

The importance of a preprint repository

Praveen Chaddah

On-line communication is changing the way in which scientific research is communicated and reviewed. Dissemination has become easy and fast, and established journals with deliberate reviewing processes are becoming a means for validation, and not so much for dissemination, of research results. Since the best of journals do retract papers for fraudulent or fabricated data, the model of quick dissemination of preprints followed by review and cross-checks by all interested subject experts, does present an alternate to limited pre-publication reviews imposed by academic journals. The profitability of academic journals is also being discussed¹, as the more traditional journals are funded by subscriptions from libraries of research institutions, while open access publishing is heavily funded by grants received by the authors. In India, both routes can, in most cases, eventually be traced back to funding from the Government. Discussion of the economic aspects is beyond the scope of this note, except to emphasize that dissemination by preprint servers is free to readers as also, so far, to authors. The cost of operating a preprint server is, for comparison, estimated at US\$ 800,000 in 2016 (ref. 2).

The impact of preprint servers on traditional journals is highlighted by the growth of the physics preprint server arXiv.org. A recent editorial in *Nature Physics*² asserts that 'Despite a sadly persistent misconception, we at the *Nature Research Journals* love arXiv. We encourage authors to post their manuscripts before formal submission and update the final version six months after publication.' The misconception is not entirely unjustified because *Nature* does not allow uploading of manuscripts modified after submission, whereas many journals allow uploading the post-refereeing and accepted manuscript version also. There is a view that preprint servers and on-line versions of traditional journals cannot coexist for long and preprint servers will dominate research dissemination, whereas the editorial above states 'we not only imagine coexisting with arXiv, but also hope for better integration'. It is clear that the use of preprint servers can only grow, and these provide a means for archiving and rapid on-line sharing of results.

Unfortunately, there are no visible preprint servers in India, despite calls for allowing preprints on national repositories³.

The DBT and DST Open Access Policy discussed in ref. 3, had called for manuscripts to be deposited in institutional repositories or in designated central repositories (dbt.sciencecentral.in and dst.sciencecentral.in) after acceptance for publication by the journal, but subject to embargo imposed by the publisher. In the case of *Nature* research journals, this is allowed only six months after publication. Uploading on a post-print repository allows free archival access, but does not help in claim of priority of initial dissemination because these provide visibility well after the journal has already disseminated the paper. More information on publishers' policies on self-archiving preprint and post-print research papers is available at <http://www.sherpa.ac.uk/romeo/>.

Some recent events highlight the need for encouraging our researchers to upload preprints on repositories of high visibility, though some benefits will follow by uploading on any repository that is accessed by popular search engines. The benefits that authors seek while disseminating are that their paper should have high visibility, and also be easily accessible to other researchers. These appear to be conflicting requirements in traditional journals as high impact journals usually have a high subscription. The repository arXiv.org has become so successful because of its committed daily readership, with free readership access even without any institutional affiliation. The benefits this offers, especially for authors from emerging bylines as in India, are tremendous and will become obvious in the examples I shall discuss. The call made in ref. 3 for setting up national preprint repositories covering various disciplines will be reinforced.

Recently, there has been a Gazette Notification 'University Grants Commission (Minimum Standards and Procedure for Award of M Phil/Ph D degrees) Regulations, 2016' that has superseded the 2009 regulation. Clause 9.4 states that 'Ph D scholars must publish at least one (1) research paper in refereed journal and make two paper presentations in confer-

ences/seminars before the submission of the dissertation/thesis for adjudication'. The reference to presentations in conferences was missing in the 2009 regulation. Why does this change emphasize (to me) the need for setting up preprint repositories?

We do want students to participate in conferences where they discuss their initial results, and these must evolve into a publication in a peer-reviewed journal. Many of our national conferences now publish proceedings through 'conference series' of international publishers. Because of this, the authors have to worry about charges of self-plagiarism. Basically, if any of the data are to be incorporated in the journal submission, the journal manuscript must cite the conference paper which is 'published' in the 'conference series', even though the impact factor is near zero. There is a distinct possibility that the reviewers may then reject the journal submission claiming that there is not enough new data or analysis. The alternative of presenting very little data in the conference is not correct because it defeats the purpose with which the student attends the same.

Since the proceedings are usually published many months after the conference, the research scholar may submit the journal paper without being able to properly cite the conference paper, and the reviewer may accept the journal paper. This is dangerous because the overlap will be on record once both papers have been published, and can be detected by hobbyists or political opponents many years later, and damage a career. Some years ago, as commercial software for checking text-plagiarism was being enthusiastically used, there was a report in *Current Science* titled 'Publish and perish' that asked 'Should one's career be ended or marked forever due to a few misdeeds?'⁴. Can we ensure that these mandatory presentations at conferences do not raise a similar spectre through the route of self-plagiarism?

Both these scenarios can be avoided if the conference submission is uploaded on a preprint repository, and not published in the conference proceedings. As brought out above, an arXiv.org upload of a manuscript followed by submission

to a journal, is actually encouraged and will not be considered as self-plagiarism. In addition, arXiv.org allows much wider dissemination than the proceedings of most of our national conferences. We should actually discourage submission of manuscripts for publication in conference proceedings.

