The 2nd SPARC Japan Seminar 2018

"Quality Control in the Age of Open Science"

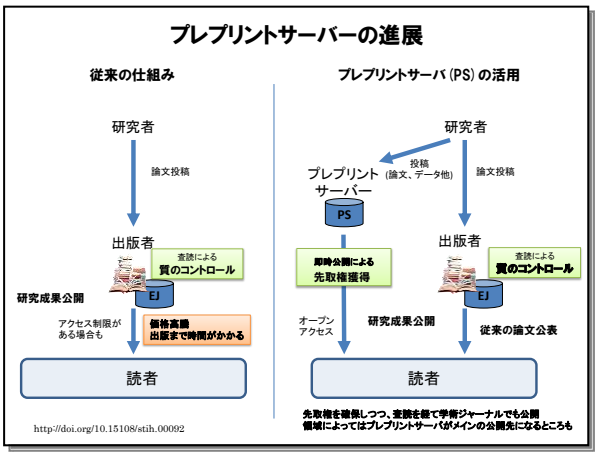
Panel Discussion

Kazuhiro Hayashi	(National Institute of Science and Technology Policy)
Rebecca Lawrence	(F1000)
Hideaki Takeda	(National Institute of Informatics)
Hidemasa Bono	(Database Center for Life Science)
Masamitsu Kuriyama	(Library and Academic Information Center, Tokyo Metropolitan University)

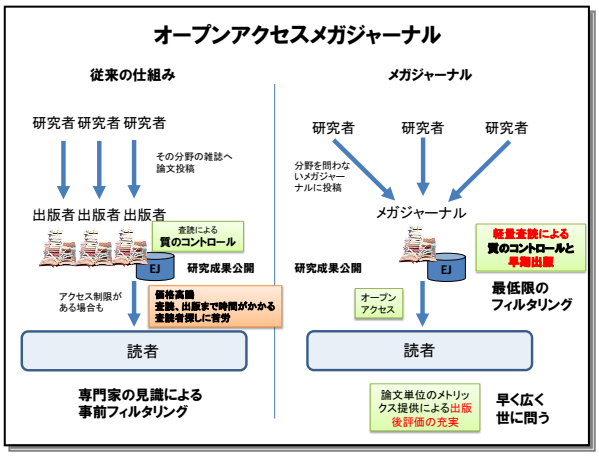
● Hayashi　Before the panel discussion, I would like to start by introducing the fundamental model that we are talking about today since we have many participants from various sectors again this time.

In general, researchers publish their papers in journals after a peer review. However, there are various challenges, such as soaring journal costs, difficult quality control, time-consuming review process, difficulty of finding a reviewer, and questionable review process. Therefore, there is a trend that researchers post their papers on a preprint server and make them open to secure the right of priority first (Figure 1). In the field of High Energy Physics, for example, they also publish their papers in journals with a peer review to get an extra track record. For areas with a short research cycle where they cannot wait for peer review, there is a growing trend of placing more emphasis on a preprint server. From there, it leads to today's theme, which is who will guarantee the quality of preprints.

Besides that, the open-access mega-journal is another game-changer (Figure 2). Instead of publishing papers in journals in the respective areas, this model is to publish papers altogether, and as introduced by Ben, papers are filtered minimally to wait for open evaluations. Some people argue how



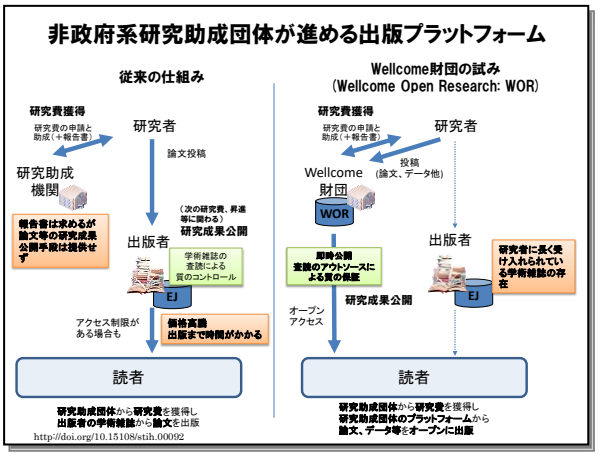
(Figure 1)



(Figure 2)

comparable the quality of such a peer review by the open-access mega-journal is with the quality of a peer review by traditional journals and how they are different.

