[3]. They then define versions between acceptance and publication by a journal as "post-prints." NISO (National Information Standards Organisation) doesn't formally define the word "preprint" in its Journal Article Version (JAV) standard[2], preferring instead to further delineate where "significant value-added state changes" occur. They break down the broad Wikipedia definition into four distinct stages, including "author's original", "submitted manuscript under review", "accepted manuscript" and any numbers of "proofs" that may emerge between acceptance and the published "version of record", a term which suffers under the dual burden of being both essentially undefinable and highly politicised.

As a further complication, the shifting roles of different players in the ecosystem have also contributed to this confusion. To "publish" a work can mean three entirely different things: the labour of preparing a work for its dissemination, to communicate or make public a work, or in the narrow sense we use in the academy, to make available through designated channels after specified social and technical processes. "Preprint" is positioned and often defined in relation to "publish", in a way that adds to the ambiguity of both terms.

---

[a]It should be noted that this is not a new problem. For many years researchers have bundled everything that has not been formally "published" under the umbrella term "grey literature", creating headaches for every meta-analyst and systematic reviewer who has had to decide what "counts" as a meaningful academic contribution.

In the past, there was a clear distinction between services that hosted preprints and "publishers" who carried out the formal process of "publication", as defined by scholarly communities. A preprint could therefore be identified by its presence on a platform that was not that of a "publisher". But today, publishers are starting to provide repositories to host preprints (PeerJ, Elsevier/SSRN, and the American Chemical Society). To add to the confusion, new forms of journals that run quite traditional quality assurance and review processes are being developed, which use preprint servers as the storage host for their articles. *Discrete Analysis* and *The Open Journal* both use ArXiv to store the PDF versions of accepted papers. A definition that depends on the historical role of any given player will fail if that role changes. Attempts to define the term "preprint" in this way pushes the confusion onto other terms that are equally poorly defined. Saying a preprint "is not published" or "is not in a journal" merely shifts the ambiguity to the question of what "published" means or what counts as a "journal."

The lack of clear definitions is a problem when discussing and negotiating important changes to research communication. Researchers today can share results earlier, in new forms, and to new communities. But the newness of such technologies means that we have not yet come up with terminology to clearly discuss the available choices. Some researchers simply see a preprint as an early notification or preview of a "formal" publication. For others it is a complete finding and a clear claim of priority in the scholarly literature. These differences are most often due to differences in disciplinary cultures. And, as in our story, the confusion is even greater with work that crosses disciplinary boundaries.

At the core, we have a fundamental issue of what "counts", and what counts will clearly depend on the community doing the counting. This is the central social and political issue on which disagreements on the status of preprints are based. We will never agree on a universal definition because communities naturally value different things. So are we fated to build walls between disciplines, between Maxwell and Darwin's tribes, never to be scaled or crossed? As research itself brings together different perspectives and types of knowledge to work on shared intellectual questions, we want to break down, not build up walls. We can in fact fruitfully engage across disciplinary boundaries and have productive discussions about preprints and the value of different kinds of scholarly communication. But to achieve this we must recognise when our differences are matters of fact (what process has an object been through) and differences in opinion and values between communities.

We present a model that will tease out one of the fundamental issues we've witnessed when research communities assess what will count and why. We do not propose a new vocabulary nor a new universal definition of preprints. This would only further contribute to our current confusion and complexity. However, our conceptual framework offers practical paths for publishers, service providers, and research communities to consider and implement, all of which will facilitate more effective discussions and better communications systems.

The State-Standing Model
State vs. Standing

While "preprints" is a referentially opaque term that make little sense in the context of an online communications environment, it is unlikely we will persuade anyone to abandon the term. Instead, we seek to tease out two attributes often elided when discussing objects in scholarly communication: "state" and "standing". We use the term "object" so as to be inclusive, as well as to avoid the further use of terms tied to obsolete technologies (see Box 1).

