Thank you to the organizers for inviting me to be part of this conference. It is an honor to be here and a pleasure to spend time with all of you, and in particular my colleagues from SPARC Japan. I will cover three things using the movement towards open access, the progress, and the challenges to adoption of open access as one way of looking at the challenges that face open science as a whole.

We consider open access to articles to be a primary pillar of open science. It is just one part of the whole open science ecosystem, but it is a very important piece of the puzzle. I will focus on open access primarily because SPARC is a global coalition of academic and research libraries whose mission is to make open the default in research and education. To ‘make open the default’ means, that you will share the results of research and the materials you use to teach unless you have a compelling reason not to. That is essentially the opposite of what we have right now.

At SPARC in North America, our primary focus is on supporting an open research ecosystem by opening up access to the materials which scholars and scientists most desperately need in order to do...
their work. This includes articles, research data, and educational materials like textbooks, courseware syllabi, and learning objects.

The Potential of the World Wide Web

We would think that in 2018, with the beautiful platform of the world wide web and the internet at our fingertips, we would have unfettered access to everything that we need to be able to do effective science, research, and scholarship automatically.

With the advent of network digital technology, we have access to much more information and opportunities to do much more with that information. It should be easy for us, that is only in theory. I think we all understand that, despite the promise of the internet, the materials that we most desperately need the freedom to be able to work with to do good research and good science are overlaid with restrictive policies, access barriers, pricing barriers, and barriers to using and reusing information in the digital environment.

Financial Barriers to Access

Since SPARC is a membership organization of libraries, a major way that we begin to think about wanting and needing change the current status quo came form of dealing with financial barriers, the cost of getting access to information to our colleagues on campuses. These financial barriers are primarily for journals. This table shows the average journal prices in US dollars from 2016 (Figure 1). You can see how expensive it is to essentially rent access to a year’s worth of articles in a journal in a particular discipline. The average is $1,788 US dollars, which is an enormous amount of money to pay for one year’s worth of access to a set of articles.

The costs are not insignificant on an individual title level, and cumulatively this is an enormous business. Profits made from these journals are roughly a $10 billion a year. It is a revenue-producing industry. Therefore, there are enormous barriers and financial burdens that face the folks who want to get access to this material, as well as for librarians who are tasked with getting access to these materials.

The cost barriers are not just theoretical barriers. Many researchers think, “It is expensive, but our libraries do get us access to everything that we need. We have the ability to pay and get the access that we need,” and it is true that we do not see faculty members protesting in front of university buildings because they are feeling there is a lack of access to primary articles. However, they actually do feel this lack of access, and we see this every day.

I believe that all people in this audience have encountered this scenario. You are doing research on a topic that is of interest to you. You do a Google search to look for articles that you may want to use to inform the research that you are doing (Figure 2). My son happens to have type one diabetes, so I am constantly looking for the most up-to-date information about how to keep his blood sugar from going low overnight. You find an arti-

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Average Cost Per Title</th>
<th>Average Cost Per Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td>$5,105</td>
<td>Psychology</td>
</tr>
<tr>
<td>Physics</td>
<td>$4,508</td>
<td>Sociology</td>
</tr>
<tr>
<td>Engineering</td>
<td>$3,244</td>
<td>Social Sciences</td>
</tr>
<tr>
<td>Biology</td>
<td>$2,104</td>
<td>Political Science</td>
</tr>
<tr>
<td>Food Science</td>
<td>$2,779</td>
<td>Biology</td>
</tr>
<tr>
<td>Environment</td>
<td>$2,718</td>
<td>Botany</td>
</tr>
<tr>
<td>History</td>
<td>$4,416</td>
<td>Veterinary Science</td>
</tr>
<tr>
<td>Sociology</td>
<td>$2,400</td>
<td>Veterinary Medicine</td>
</tr>
<tr>
<td>Mathematics</td>
<td>$2,319</td>
<td>Neuroscience</td>
</tr>
<tr>
<td>Agriculture</td>
<td>$2,221</td>
<td>Medical Sciences</td>
</tr>
<tr>
<td>Arts</td>
<td>$1,856</td>
<td>Pharmaceutical Sciences</td>
</tr>
<tr>
<td>Business &amp; Economics</td>
<td>$1,801</td>
<td>Pharmacy</td>
</tr>
<tr>
<td>Health Sciences</td>
<td>$1,772</td>
<td>Library Science</td>
</tr>
<tr>
<td>Computer Science</td>
<td>$1,713</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Language &amp; Literature</td>
<td>$1,474</td>
<td>Law</td>
</tr>
<tr>
<td>Business &amp; Economics</td>
<td>$1,474</td>
<td>Philosophy &amp; Religion</td>
</tr>
</tbody>
</table>