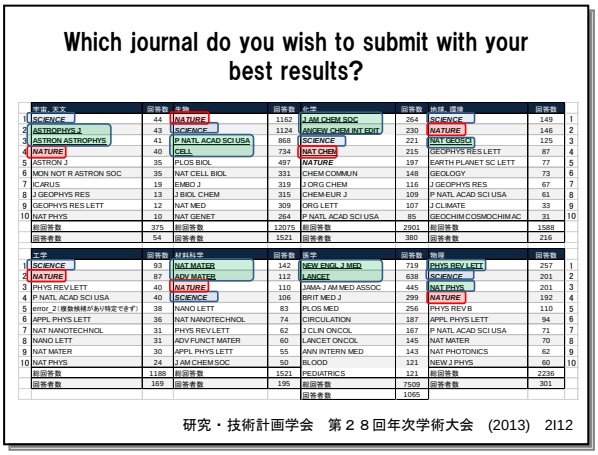
In addition, one of the main topics that Dr. Laurence talked about today was that research funding agencies have their own servers (Figure 3). In the past, they simply received reports, and the outcomes were published by existing publishers. In contrast, researchers publish on a server of, for example, the Wellcome Open Research, which is the equivalent of preprints first to get a time stamp to make it immediately accessible. After that, the peer review is so-called outsourced to the Faculty of 1000 for quality control. Once the paper passes quality control, it will be indexed in PubMed and other archives.



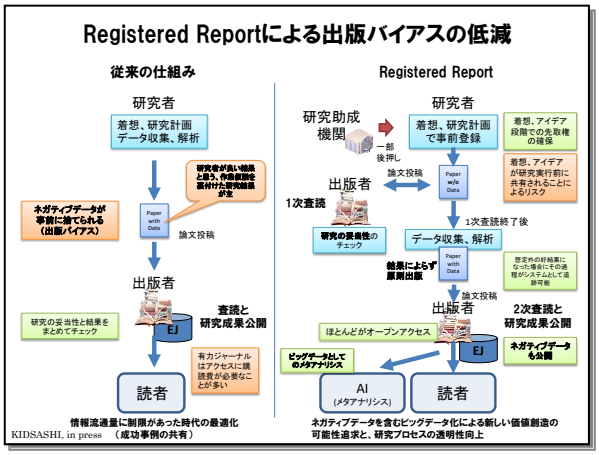
(Figure 3)

How well will this method work? According to Ben, we cannot ignore the presence of academic journals that have been accepted by researchers for many years and the impact factor. Dr. Bono also mentioned that.

In this respect, I have one back data, and I will show you since it is a great opportunity (Figure 4). In 2013, I asked researchers of eight areas, "Which journal would you like to publish if you achieve good results?" Basically, I wanted to know how many researchers support *Nature* and *Science*. As working hypotheses, I also think that the top journals published by international symposiums in the United States and European countries have a strong presence in some cases. In fact, depending on the area, I found that there were other journals in which researchers of each area want to publish their papers rather than *Nature* and *Science*. I felt a sense of relief in a sense. For researchers of engineering, earth and environment studies, on the other hand, it was found quantitatively that they prefer *Nature* and *Science*. These are the results we obtained from researchers from around the world, not just Japan. In reality, the Nature Index is created based on these data. There is a debate on how to handle preprint servers while there are



(Figure 4)



(Figure 5)

specific journals where researchers want to publish their papers.