Box 1. Attributes of a Research Object

State - the external, objectively determinable, characteristics

Standing - the position, status, or reputation

The "**state**" of a research object is comprised of the external, objectively determinable characteristics of the object. This includes records of claims made about the object, metadata, statements of validation processes the object has undergone, etc. An object submitted for peer review undergoes a wide array of state changes as multiple players interact with it in the process of submission and publication: technical checks and validation, editorial assessment, assignment of editor and reviewers, referee review, editorial decision, typesetting, author approval and corrections, publication accept, content registration/metadata depositing, front matter editorial posting, publication commentary facilitation, retraction/correction processes, publication event tracking, etc. This includes explicitly modelled metadata elements within strong schema (such as "indexed in PubMed"), as well as unstructured and vague terms. It also includes a description of groups that have access, including "the public"[3]. With "state," there *can be* an explicit record made even if it is not exposed. Such records may be hidden within publisher systems or may even be private information that is unethical to share. The record might be in third party systems, such as Pubmed Central or ORCID. Some elements may be badly recorded or lost and thus inaccessible.

If an object changes state, it may also undergo changes in perceived value or intellectual status. The "**standing**" of a research object is the position, status, or reputation of an object. It is a consequence of its history and state. There are various forms of standing recognised by different groups, for example: "has been validated by (a traditional) peer review process", "establishes priority of claim", "is appropriate for inclusion in this assessment process," "is considered appropriate for discussion and thus citable", etc. These are judgments about the recognition or value of the output. Standing is conferred by a group, not an individual, and is therefore distinct from any individual's opinion of the work[b]. It is also conferred not directly to individual objects, but to classes of objects that share attributes of state.

---

[b] We use the general term "group" to refer to communities, institutions and other parties that confer standing. "Groups" therefore includes disciplinary communities, universities (and their departments), funders, but also potentially entities such as main stream media venues, as well as specific publics.

Nonequivalent states, nonequivalent changes

With a conceptual barrier between state and standing in place, we can investigate their relationship as the scholarly output changes over time. A state change may lead to a change in standing, but not necessarily and not in all cases. A change in standing, however, only occurs as a consequence of a state change triggered by some external shift that has led to a reconsideration of value.

Standing is independently conferred by each group for whom the research output has meaning. While *similar* forms of standing between groups might arise, they cannot be identical as such. What matters most in this model is the possibility that a particular community may confer a different form of standing than another on the same type of research object (i.e., with the same state).

Figure 1 illustrates how changes in the state of research objects may result in *different* changes of standing between two communities: physics and life sciences. Both may consider research validated and part of the formal record at similar stages of the publication process. But there are also key differences. When a preprint is posted by a physicist, they have established the priority of claim in that community, and it is considered worth of citation. However, for the life sciences community, claim priority is generally established when a manuscript[c] is submitted to a journal. It is only appropriate to cite the article even later, when the text is made available online (Advanced Online Publication or online publication).

The conditions that prevail in the conduct of research are naturally tied to the type of research itself. As these vary widely, so would the influence they have on the communication culture of the group and how they confer status. That certain fields in

---

[c]We acknowledge that 'manuscript' is as much a retronymic synecdoche as 'preprint'. However, we use the term here as the most appropriate in context.

physics share equipment, work in very large groups, etc., has been often mentioned as a contributor to their predilection for preprints. On the other end of the publication event, research may expand its reach and utility beyond the academy. This introduces other possible entities that begin to serve as a conferrer of status (e.g. university office of technology transfer), and it will vary by field and discipline depending on the opportunities possible. Both Maxwell and Darwin are awarded for their work in acknowledgment of their contributions, but given that the research was taken up by the physics community earlier, it would not be surprising to see time differences in the subsequent accolades offered to each by their respective disciplines.

Applying the model to preprints

Prior to the development of the web, some segments of both Economics and High Energy Physics communities shared a similar practice, the circulation by mail of manuscripts to a select community, before submission for formal peer review at a journal. As the web developed, both communities made use of online repositories to make this sharing process more efficient and effective. Paul Ginsparg initially created ArXiv as an email platform, but then migrated it onto a web-based platform in the early 1990s. In 1994, two economists created the Social Sciences Research Network (SSRN), a platform that shared many traits with ArXiv. In both cases, researchers submit digital manuscripts, which undergo a light check prior to being made publicly available on the platform. These manuscripts have not been subjected to any formal version of review by expert peers. Furthermore, there is a common expectation in both repositories that most manuscripts will go on to be formally published as journal articles or book chapters. That is, the *state* of objects in both ArXiv and SSRN is very similar.