(Figure 1)
Enabling Open Research

National Institute of Informatics

The 3rd SPARC Japan Seminar 2017

Feb. 21, 2018

...ple that you think that might be of interest to you (Figure 3). You read the abstract and think that it sounds good. You want to take a look at it, but when you try to access the full text of the article, you run into a paywall (Figure 4). It happens to all of us.

Overcoming Financial Barriers to Access

Despite this fact, we are not protesting. What we are doing is very important for us to understand changing behaviors that are moving towards open. As librarians, when a researcher is unable to access the content that they need, we would love it if they would call us and ask to get a copy of the article through an inter-library loan. Sometimes that happens, but more frequently, people will go directly to the author, or they will talk to a colleague at another institution that has a subscription and get a copy. A lot of our younger students and early career folks are using the hashtag #ICanHasPDF and asking for copies on Twitter, which is very creative (Figure 5).

We are also starting to see people create small apps that search repositories and the web for free copies of articles that are behind paywalls (Figure 6). We are really seeing is an enormous black market of pirated articles springing up. Sci-Hub is a completely illegal database of 58 million pirated scholarly journal articles. Science Magazine did a study to see who is accessing these illegal articles (Figure 7). As you can see from the illuminated pieces of red, they were coming from all...
over the world. This is a problem that is under the surface. Even though we may not be protesting on our campuses, we are certainly seeing the net effects in our daily lives.

What is worse, we hear over and over from scientists, scholars, and researchers that, when they hit a paywall, they automatically calculate in their head, and make the decision to skip that article and move on to something that they have access to rather than making an effort to find a free copy of the article. That means that we are doing research based on what they we access to, rather than on what we actually need to know. What is worse, we teach our students and provide information to other people based on what we have access to rather than what we might really need to know.

**Cost Compounded by Limited Utility**

Essentially, we have been operating in this beautiful internet-enabled world in a system that forces us into using workarounds to get the basic information that we need to do good science and good research. As we think about getting into open as a potential solution, we should recognize that it is not just getting access to these digital materials that is important. We want to be able to access these digital articles and their underlying data and be able to use them fully and freely in the digital environment.

As an example, when a discipline moves into conducting research digitally (such as what happened in genetics and genomics with the digitization of the human genome), you end up with a corresponding exponential increase in the amount of attendant data and information (Figure 8). When the genome was digitized in the late 1990’s, the amount of data deposited into the Genbank genome database went up on an exponential curve. This is not special to genomics. Every discipline that moves into a digital operations environment has seen (and is seeing) this kind of increase in information.
It is critical to remember that our ability as individual humans and scientists to deal with this information has stayed the same (Figure 9). My friend Cameron Neylon said that he is being asked to deal with so much more information, but that he only has one brain.

What this means is that we really need to enable computers as a category of readers of scientific and scholarly articles, as well as the attendant data that forms the basis of those articles. We need to do so in order to make sense of the increasing amount of information we have in front of us. We want to be able to text and data mine and do computational analysis not just on a subset of articles or the ones that we can get access to, but all articles and everything that we might need to inform our work in order to truly unlock the power of the information contained in digital material.

We also are operating increasingly political environments where the validity and value of science is in question. We want to be able to demonstrate the validity of conclusions and articles. We need access to the articles as data as well as the underlying data to verify, reproduce, to validate the efficacy, and the truthfulness of science on a regular basis. This is really important.