Additionally, Dr. Laurence mentioned that it does not necessarily have to be published in journals any more. There is also a registered report (Figure 5), which is the most advanced in a way and as one of the attempts. It is to publish an experiment plan even before conducting any experiment. The articles are peer-reviewed prior to any data collection and then published after testing and collecting data, regardless of the positive or negative results against the working hypotheses. Such a medium is already in action in some areas. I understand it is mainly in clinical practice. In a way, this might be a gimmick that fundamentally breaks the journal systems. In a positive sense, there is a movement to researchers publish research results neutrally and openly after conducting research based on working hypotheses without the so-called publication bias.

The examples I have introduced so far could never be achieved with paper media. Who will guarantee the quality of these contents when researchers become able to place their knowledge on the web to share with everyone and making information accessible by anyone? Or, who will pay for the cost of communication? When a new model is in sight, how should things be changed for it? I provided an additional explanation because I would like to discuss on these matters from now.

Now, we will start the discussion. At first, let me ask Dr. Laurence about her feedback on the talks given by Prof. Takeda and Dr. Bono in the latter half.

●Lawrence Both of your talks about the preprints were great. arXiv particularly has led the way. It is interesting to see that despite the fact that a lot of physics has been preprinting for long time, there is still growth. In fact, more and more areas are getting behind preprinting. bioRxiv in biology are still covering tiny fractions in reality of all articles that are published, but again it is a significant growth. Preprints are a major shift forward, but it does not make sense to have preprints and journals separately. Layering on top of preprints things like peer review, curation, and other different types of checks that you might want to do is a better way forward.

●Hayashi Thank you.

I am also grateful that we have received many questions. There are surprisingly so many questions.

I have a question for Dr. Laurence. If rejected papers are made available for viewing, there are risks of scooping that other people publish a paper after correcting minor mistakes. How will such a problem of bad manners be handled?

●Lawrence Bear in mind, even if it is 'rejected' (or 'not approved'), it is published. You cannot be scooped on what you have published. If you identify issues, you can build on it and revise it. This is what science is about. We see something published, we tweak it. You think of an idea, try it, and then publish that. If you have generated new knowledge off the back of an article, that is fine. You should clearly follow standard scholarly practice and cite the original article, but I do not think that is a problem. That is not scooping. Scooping would be if you are trying to beat them to claim the same point. Just like with preprints, you get priority because you have already published it. There

is of course also benefit to replicating a finding and showing that it can be replicated successfully but it should be clear to all if this is what you have done.

● Hayashi I have a technical question for Prof. Takeda. Will you read and answer the question?

● Takeda Sure. About IDs, for example, there are arXiv IDs for arXiv, which are included to the bibliographic database of the arXiv community. However, although the arXiv IDs can be used and connected in the world of arXiv, people on the outside do not know. That is why not having DOIs is a disadvantage.

The reason why DOIs cannot be given in arXiv is due to system and money issues. CrossRef charges a dollar to add a DOI for each case, and it costs money to add the needed DOIs. In addition, corresponding metadata must also be provided. What is more, the system needs to be fixed, and that is why DOIs have not been added.

Someone asked that the system has not been fixed because of a money issue. For one, it is largely due to the financial front. We are now saving money to fix the system, but we still cannot fix it fully. Because the system became complex in the history of 30 years, another reason is it is quite difficult to switch the system to change the function.

As for the copyright of arXiv, authors have their copyrights because arXiv does not claim a copyright.

There is a question that says what flaws of metadata are. I do not know much about that. I heard a person in charge of the system saying enthusiastically that is something difficult.

The next question is important. It is about why papers are preferentially published in arXiv. For the traditional arXiv for physics papers, a preliminary report is published first, and then a final paper is eventually published in a journal, indicating that the arXiv has a role of being the first step. For computer science, there is the purpose of disclosing research before anyone at any rate. Some researchers publish their papers in a journal after publishing in arXiv. However, they publish their papers in arXiv before anything regardless of whether they publish their papers in a journal or not. What is important here is that people notice papers instantly once publishing in the well-known arXiv instead of publishing on their own websites.

Yesterday, I went to some computer science lab in Japan to give a lecture on open science to students. I also talked about arXiv there. A faculty member said to the students that "You always check arXiv anyway before doing anything, right?" Even in computer science, it is becoming customary to check arXiv first to look for similar research. In that sense, the incentives of publishing on arXiv are that researchers can publish their research early and it is easily findable by people.