Nonetheless the *standing* of these objects for these two communities are quite different. For the High Energy Physics community (and others in theoretical physics), posting to ArXiv establishes the priority of claims and discoveries. In many ways, ArXiv preprints are seen as equivalent to formally published articles, and
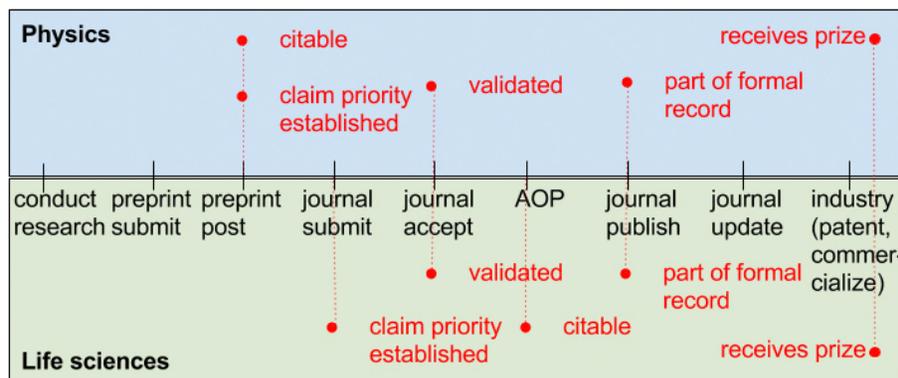

Figure 1. Differences in standing (red) between Physics and Life Science communities. AOP stands for advance online publication.

many physicists will preferentially read articles at ArXiv rather than find copies in journals. Indeed, for those disciplines where use of ArXiv is common, the formal publication is the point at which citations to the manuscript start to drop off[4]. The question of why physicists continue to publish in journals at all is a separate one and beyond the scope of this article. However, our model can help: clearly the community, or communities that matter, do grant some standing to journal articles, which is both different to that granted to preprints and important in some way. The question of what that standing is and why it continues to matter is separated in our model from the equivalence of state that journal articles in physics share with those in other disciplines. As Maxwell and Darwin found in our story, physics and biosciences are different in important ways, even when their publication processes are very similar.

By contrast, working papers on SSRN are seen much more as works in progress. They are frequently posted well before submission to a journal, unlike ArXiv where posting is frequently done at the same time as submission. Observers from outside these communities, including those interested in adopting physics posting practices for the biosciences, often make the mistake of seeing two similar repositories with similar requirements and assume that SSRN working papers and ArXiv preprints can be equated. The differences are not obvious from an examination of *state*, but are situated in differences in standing. Working papers and preprints have a different *standing*, and serve quite different functions for their cognate communities, despite being quite similar in form. Separating the two concerns allows us to be much clearer about what is similar and what is different between the two cases.

Further applications in the publishing life cycle

The uses of our model are not limited to preprints. It is a useful heuristic for isolating the questions that require answers from a community, from those that can be answered by auditing the process an object has been through. That is, it is helpful to separate the question of whether something has been done, from the question of whether any community cares.

We believe this separation of concerns will be valuable for discussions on a wide range of outputs, including software, data and books. Indeed all types of research outputs go through processes of validation, dissemination and assessment, which are accorded differing degrees of importance by different communities. Discussions of the details of options for differing modes of open, signed, partially open, single blind, double, or even triple blind, peer review will benefit from separating the description of process (and testing whether the stated process has been followed) from the views of any given community of objects that have been through that process.

Until recently much work on peer review was done within disciplines with little comparative work. The role of peer review processes in community formation is now gaining greater interest, as is the detailed study of the history of peer review processes. Some communities have strong affiliations with double-blind peer review processes, and some with single-blind, or

increasingly non-anonymous or signed reviews. Today, questions are raised as to whether processes that do not blind referees to author identity (a process described by specific state changes) can be expected to be unbiased and therefore valid (a question of standing). Pontille and Torny[5] in examining the complex history of these views quote Lowry[6] to showcase the view that "...a man's[sic] name is important and...can be used as a basis for judging the reliability and relevance of what he says". Separating the value-laden discussion of what judgements are necessary or reliable from the details of the process that support them can help to uncover and illuminate effective paths forward in deep-seated disagreements.