We must worry that our research output is not plagiarized or appropriated by someone who has access to our work before it is released in public domain. As we shall see, such misuse can occur even for manuscripts submitted to a journal, where access is restricted and can be traced. Many organizers require submission of the manuscript well before the conference and access to the manuscript is not as well controlled as for submissions to journals. There is a distinct possibility of unethical misuse in the time between this submission and the presentation in the conference. Since the authors uploading on the preprint archive decide when and how their research results are announced, they can ensure that this is done before submitting the manuscript to the conference organizers, and prevent any possibility of unauthorized usage of their results.

Such unethical behaviour was noticed when a manuscript submitted from JNU, New Delhi in 2005 was rejected, but significant portions of original text from the unpublished work appeared in another paper from the University of Bradford, UK, clearly showing unethical behaviour from someone having access to the submitted manuscript. IEEE issued a notice of violation of ethics, by D. Kouvatso and S. A. Assi in 2007. But the correction was incomplete since the original manuscript was not in public domain. Eventually, in 2010, the JNU group uploaded their manuscript on arXiv.org and the notice now stated ‘This paper contains significant portions of original text from the unpublished work of the authors cited below. The original text was copied without attribution and without permission. “Long Tail Behavior of Queue Lengths in Broadband Networks: Tsallis Entropy Framework” by Karmeshu Gupta and Shachi Sharma, July 2005 <http://arxiv.org/abs/1012.2464>.’ While this brings out highly unethical behaviour in which Indian researchers (Sharma was a doctoral student) were victims, it also highlights the role of a preprint archive in ensuring corrections to unethical behaviour.

There is sometimes an uncomfortable feeling that reviewers of journals are more skeptical towards new ideas from emerging bylines, and many experience that the time interval between the first submission of a manuscript from an emerging byline and its final acceptance is longer than average. In one such case we uploaded our manuscript on arXiv.org because the initial reviewer comments were negative⁵. This gave us a six-month lead over our journal publication, and some visibility. Some years later, when these ideas were accepted, an established group made us victims of idea-plagiarism and usurped our idea in their publication in a reputed journal. The journal did get those authors to publish an apology, but the correction was not prominent. Since we had used arXiv.org, and so did those authors, we brought this to their attention. A search with the name of any of the errant authors now gives the title of their paper with the entry ‘Comments: 17 pages, 5 figures. Erratum Doi: 10.1103/PhysRevB.84.059904: “The authors did not cite a relevant and important reference. The cooling and heating in unequal field (CHUF) protocol that has been used and described in Sec. III B 2 of this paper was first published in ref. 1 [arXiv:0805.1514]. We apologize for this omission”.’ Another example that highlights the role of a preprint archive in ensuring corrections to unethical behaviour.

We now present a recent example of how this preprint archive is helping an Indian group resist unethical attempts to usurp their discovery.

A group from IISER Mohali, and collaborators from Mohali and Delhi, reported⁶ (the paper was released on arXiv.org on 9 October 2014) a claim of unconventional superconductivity in the material Cd₃As₂. On 2 January 2015 a Chinese group reported similar results on arXiv.org, on the same material⁷. Since the first positive results were known, the Chinese group put in efforts to grow single crystals. It is normal in materials research that major new results are first discovered in polycrystalline samples, or even in samples that may have coexistence of multiple compounds, as happened in the Nobel Prize-winning discovery of high-temperature superconductivity.

Both the groups also submitted manuscripts to the same journal where they eventually appeared in the same issue,

viz. *Nature Materials*, vol. 15 and both were published on-line on 2 November 2015. The two papers were published simultaneously, but had been received by this journal at very different times. The journal records that the Indian manuscript was received on 21 October 2014, whereas the Chinese manuscript was received on 11 January 2015. This confirms the three month precedence for the Indian group’s discovery that they had anyway established with their October 2014 preprint on the widely read preprint server arXiv.org.

In a recent paper⁸ uploaded on arXiv.org in July 2016, the Chinese group has claimed that ‘We two groups independently studied Cd₃As₂ by using hard point contact spectroscopy and reported tip induced superconductivity on Cd₃As₂ in *Nature Materials* simultaneously’. The reporting was clearly three months apart. The claim of simultaneous reporting by the Chinese group, against the actual simultaneous publishing, is an attempt at unethical usurping of credit. Fortunately, arXiv.org provides hard evidence of priority in reporting, as has been highlighted in the rebuttal on arXiv.org by the Indian group⁹.

To conclude, preprint archives allow presentation of detailed results in conferences, and follow-up submissions to journals, without the possibility of a charge of self-plagiarism. They also protect against unethical attempts to usurp priority of ideas or results. Finally, as shown by three examples involving Indian groups, a preprint archive allows correction for unethical attempts to usurp credit by ignoring earlier work. Thus we should create national preprint repositories.

1. Balaram, P., *Curr. Sci.*, 2011, **101**, 1259–1260.
2. *Nature Phys.*, 2016, **12**, 719.
3. Chaddah, P., *Curr. Sci.*, 2016, **110**, 294–295.
4. Malhotra, R., *Curr. Sci.*, 2011, **101**, 476–478.
5. Banerjee, A. *et al.*, 2008; arXiv:0805.1514.
6. Aggarwal, L. *et al.*, 2014; arXiv:1410.2072.
7. Wang, H. *et al.*, 2015; arXiv:1501.00418.
8. Wang, H. *et al.*, 2016; arXiv:1607.02886.
9. Gayen, S. *et al.*, 2016; arXiv:1607.01405.

Praveen Chaddah lives at Flat 702, Block 24, Heritage City, Gurgaon 122 002, India.
e-mail: chaddah.praveen@gmail.com