● Hayashi Someone is asking Dr. Bono a direct question, "A paper on a preprint server will be acclaimed ?" What are people's reactions?

● Bono I think it depends on what it means to be acclaimed, but for me, it feels very good to find my research on a preprint server when I see it. The fact that papers uploaded on a preprint server will be acclaimed scholastically. Even so, I do not know if it will be acclaimed in other aspects. For me, I am simply satisfied that my papers have been uploaded. Simple as that.

● Hayashi There are some principles of evaluation.

This is a tough one. In the case of a doctoral thesis review, there is a rule that students are required to publish a certain number of papers with peer reviews. It means that preprints might be difficult for young researchers working on doctoral dissertations to use, which is another question. Dr. Laurence, will you comment on this?

●Lawrence Young researchers have not had a lot of time to produce a big study. Some will get lucky and find big findings, but others may not. One of the challenges that early career researchers (ECRs) have is that they may have found lots of smaller findings that are very difficult to publish but they also need to get on with their career and demonstrate what it is that they have done in the last couple of years. Again, the slow journal process makes that difficult. Preprints and models like ours where you can get the paper out and demonstrate something that you have done, and ideally if have it peer reviewed within the timeframe too so you have expert comments, really helps ECRs to achieve those points.

●Hayashi Is there anything you would like to say?

●Takeda Supposedly, requiring students to publish their papers in a journal for the review of doctoral degrees is essentially the abandonment of the job of the review committee. Although the review committee must review papers in the first place, they use external reviews instead, which means they are not doing their job.

So, essentially, they can simply evaluate papers based on the content. Journals should be originally used as supplemental information. In that sense, such journals can be used in a positive way. We will think about what to do if this is not the case.

As a reference, in the field of computer science, publishing papers in a journal is not highly regarded as a whole. My association only requires members to participate in a conference that fare strongly in conference rankings. I think the acquisition of a doctoral thesis is a good opportunity to review the original problem.

●Kuriyama Even so, we always tell students at my graduate school to publish their paper in a journal early. There is a similar trend for social science, even it is arts and sciences.

I mean no disrespect, but I would like to add something here. I think Dr. Bono misunderstood the green open access slightly. Green open access is where an author publishes their papers on the Internet after posting to a journal with peer review, such as *Nature* and *Science*. That is why they rely on a journal with peer review for evaluation. Unless the conventional evaluation system remains unchanged, I feel that green open access should be recognized more as one of the open access methods.

●Hayashi Japanese research communities have not fully noticed this issue yet even though it is extremely important to talk about how the learned society sees the quality of preprints.

Dr. Laurence, do you listen to the opinions of people who are members of learned societies or research communities? What type of discussions do you engage in? I believe the members of learned societies assume a role of quality control in respective area including preprints.

●Lawrence We talk a lot to not only funders, but also talk to researchers, and particularly ECRs are desperate for a change in the system. They see what is before them and do not like it. There are a lot of initiatives like Future of Research which started in Boston and has now spread more broadly. It is a growing group of post-docs who want to change the system and want to try and encourage that. We obviously need to change the whole system. We need a lot of education from not just ECRs but also to more senior researchers. The challenge is that those doing the assessment are the senior researchers who have got to their present situation through the current system. Therefore, they may have less incentive to change it. You increasingly see funders and others who are recognizing preprints. Wellcome and NIH and many others are clearly encouraging listing of preprints. But without any peer review on those outputs, you can see what you have done but not necessarily the quality of what is it that you have done. However, it is shifting, and the upcoming generation is keen to change this.

●Hayashi In a sense of changing how to communicate with learned societies or with researchers in the future, do panelists have any comment on this?

●Takeda What caught my attention about the activities of F1000 is whether they do not need learned societies as communities and journals as communities. I am not really familiar with life science, but in other areas, I think a learned society or a journal published by a learned society forms a community and gathering together in the community leads to a type of a guarantee for quality. Does that mean you think it is okay to be more flat? Or, will F1000 create more communities within the platform?