It may be the case that much of the confusion around newer forms of scholarly sharing, including efforts to make certain scholarly outputs "matter" as much as traditional narrative publications is due to a similar confusion. New forms of output seek to co-opt the expression of forms of state, without putting in the required work that connects the social machinery of state-standing links. As a result, they frequently fall into an "uncanny valley", objects that look familiar but are wrong in some subtle way. The most obvious example of this are efforts to make new objects "citable", i.e. making it technically feasible to reference in a traditional manner through provision of specific forms of metadata, most commonly via DOIs. To actually shift incentives, this work needs to be linked to a social and political shift that changes a community's view of what they should cite, i.e. what gives an object sufficient standing to make it "citation-worthy".

A similar debate is that which rages between traditional publishers and advocates of a shift towards "publish-first, review-later" models of research communication. On one hand, advocates of change often remark on the seeming lack of improvement made to the *text* of an article through traditional peer review. For example, Klein *et al.* found that text content of ArXiv preprints only undergo minor changes between the initially submitted and finally published versions[7]. Of course this neglects state changes in the validation process that may be important, but are not necessarily reflected in the character-stream of the article, such as ethical or statistical checks that were managed by the publisher.

On the other hand, publishers have established practices that they consider important, captured in the JAV vocabulary[2]. JAV details a number of different stages (with different states) that a manuscript might undergo. Many of these are invisible to authors. For instance, Author Original and Submitted Manuscript Under Review are identified as distinct states. An author would consider these to be the same document, but a publisher needs to record the manuscript's transition into the peer review pipeline. At the same time, JAV ignores changes that are likely of concern to authors by failing to record them. For example, it has no concept of the distinct revised versions of a manuscript submitted during review cycles.

This distinction may be useful in looking backwards as well as forwards. A growing interest in the history of scholarly communications reveals that processes of selection and publication in the 18th, 19th and early 20th century could be very different to our current systems. For instance, Fyfe and Moxham[8] discuss a shift

in process at the Royal Society in the 19th century. They trace "a transition from the primacy of [a paper being read at] face-to-face scientific meetings...to the primacy of the printed article by the end of the nineteenth century". The processes changed, as did the status granted them. Presumably our current views of standing, and their ties to current processes of state change, evolved together. Separating the processes and state changes from the standing granted by historical communities (if this can in fact be determined from archival records) can only help us to understand how our current processes and values evolved.

It is also not just deep history that could find the distinction helpful. The primacy of the reading of a paper at a meeting will be familiar to many scholars in Computer Science, where conference proceedings remain the highest prestige venue for communication of results. The state changes of an object in computer science have some similarities to the historical state changes at the Royal Society. An examination of how similar standing is in these two cases, and more particularly how the primacy of conference presentation arose in Computer Science could benefit from analysis in terms of our model.

Here, the issue is a difference in focus on what it is that matters, what kinds of standing are important. Changes in state that are important markers of shifts in standing for one group are ignored by the other and vice versa. Until the full set of state changes that are relevant to all stakeholders are transparently visible, discussions of standing are unlikely to be productive.

This illustrates a crucial point. Our model exposes the need for high quality metadata that is well coupled to the record of processes that a work has experienced. If what is contained within the scholarly record is a question of standing, then the formal record of state is a critical part of supporting claims of research.

Conclusions

To engage in productive discourse on new (and traditional) forms of scholarly sharing, we need to gain clarity on the objects themselves. We propose a model that explicitly separates the *state* of a work – the processes it has been through and the (objectively determinable) attributes it has collected through those processes – from the *standing* granted it by a specific community. It is not only a formal framework, but a practical apparatus for navigating and negotiating the ongoing changes in scholarly communications. By distinguishing two attributes we can isolate aspects of objects that can be easily agreed on across communities, and those for which agreement may be difficult. These have clouded discussion of community practices, particularly those around the emerging interest in "preprints" in disciplines that have not previously engaged in the sharing of article manuscripts prior to formal publication.