What we talk about at SPARC all the time as a motivation for moving our community into an open access and open research environment is about optimizing the system sharing the results of research and attendant educational materials to better suit the needs of the folks who are doing the science and who are using the science. I should say ‘science and scholarship’ because we are talking about open science in this symposium. At SPARC, we are talking about open scholarship. In digital humanities and social sciences, we believe the work that is done on our campus in education benefits from utilizing the strategies of full accessibility and fully reusability.

Open Access

Talking about the background of open access, 15 years ago a group was convened by the Soros Foundation in the Budapest Open Access Initiative, and tasked with answering the question: If we could build and conceive of a system for researchers to share the outputs of the research fully and freely and allow people to use them to its fullest, what would that system look like? What would we need to do to make that happen? The idea of open access was coined during the Budapest Open Access Initiative. It talked about bringing the new technology of the internet together with the tradition of scientists and scholars freely sharing their ideas with each other as well as the results of their research, and bring these two things together for the first time to form a powerful and unprecedented public good, something that would have a net positive effect on not only science and individual scientists, but societies as a whole.

To quote directly from the Budapest Open Access Initiative, “The public good they make possible is the world-wide electronic distribution of the peer-reviewed journal literature and completely free and
unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds.” SPARC was privileged to be at the table and helped to draft this statement. The Budapest Open Access Initiative did not only theoretically talk about enabling an open research environment, but it outlined the definition of the framework of how to get there.

When we are talking about opening access and making things free, open means different things to different people. For open access to the materials, we need to do science and research. The definition in the Budapest Open Access Initiative is very specific: “By “open access” to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers.” We want to not just be able to get access to an article and read it in PDF form on the web, but we want to be able to bulk download the digital data on the open web and text and data mine that data, distribute it, crawl it for indexing, and use it for any legal purpose under the sun.

It is important to remember that is that open access is a two-part definition: Free online accessibility coupled with the rights to use that material fully in the digital environment. We cannot have true openness in open science or open research without having this definition fully realized. What has been truly important as we at SPARC have continued to tried to move the needle in this direction is how we talk about open matters. If we simply say to researchers, scholars, funders, publishers, and all the important players in the science and research ecosystem that they should push for open access because it is good, or that open is better than closed, or we are open access advocates, it makes us sound like all we really care about is open. However, that is not really the case. We want people to really understand that we are not asking for an open science system and open access system because open is better than closed. We want to use this openness, free accessibility, and the full utility as an enabling strategy to enable everyone to make important progress in our daily work. Whether we are scientists, students, or funders, we need to think about what open can do to help us achieve our end goals more efficiently.

We knew this when the Budapest Open Access Initiative was drafted. We had the aspirational goal of the system set out. We had the framework of how to get there, the definition, and we talked about open in order to accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as possible, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge. It is about open in order to do all of these things, not open because open is better than closed.

Now we are starting to talk more with researchers on campuses about why they should participate in making materials open in the open research environment, such as sharing data more quickly in order to prevent a potential mosquito-borne pandemic like the Zika virus from spreading. Opening up access to your research data and articles will also improve transparency and reproducibility. By putting yourself in the shoes of the person who you are talking to, you can help them understand that open can be an enabling strategy to
achieve whatever is important to them.

At SPARC, we work on all aspects of opening up access to materials in this ecosystem, and that includes textbooks. This means opening up access to textbooks in order to make higher education more affordable for all students.

‘Open in order to...’

The idea of filling in the blank in ‘Open in order to...’ will be familiar to those of you who participated in Global Open Access Week. This was the theme of 2017’s Open Access Week. We asked the community to think about talking to folks, and encouraging people to ask others to fill in the blanks. We wanted to draw out of stakeholders what is important for the ultimate success of an open science and open research enterprise. We did so to get them invested to recognize why open could be an important strategy for them.

We had some wonderful results. This is a picture of a chalkboard in the lobby of the Gates Foundation, an enormous private foundation providing funding for research based in the United States (Figure 10). During Open Access Week, in their lobby they drew this on their chalkboard and asked their grant makers to answer the question, “What can open access do for you?” The Gates Foundation folks actually filled in the blanks. They have enormous investments in global health initiative and public health. Their ‘Open in order to...’ sentence said: “The Gates Foundation supports open access in order to save lives through our funding of global health initiatives.” Therefore, they really saw themselves in the end game. Funders embracing open and understanding the power of open is an enormously important piece of effectively enabling an open research enterprise.