●Lawrence The platforms span different domains, but I am not suggesting we get rid of communities. Communities are important as are academic associations and societies. All I am suggesting is slightly different roles for those groups, and the key thing is that sharing of findings is separated from the rest. Those academic communities, reviewers, and societies can still do the curation and evaluation for impact, quality, importance, and novelty. We just do not need journals to help enable the publication. arXiv is a fantastic base, and if we could expand arXiv to all different areas and layer on top peer review and then separately curation on top of that, that would be perfect and have different communities to look at different areas within it.

●Hayashi What about the learned societies?

●Bono As with Mr. Hayashi's investigation, researchers of the field of life science ignores journals published by learned societies. Because of this, I do not think learned societies on life science create their own community. It seems to me that members only participate in alumni associations rather than otherwise.

●Takeda Does that mean learned societies and journals are originally separated from each other?

●Bono I think it depends on the area. I mean I only belong to learned societies on life science with an enormous number of members.

●Takeda I see. Do you mean that the role of a learned society on life science is just that, and the learned society does not publish its own journal?

●Bono Larger learned societies do have their own journals. While some researchers publish their papers in these journals, some belong to various learned societies and publish in smaller journals in their respective specialized fields. In that case, there is also a trend of publishing in *Nature* or *Science* first, which I mentioned two years ago.

●Takeda That should vary greatly between disciplines. In the area where journals published by learned societies are valued, as Dr. Laurence has stated just now, learned societies might take the initiative to change such things. Even with arXiv, people from the ACM (Association for Computing Machinery) took part in the Member Advisory Board (MAB) this time. Although arXiv used to bring universities on board basically, it has decided to include learned societies as members from this year. While it was more like arXiv was maintained by universities in the past, it will also be joined by learned societies as maintenance members, which is a major change.

In fact, researchers conduct experiments with ACM and learned societies work together. Especially, learned societies were saying that they wanted to work with arXiv, unlike a commercial publisher, and they thought it was a good thing for learned societies. I certainly agree with Dr. Laurence that learned societies can do so if they want to.

●Hayashi That is interesting. The manners of researchers change based on preprints, and if their manners change, the manners of researchers' communities change at the same time. I now have a clear understanding that it varies between domains and is not the same across the board. Now it seems like we are at the phase where we need to think how discussions will evolve once it is about your own domain.

Now, since we are talking about learned societies, let us move on to the next topic about libraries. We also received a question that when the time comes for preprint servers to be used, how should libraries and librarians handle that. This question is about the point of an argument that is also deeply related to the roots that SPARC Japan was created. I would like Prof. Kuriyama to comment on a topic useful for librarians' future activities. After that, Dr. Laurence can follow up on that.

●Kuriyama In regard to a library to maintain a preprint server, I think it is an extremely rare case that arXiv was maintained and managed by the Cornell University Library. Since regular libraries were institutional repositories, they worked as a receptacle for green open access. As long as today's publishing trends remain the same, although green open access has not been doing so well, I believe it has the significance of existence at least.

Another thing is that libraries themselves embark on the publishing business. There is a university's bulletin, which I heard it is unique to Japan. A university's bulletin is a journal where a university pays money to publish their faculty's research results. In other words, it is a type of gold open access. If proper papers were published, I think it can make sense as one library operation. In reality, I sometimes saw papers of ridiculous quality on a university's bulletin, so it is necessary

to be responsible in that sense.

I have experienced a case where some paper was suspected of violating a copyright, and when I sought legal advice from a professor, we said it would be a problem. For this reason, while an editorial board for a university's bulletin is directly responsible, when a library acts as a repository to publish such a paper, it will obviously be held responsible. I think libraries should prepare themselves for the worst and act like publishers.

●Takeda It is not for all university libraries, but for example, some Japanese universities operate a system equivalent to a subject repository. I think it is another option that anyone who is in the country can place their articles as long as the area matches, such as mathematics, even though they are not university faculty. What do you think of it?

●Kuriyama Yes, of course, that is also possible. In that case, universities need to work closely with learned societies in the corresponding fields.