In proposing this distinction, we are foregrounding the importance of social context in the community-based processes of scholarly validation. The importance of social context in scholarly processes is of course at the centre of many of the controversies of the late 20th century in Philosophy and Sociology of Science, Science and Technology Studies, and other social studies of scholarly and knowledge production processes. Our proposal follows in those traditions. In our model, what is to our knowledge novel is that it provides a way to link the conversations, focusing on process and metadata, that occur when researchers and publishers discuss scholarly communication, with the social context that they occur in. By connecting state and standing, and recognising that each has an influence over the other – state directly on standing, standing by privileging certain changes of state – we aim to show how the intertwined relationship is at the core of conferring value across scholarly communities.

How does our model help the young Darwin and Maxwell? Well, it makes explicit the changing nature of discovery across disciplines, and provides a way of differentiating between changes to the object and changes to the perception of the object. Questions of standing will be inherently difficult to discuss across community boundaries, and while the model cannot solve the underlying *social* challenge that different research communities simply value different things, and in particular different parts of the overall life cycle of a research communication, it does offer a way of talking about and analysing those differences. To bring the culture of manuscript posting to the biosciences, Darwin would be better served by identifying the different goals that different people had as well as discuss the concerns that more traditional researchers have.

Our model does not, and cannot, solve the problem of differing perspectives between research communities. It does, however, have clear implications on how the various players in research communications can better contribute to an effective and efficient conversation:

Publishers, including preprint repositories, can better serve their communities by making state changes much clearer, more explicit, and transparent. It is impossible for us to make progress in discussing standing when we cannot clearly define what the state is. We cannot discuss the difference in standing between a preprint, a journal editorial, and a research article without knowing what review or validation process each has gone through. We need a shift from "the version of record" to "the version with the record".

Service providers, including publishers and repositories, but also those that record other processes, need to pay much greater attention to recording state changes. Currently many records of state are focused on the internal needs of the service provider rather than surfacing critical information for the communities that they serve. Principled and transparent community evaluation depends on a clear record of *all* the relevant state changes.

Finally, **scholarly communities** must take responsibility for clearly articulating our role in the validation and to recognise that this is a fundamentally social process. It is our role to grant standing. We need to explicitly identify how that standing is related to a clear and formal record of changes in state. The current discussion arises due to confusion over the terminology of preprints, but the issue is much more general. By making explicit both the distinction between social processes and the record of attributes that results from them, and explicitly recognising the connection between state and standing, we will surface the processes

of scholarship more clearly, and re-centre the importance of our communities deciding for themselves what classes of object deserve which granting of standing.

Author contributions

All authors contributed equally to the conception, authoring and editing of this paper.

Competing interests

Grant information

References

1. Wodak SJ, Mietchen D, Collings AM, *et al.*: **Topic Pages: *PLoS Computational Biology* Meets Wikipedia.** *PLoS Comput Biol.* 2012; **8**(3): e1002446.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. NISO/ALPSP Journal Article Versions (JAV) Technical Working Group: **Journal Article Versions (JAV): Recommendations of the NISO/ALPSP JAV Technical Working Group.** NISO Recommended Practice 2008–08. 2008.
   Reference Source

3. Neylon C, Pentz E, Tananbaum G: **Standardized Metadata Elements to Identify Access and License Information.** *Informations Systems Quarterly.* 2014; **26**(2): 35–37.
   Publisher Full Text

4. Gentil-Beccot A, Mele S, Brooks T: **Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories.** arXiv: 0906.5418v2. 2009.
   Reference Source

5. Pontille D, Torny D: **From Manuscript Evaluation to Article Valuation: The Changing Technologies of Journal Peer Review.** *Hum Stud.* 2015; **38**(1): 57–79.
   Publisher Full Text

6. Lowry RP: **Communications to the editors.** *Am Sociol.* 1967; **2**(4): 220.
   Reference Source

7. Klein M, Broadwell P, Farb SE, *et al.*: **Comparing Published Scientific Journal Articles to Their Pre-print Versions.** arXiv: 1604.05363v1. 2016.
   Reference Source

8. Fyfe A, Moxham N: **Making public ahead of print: meetings and publications at the Royal Society, 1752–1892.** *Notes Rec.* 2016; **70**(4): 361–379.
   Publisher Full Text

Open Peer Review

Current Referee Status: ✔ ✔

✔ **Todd Carpenter** (iD)

National Information Standards Organization (NISO), Baltimore, MD, USA

There are many challenges in our current scholarly communications and assessment environment. This article draws out an important distinction between the objective status of a piece of scholarly content and the value to which the community assigns to that content. This is an intricate, intertwined, and sometimes confusing interplay between the two concepts. The authors do a commendable job of describing the current state and outline potentially valuable model for distinguishing between the two concepts.