The World Bank also participated in Global Open Access Week, and they filled in the ‘open in order to...’ in this way (Figure 11). Part of their core mission is to end extreme poverty. They are opening access to all of their publications and data over the last five years. They have adopted an open access policy in order to help them achieve this portion of their core mission. Bringing in the source of research funding under the open tent, or getting people to understand the importance of openness is something one which we have been working very hard.

Joe Biden, our former vice president in the United States, he is a great example of somebody who came to understand the power of open and open science intuitively (Figure 12). When his son was diagnosed with a brain tumor, Vice President
Biden came realize the importance of open when he wanted to move his son from one treatment facility in the United States to another in order to get an experimental treatment. The first research facility had to transfer the electromagnetic brain scans to the other research facility. The vice president was told it would take 24 to 48 hours for the scans to be read because the data was in a format that the other research institute did not have. They had to convert the data in order to use it. Vice President Biden gave a speech to the American Association for Cancer Research where he talked about the unconscionable delay of 48 hours that was needed to translate data from one format to another. He did not understand that what he was talking about was interoperability, of using standards to make sure that scientific data is in a format that can be openly, easily, and immediately read and fully reused, but that is what he was getting to intuitively. Later in his speech to cancer researchers, he said that it was not just the data be in a form that is available on day one that is important, it is the articles that provide insights into the data that we need access to immediately. He just came to that conclusion organically that we need open in order to speed up progress towards treatments and potentially to find a cure for a specific condition or disease.

If you are interested in looking at other people’s examples and what they have said, the website is called openinorder.to. There is a collection of stories from Open Access Week to help folks in the community talking about open science and open research have more material to form a base for conversations.

**Barriers to Open Access**

We are making progress. Open is within our grasp, but I think it is important for us to look at why we have been talking about this for 15 years, and yet we are not there yet. Some will blame commercial publishers. Although there is pushback from commercial publishers to a certain degree, it is not the main issue. The main issue is that we have individual scholars and researchers who are not convinced or do not understand that moving into an open science or open research environment is optimal. It is probably no surprise. Even if people are aware that there are options to share articles openly, they often do not know how they can tell whether a journal is really open access.

The Creative Commons licensing scheme is very confusing (Figure 13). We are asking them to learn another short-handed legal language in order
to know whether something is truly open or not.

There is a real influx of predatory publishers who seek to article processing charges where authors pay a fee to publish their articles. Researchers tell us that they get 5 to 10 emails a week asking them to publish articles in journals that they know are fake. This makes researchers think that open access is a fringe thing, that it is not legitimate, and that it is not okay to publish in open access journals. They are confused about what to do with their data. We are trying to provide mechanisms to help them understand how to make their data openly available or share their data by explaining how to do it, but they are still confused. There are so many details that they need to try to understand.

To finish up, I would like to focus three things we need to do in order to really make progress. The main thing that researchers tell us that they are worried about is that they are not rewarded for making research outputs openly available when it comes to tenure, promotion, funding, and evaluations in the same way as publishing in journals that have high impact factors. We have to grapple with this issue and deal with it head on. Therefore, we have institutional incentives that do not reward the creation and use of open materials and practices. These institutional incentives are found on our college and university campuses, as well as in our public and private funding bodies.

**Addressing the Barriers to Open Access**

I want to close by giving three suggestions for how we might address the problems that our researchers face, and the problems that we have in trying to move into an open environment. First, we need to realign the incentives that underpin our scientific and scholarly communication practices to contribute those that are contributing to the core mission of the university. If we look at university or funding agency mission statements, we often see that they provide direction for the diffusion of knowledge, such as this example from Chiba University (Figure 14). Despite this fact, our incentives are rewarding publication outlets that are not reaching the widest possible audience. Therefore, we need to realign incentive and reward behavior to move the university or funder in the direction of their core guidelines.