●Lawrence On a different point, another role that libraries could take is around supporting researchers, particularly with data sharing, where researchers do not know a huge amount about the different options, about what you can and cannot do. There is an important role there that librarians are in a perfect position to help with. Also more broadly, with open science education around all those different issues - how to do it, what are the pitfalls, how to avoid them, and general education - there are some crucial roles that the librarians could take.

●Hayashi Thank you. Does anyone on the floor have any comments?

●Floor 1 I am a librarian from the Chiba University Central Library. What shocked me the most today was that libraries were not mentioned at all in any presentations. I was shocked that even Mr. Hayashi's summary slide did not mention libraries. When a platform introduced by Dr. Laurence appears, I was wondering about what roles libraries should play in the future. I was very pleased to see you point out what libraries could do to help data sharing and for education.

Another thing is that I would like to hear from Dr. Laurence if it is possible to archive articles with guaranteed quality to some extent, for instance, in institutional repositories.

Also, I would like to ask Dr. Bono about the difference between preprints and institutional repositories. I would like to know if institutional repositories can work as a preprint server, not just to place papers for green open access. Do you want to publish in institutional repositories? Why do you want to publish in bioRxiv?

●Lawrence The way I imagine is that publishing an article and data that goes with it is a set of different tasks and they do not all need to be combined. There is a set of different services that needs to be done. One of them is around quality control checks, which could easily be done at the library. They could be done even before an article is submitted to a preprint or elsewhere. You could imagine all sorts of different groups doing that to which a library would be a perfect group.

Equally, the data sharing could be done as a separate set of checks. Presently, our editorial

team at F1000 spend most of their time doing checks and supporting researchers particularly around data. It is a huge amount of work that needs doing there. All of these are tasks that all sorts of groups could do it. It does not have to be a publisher providing the services. It could be a group of post-docs. The better the service they provide, the more likely people are going to use those groups.

●Bono If you are willing to do it, it is a very good attempt, but in reality, I think it is difficult. I said a paper was published in bioRxiv in just two days. The question is if it is possible to operate at such a speed. Also, bioRxiv does not simply archive articles. It also posts tweets on Twitter for each area. Things like that are necessary. The thing is if it is possible to earn that many followers from around the world. bioRxiv has an enormous number of followers. That is something difficult.

Its searchability is also crucial that users can always find what they are looking for. There have been many people who avoid using bioRxiv because articles on bioRxiv cannot be searched in PubMed. People should be able find papers using search engine that you use. Since the data is big, it is ideal to have a repository on your own in that sense, but I personally think it requires a lot of work.

●Hayashi Does anyone want to comment on libraries?

●Takeda As shown on my slides, there are various ways to publish papers today. Of course, it varies between disciplines, but there is a common way that works for any discipline. Faculty and students will be pleased if a library educates them on what methods are available when they want to publish papers. As Dr. Laurence said, young researchers especially do not know that other channels are available. It would be great if they knew through a library. In fact, not just young researchers but regular researchers do not know these things. Although I know all the journals in the field that I specialize in, but when it comes to different fields, there are many things that I do not know. I think researchers would find it extremely useful to get that kind of information.

●Hayashi I think so. Even when an issue of article processing charge (APC) was mentioned at the previous discussion of the SPARC Japan seminar, we talked about what a library could do to support selecting open access journals to publish carefully. These days, there is a heated conversation that we should educate researchers not to publish their papers in predatory journals. Through today's discussion, I think that if librarians are asked about the choices of preprint servers, librarians can play a role in educating researchers in a way similar to a reference service.

Today we have a guest here who has been involved in institutional repositories for many years while conducting research in mathematics at Hokkaido University. Can we get your comments on the discussion up to this point from such a position? As a researcher, he was deeply committed in the initial activities for institutional repository.

●Floor 2 I am a researcher from the math lab of the Hokkaido University. As Prof. Takeda mentioned in his lecture, the culture of quoting arXiv preprints in a paper has been normal practice in mathematics and theoretical physics. Under such

a circumstance, I would like to know if it is possible with current technologies to later add DOIs for journal articles to arXiv preprints cited in references after publishing.