While the article does have much to recommend, I have several areas of concern. First, the article provides a quaint description of two scholars in two distinctly different fields of inquiry, physics and biology, who collaborate on a research project. One of the researchers, the physicist, receives "credit" for a joint paper, while the other, the naturalist, receives none. The article takes the reader on a journey where one researcher succeeds while the other fails because of a lack distinction between the two social responses to the same content form. This illustration describes the sometimes critical differences between domains and the different weight given to forms of distribution and publication. In the article's conclusion, the authors note that their framework isn't meant to address the social differences between domains that are at the root of these differences, simply to describe them. While distinguishing between "state" and "standing" might provide some method to identify the objective and subjective status of a content object, the article lacks consideration of the criteria or suggestions about what characteristics might contribute to their notion of standing. At the heart of the illustration is an environment where different domains confer different meaning or value, the objective status may or may not influence the subjective response. Distinguishing between the two seems obvious.

While the distinction between "State" and "Standing" as described in the framework appears a useful distinguishing characteristic, it is not clear to me that the examples of "state" changes are in fact "objectively determinable characteristics of the object" that are intrinsic to the object itself. There is no way to know by examining the object whether it has undergone any particular state change. To consider a real world example, take an article in ArXiv (https://arxiv.org/abs/1509.06859v2) by Sébastien Gouëzel (LMJL). This paper has an earlier version (https://arxiv.org/abs/1509.06859v1) and was updated with the current version in May, 2017. Viewing this from ArXiv, there is no indication that this object has gone through any vetting, nor any editorial review, nor any validation processes, nor any copy editing, nor any of the other state changes mentioned in this paper. However, the paper has been included in the online journal *Discrete Analysis* . There is an editorial introduction with a DOI (10.19086/da.1639, which oddly didn't resolve) and it isn't clear that the journal "publishes" the article or the introductions. Presumably the article itself went through a peer review, an editorial review, and possibly edited and then was revised and resubmitted to ArXiv. The date on the ArXiv revision is 5 May 2017, four days before the Discrete Analysis

paper was posted on 9 May 2017. If a user views this article through the wrapper of journal, it may be clear that these "state changes" as defined in this paper might apply, but the same content viewed through the lens of the paper directly on ArXiv they are not. The state changes exist in one environment but not in another.

The authors respond to this situation, as they note in their conclusion, that this is simply a failing of metadata and that if only the state changes were recorded in the metadata, this problem might be addressed. The problems of metadata quality is well known and much discussed in the community. Properly assigning metadata to a final version of record is challenging enough, without retroactively populating metadata or ensuring a string of provenance data is included with the current object to support this chain of awareness of the current state of a content object. For example, if someone were to come across the first version of Gouëzel's paper in the example I noted, how would anyone know the current state of the previous version? Without forward linking, there is no way to know that the preprint version (or authors original, using the JAV terminology) was followed by another version?

With this example in mind, the paper would be strengthened through a more robust description of state changes, and what would distinguish a state change from something less substantial. Since many of the changes that would might take place to an article may or may not be significant. Also, many state changes might not lead to notable changes. For example, (in a closed peer-review environment say), I may have read this article without recommending any changes. The act of reading and saying "Yes, this is OK", is completely external to the object and failing quality metadata to describe the review. These external acts related to a piece of content are critical to the process of developing standing, but aren't necessarily externally obvious, as the authors note.

A minor point about standing to which I quibble is the notion that standing is not something that can be conferred individually. There are many instances when standing could be individually confirmed. Many journals are editorially run by a single individual, who might review, take a decision to publish or not, or approve for publication. A department chair, may determine that a piece of content is appropriate for inclusion in a promotion or tenure decision. I am sure there are countless other examples of this. One might say the editor is speaking on behalf of a community of subscribers, but in reality it is just one person taking the decision.