There are too many tenure and promotion guidelines in the US and around that world look like this (Figure 15). In the first column, you can see that professors are rewarded for publication in 10 to 15 journals with an average impact factor of X.
or above. There is no mention of the breadth of audience that it reaches or of the utility of the work in the digital environment. That needs to change. We need to look at recreating and realigning incentives to better support open.

Incentives are different in different disciplines. How we get to open will differ between genetics and genomics and other disciplines. Think about what those kinds of rewards and behaviors could look like in your disciplines.

We can consider employing article-level metrics that trace all kinds of downstream uses and potential impacts, and incorporate those in the evaluation process. I am not suggesting replacing impact factors with alternative metrics, but to supplement them. We should have additional measures that reward the behavior of open sharing of materials.

Research funders such as private foundations and private funders play a very important role in the United States, and they are playing an increasingly important role as our federal government wrestles with the importance of science and the continued funding of science. Private foundations in the US provide a lot of money to our campuses to continue research, and they are beginning to adopt the idea of open as a core enabling strategy in helping them achieve their primary missions, such as the case with the Gates Foundation.

The Open Research Funders Group is a group of funders that SPARC has been helping to convene for the last two years that is comprised of some of the largest private funders of research in the United States (Figure 16). They are looking at ways to say that they are supporting researchers. They want to leverage their return on investment. They believe that opening up access to results, data, and articles related to that research will help them to achieve their missions. They want to work with people in higher education to talk about realigning incentives to make this more of a possibility. This is a very promising strategy to begin to move the needle towards open.

Second, we really need to rethink the business models that support open access publications and also open access outputs in general. It is really important that we recognize that scholarship and science is a global enterprise. Our first two speakers really illustrated the global nature of research and the need for building financial models that support open sharing, particularly of journals. They also illustrated that we should make those models support local knowledge creation, encourage contributions in developing countries, from smaller institutions, and from all different types of contributors, as well as reward global dissemination.

There has been a push lately for the open access community to think about making a large-scale transition from supporting subscription-based journals to article processing fee-based journals. That is an interesting model that will get us to having more open access articles, but we have to think about the real effect that doing so will have on countries that currently cannot afford to pay subscriptions and are being shut out from accessing...
science. They are not going to have the money to be able to publish their articles in open access journals. We therefore move on to a problem of participation in science. We need to think about equity and fairness in the business models that underpin open research communications channels.

The third and final topic, in the same spirit of equity and openness in the channels that support journal publications, the infrastructure that we use in scholarly research environment needs to be open to a great degree and to be what we call ‘community-owned’ or ‘community-controlled infrastructure’. I am thinking specifically about institutional repositories in this regard. Elsevier recently acquired the institutional repository platform bepress. In and of itself it is fine since bepress was a commercial company that was acquired by another commercial company. We should have not been surprised in the library and research communities when this acquisition took place, yet the alarm bell that it rang in many librarians’ heads was that a platform that libraries rely on to freely and openly share articles on campuses is no longer ours. It is owned by a commercial entity. That may or may not be problematic, but our ownership of that channel was something that we took for granted in the library community. We thought that it would always be part of the open landscape. We need to be sure that we are being smart about the terms and conditions when we are negotiating with commercial vendors of infrastructure to make sure that the ownership and openness of that infrastructure remains intact, even if the ownership transfers to commercial players. We are not going anywhere. We are going to be part of the landscape, and we need to be smarter about how we are working with commercial entities.

Finally, and most importantly, we really need to continue to build collaborative strategies that reflect the global nature of the research enterprise. If we could reimagine the system from the ground up to optimize the system of sharing research for all folks involved, what would it look like? We need to keep asking ourselves that question and keep answering it together. We then need to work towards building a system that uses open as an effective enabling strategy to help us achieve the goals that we share.

● Takeda Although preprint servers have a long history, 10 years ago it was a very limited domain. Now preprints are growing at an incredible pace. Since you mentioned that we have to maintain control of the access environment, and considering that some preprint servers are managed by communities and others not, how to place preprint servers in the open access landscape?