To comment on the culture of preprints, for example, before arXiv, researchers used to print out several hundreds of copies of a preprint in paper media and send them to research institutions around the world to secure priority. In this way, a certain research project develops around the researcher who created a preprint to establish a very closed community. We can also say that digital technologies are well utilized to resolve these situations. I hope to share the debates on preprint servers once everyone here understands such a background first. Is that okay?

●Takeda With arXiv, we can update a paper by adding a version number, like a version one and version two. For example, once a paper is updated to be a version two, it is possible to manually add a DOI to accepted papers.

However, there is no way to add DOIs systematically. With bioRxiv, once a paper is posted to bioRxiv, a link is generated automatically and systematically when it is posted to a journal. It is a win-win for both journals and bioRxiv, and links are properly secured.

On the other hand, arXiv does not have such an automatic mechanism. It does not even have a mechanism for metadata. Due to this, arXiv is faced with the criticism that it is not within the ecosystem of open access. Although it should be simply fixed, it seems like it is a difficult thing to do, since arXiv has the tradition of 30 years. Today, the system displaying pages has been replaced, but the appearance has not changed at all. It is perhaps because they value the tradition to use a design that does not change much. I felt that the history of arXiv is hampering its development to some extent. That is how I felt since I am new to arXiv.

●Hayashi If so, will preprint servers be subdivided in an evolutionary sense? I think there is still room to create a new design without feeling restrained by the tradition or to consider UI/UX for new archives in bio-related studies. Is there such a sign? You haven't compared bioRxiv with arXiv, have you, Dr. Bono? Are there any concerns?

●Bono For life science, researchers have options to publish not only in bioRxiv but also PeerJ and BioRN. I think the reason why they post to bioRxiv is that they think many people see articles on bioRxiv because bioRxiv inherits the tradition from arXiv. It gets a bit confusing if various other archives come up again. That is why I do not want to consider others. Because it is a tradition to publish articles in journals from bioRxiv, they just follow the path. In my opinion, there will not be many gateways to preprints.

●Hayashi Well, what this argument evokes in reverse is that in order to realize the de facto standard for preprint servers, who will make such an appeal? The person who will start becomes very important? Even if a well-known person in the community were not necessarily an impactful researcher in the respective domain, it seems to me that people would gather naturally when it is started by a person with a certain centripetal power.

●Bono I see. I use a computer to conduct exper-

iments in a very minor area of life science. Researchers of life science conduct experiments to present their data in a journal. It is the field that journals do not easily accept papers without such data. I would like to use bioRxiv in the same way as posting to arXiv for computer science.

As for the matters we talked about today, I personally do not think about publishing to a journal with peer review and that is it. I will turn a situation into making new discoveries by using the data obtained from there. I am using bioRxiv for a new purpose instead of the original purpose. bioRxiv is an important place where we can try such new usage.

●Lawrence In terms of encouraging more people to use preprints, I am sure funders will help to some extent, but if other types of outputs beyond a standard narrative research article published in a journal are recognized when it comes to what have you done with their money, and applying for your next funding or your next career move, then that will make the largest difference. We saw this to some extent when the NIH and others started to say they recognize preprints. Suddenly there was a much faster growth than when it was a small group of well-known researchers preprinting. Ultimately, it comes down to if you are rewarded and incentivized by your institution and your funder to share not just preprints, but data, methods and materials, and also a much broader range of activities, such as peer review activity, teaching, all sorts of things that are very important to the scholarly ecosystem.

●Hayashi Thank you. This time, we planned on focusing on preprints for research papers to clarify the discussion. Quite appropriately, Dr. Laurence's last comment pointed us in the direction that thinking about preprints will widen the scope of research results besides papers to further develop science, and in other words, it means to move ahead with open science. I arbitrarily interpret her words as great concluding remarks. Under the original plan, I would ask each panelist to share some final words, but we have run out of time. We will close the panel discussion now. Let us give applause to all the panelists again. Thank you very much.