In practice, the framework outline in this article builds upon the structure outlined in the NISO Journal Article Versions (NISO JAV) Recommended Practice, which defined a structure of changes for the constrained scope of journal articles. That effort settled by "identif[ing] a significant value-added "state change" in the progress of a journal article from origination to publication." While these state changes in NISO JAV are explicitly focused on the formal publication process, the concept of state change applies across all forms of content, again as noted in the introduction of NISO JAV. It should also be noted that the working group that developed NISO JAV structure intentionally did not extend it's scope to other forms of content, nor to extend the resulting vocabulary to every instance in the content creation process.

This article, or potentially in subsequent work by the authors, would be strengthened by a discussion of the types of elements that go into standing. The description of potential changes to a content object's "state" are comparatively robust, but the components of what constitutes "standing" is decidedly weaker. Especially since this appears to be the core argument of the need for this framework, this lack of discussion around those details glosses over the difficulty in that side of this environment.

Inherently, this second domain of the meaning of and definitions of "standing" are incredibly fraught and fungible across the academy. What has standing in one domain does not in another, often without rhyme

or reason. There is no fault in the authors avoiding these very granular and thorny questions in this article, nor does it diminish from the value in trying to distinguish between the two. However, without the understanding of "standing" there can be no resolution to the problems that Darwin faces in the article's opening illustration.

I look forward to the continued discussion around these issues and encourage the authors to continue to develop their work in this area.

Is the topic of the opinion article discussed accurately in the context of the current literature?
Yes

Are all factual statements correct and adequately supported by citations?
Yes

Are arguments sufficiently supported by evidence from the published literature?
Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?
Yes

Competing Interests: CrossRef is a member of NISO and an occassional sponsor of NISO events. At various times, Neylon, Bilder and Lin have participated in NISO initiatives. The NISO Journal Article Versions (NISO JAV) Recommended Practice, note in the article and the review is published by NISO.

Referee Expertise: Information science, standards, metadata

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 19 May 2017

Kathleen Fitzpatrick 🆔
Modern Language Association, New York, NY, USA

This article focuses on the need (most immediately seen in recent discussions of the challenges presented by "preprints") for distinguishing between the ways we represent the state of research outputs and the ways we represent the standing those outputs have in accordance with that state. The importance of this distinction, as the authors point out, is in recognizing that different communities of practice accord different value to outputs that are objectively in the same state. By "separating the description of process (and testing whether the stated process has been followed) from the views of any given community of objects that have been through that process" (5), we might be better able to speak across disciplinary boundaries about the value of the work we do. The distinction between state and standing is especially crucial for those who seek to change scholarly communication practices, for instance by allowing a greater range of outputs to "count" in hiring and assessment processes, in order to make clear that the transformation that matters lies in "the social machinery of state-standing links" (5). By encouraging the disambiguation of state and standing, the authors are able to advise publishers, platforms, and scholarly

communities on ways they might contribute to better conversations about the value of particular research outputs.

The article begins with a highly engaging opening illustration of the stakes of the non-universality not only of language but of practices in scholarly communication, and continues through careful and well-documented argumentation. The authors look carefully at distinctions not just in terminology but also in values across different fields. They are careful to note that they are not recommending a new vocabulary, nor a guiding framework for how scholarly communities should negotiate the current changes in their communication practices, but they do our fields a great service by exposing the reasons for much of our mutual incomprehension across fields. They also go a long way toward explaining why "we have to make preprints 'count'" is a very heavy lift in some communities.

I do hope that the authors will continue their research in this line. It would be great to have their input, for instance, into the construction of metadata that can help clarify changes in a research object's state, enabling better judgment in communities about its standing.

A few very small copyediting notes:
- The phrase "different to" is used a couple of times; I'm honestly not sure if this is a UK/US distinction, but I'd argue for "different from" instead.
- On page 3, column 2, line 1, "that make" should be "that makes".
- On page 4, AOP is glossed as "advance online publication" in the caption for Figure 1, and "Advanced Online Publication" in the text.

I am grateful for the opportunity to have reviewed this article, and I look forward to the discussions that it might inspire.

Is the topic of the opinion article discussed accurately in the context of the current literature?
Yes

Are all factual statements correct and adequately supported by citations?
Yes

Are arguments sufficiently supported by evidence from the published literature?
Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?
Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Scholarly communication, humanities

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.