● Joseph Yes, it is a mixed situation in terms of who is controlling or owning preprint servers. I think they need to follow your advice that the ID is the infrastructure. We need to have DOIs for preprints. We need to have ORCID identifiers on the authors so that that material can be located on-par with articles that are published in journals or articles that are in institutional repositories that are carrying DOI’s. The idea of identifiers as a critical piece of infrastructure rings right to the heart of what will make preprint servers successful.

● Takeda Therefore, a preprint server is one of the key components in creating an open access environment.
Joseph Absolutely, and it is very encouraging that the use of preprint servers is increasing. Although in physics they have always used preprint servers, over the last two years, preprints have finally emerged as something that people think is possible. There is behavior change that is happening where researchers are beginning to think that preprints are beneficial to them to get their work out faster. They are realizing that, although they may not be immediately rewarded for putting something in a preprint server, they are not going to hurt by doing so. That is a really important psychological step for behavioral change in the community.

David Going back to the idea of ‘Open in order to…’ virtually all of the foci of the activity can be reduced to creating an externality. An externality can be bad like burning fossil fuels, or it could be good such as making public something that can be used by others or making clean water available. We depend on a lot of externalities, and they cannot be brought within the calculus of the existing way of allocating resources. Therefore, it is important to know that there is a very large literature in economics that deals with trying to provide externalities that are good, and how to stop the generations of externalities that are bad. This is not an easy terrain to work in, but I think you might reach out and find some of the leading or younger people who have learned in economics about externalities and find out about that mechanisms that have been used to elicit behavior that is beneficial to a large number of people and involves some way of getting something back. This will induce people to contribute to it something that is an externality of interest only to a small group. The question is whether you can harness the activities of a small group by getting them to cooperate in a way that creates an externality for a very large group. That is an interesting approach to take, and it has been taken. For example, people are working on fixing the climate problem by thinking about the mechanisms we use to fix the ozone hole.

Joseph That is a great point. There are two things I want to pull out of that. One is that we are thinking about ways to demonstrate some of these things, and the idea of working with younger early-career researchers is something that SPARC has a huge focus on. We convene a yearly meeting of early career researchers from around the globe called OpenCon. We have about 200 participants each year. Last year, we had 13,000 young researchers from around the world apply to be a part of the meeting. We can only take 200, but we took 200 from 62 different countries, and they all came to the meeting. Part of the activities of the meeting was to identify something that they wanted to work on individually or in groups, commonalities that they wanted to look at, utilize, and open as an enabling strategy to move forward. Out of that meeting, we now have these growing groups of early-career folks who are forming communities around using open as a way to advance progress in their specific discipline or interest area. Therefore, I think it is a beginning. We did not think about it as using the roots of economic theory. We looked at it as a community of practice and trying to build community activities in that way. It is a tremendously important and fruitful way to think about this.

Murayama My question may not be on the central focus of your talk, but I am interested if you
have any finding of the open access practices in various countries in your experience, that publication industries should have different culture, different industrial behaviors in the different countries maybe based on the different legal systems. My focus is just in Asia, but if you have any example of that kind of experience.

Joseph I would not presume to tell the commercial publishers how to operate. I think they are very smart, and I am sure they are looking at different operational modes in different countries. From our perspective, we are looking at different models that are working in different regions. For example, when we are looking to find an alternative to article processing charges for, we are looking at the culture in Latin America of institutions supporting local content through institutional funding. There is not a culture of publication in large commercial journals or a high emphasis on impact factor in Latin America because there is a culture in many of the countries of valuing locally-produced content. The university's imprimatur is what is important, not the journal name, so we are looking at those types of cultural differences.

We are also looking really carefully at places where the production of local knowledge is not valued in the same way because people are asked to publish in high-impact journals. Along with colleagues, we recently founded a presence for SPARC in Africa. In many of the countries, researchers are asked to publish in commercial journals or high-impact journals that use English as the primary language, which disadvantages researchers at the get-go. Second, the subject matter that is of interest of those journals is very different from the subject matter that is of import to many of the local researchers. There is not an interest in indigenous or local research. Our African colleagues call for the decolonization of scholarly communications and publishing.