4-11-2017

What's Up with Open Research Data

Heather Joseph
SPARC, heather@sparcopen.org

What's Up with Open Research Data

Heather Joseph
Executive Director, SPARC
University of Denver

April 11, 2017

Michael Levine-Clark: Good morning, everybody. And good morning to those of you who are watching this via Zoom. I'm Michael Levine-Clark. I'm the Dean of Libraries here at the University of Denver. I'm delighted to welcome you to the second day of the Colorado Open Scholarship Series, which is a series of presentations and meetings with campus leadership on all things open: open access, open educational resources, and open data, which is today's topic. These events are jointly sponsored by the libraries at DU, at the Colorado School of Mines, and across the University of Colorado System. I know that many of you are watching and listening today from those locations. There's information about – There's a poster outside the room that's got information about additional events throughout the week and you can ask Jack or Jenelys -- raise your hands -- if you have questions about these and you can watch tomorrow and Thursday's events via Zoom, I think. Again, ask Jack or Jenelys for information about that.

I'm thrilled to introduce Heather Joseph, who's the Executive Director of SPARC. SPARC is the Scholarly Publishing and Academic Resources Coalition. She's here today to talk about the importance of open research data. It's vital for libraries and the researchers that we support to preserve and make accessible the data that underlie that research. They can be used to validate and build upon the research and the scholarship that is reported in the journals and the books that we collect. Open data -- data sets that are accessible and available to all -- allow other scholars to test and replicate published research, to challenge findings, and to undertake new research that expands the scope of knowledge. Having that data available and open to everybody is fundamental to what we, as universities, do.

Heather Joseph is here today to talk about that topic. Nicole Allen, her colleague, will be talking about other topics tomorrow and throughout the week. Heather is the Executive Director of SPARC, where she leads strategic and operational activities of the organization. She's focused SPARC's efforts on

supporting new models for the open sharing of digital articles, data, and educational resources. Under her stewardship, SPARC has become widely recognized as the leading international force for effective open access policies and practices. A firm believer in collective action, she has bolstered SPARC's mission through the development and leadership of effective coalitions. She convenes the Alliance for Taxpayer Access and the Open Access Working Group, broad coalitions of university, library, advocacy, and consumer groups that serve as leading voices on U.S. open access policies. She supported the creation and launch of SPARC's student Right to Research Coalition, an international advocacy group organization that now provides nearly seven million students worldwide with opportunities to actively engage in the open sharing of research outputs and educational materials, often in partnership with the library. Prior to joining SPARC, Heather spent 15 years as a publishing executive in both commercial and not-for-profit organizations. Now, I'm going to turn it over to Heather, who's going to talk about open research data. Thanks.

[Applause]

Heather Joseph: Every time I hear my bio, I start to feel older and older and older. If I did that for 15 years and I've been at SPARC for 12, what does that start to add up to? But let's not dwell on the negative here. Thank you so much for inviting me to come and join you today on your gorgeous campus. Thank you all in Zoom. I don't know if you can see us or how this works, but I'm waving at you from up here in this beautiful Academic Commons room.

As was just said, the topic is open research data, but if you all have questions, remotely or you guys here in the audience, about any other aspect of the open movement, writ large -- if you want to ask questions about open access journal articles, open educational resources -- I'm happy in the questions and answers to divert off into those topics, as well. Nicole is actually going to be talking explicitly about open education resources, open textbooks, and really digging into that incredibly vibrant and growing movement in her talks at Boulder, tomorrow. She's much more expert in that area than I am. So if you really have

burning questions, she's the person to talk to. Open access, I'm your gal. Hopefully, today, when we dig into open research data, what I'm going to try to do is start out with a little bit of a wide angle lens and talk about some of the factors that are driving this real global movement towards opening up access to data and research data, in specific. It's a front and center issue in the research community, in civic government worldwide. Then we'll finish up talking about how that is beginning to really impact our campuses, what opportunities we have in front of us, collectively, to engage productively to help drive the movement in a positive direction and also expand the role that the library is playing and can play with this incredibly important layer of information.

Ok, advance. My slides are not advancing. I'm looking for my little AV person. It advanced when we tested it, but it's not budging. Oh, there we go. What did you hit? Ok, perfect. Thank you.

Just two seconds of basics on SPARC. We are a library membership organization. We have members in the U.S., in Canada. We have sister organizations in Europe and Japan and we recently launched SPARC Africa, which we're really excited about. We want the membership of our organization to reflect the reality that research and education are global enterprises and that they are, in fact, collaborative global enterprises. So, we work very hard to make sure that the membership of SPARC represents the global nature of those initiatives.

Our primary mission is – we were founded to really support, in an action oriented way, a move to a more open and equitable system for scholars and scientists to communicate the results of the research that they're working on. Our end game is using open as an enabling strategy and really working to make open the default in research and education. Right now, especially for research outputs, our expectations are that we're going to hold onto our data, we're going to hold on or keep our articles in subscription access journals unless there's a compelling reason to make them open. What we'd like to see is that dynamic flipped and have scholars, educators, researchers really approach this

enterprise with the thought that science research advances when we share the results of what we found. So, the expectation is that things are open unless there is a compelling reason to make them closed. As we'll talk about in this data environment, there are lots of compelling reasons that you might not want to share every piece of data that we collect and we might not want to share it right away. But resetting that expectation is what we're working towards as our final goal at SPARC.

As we know -- no secret, right -- the farther along we get in the 21st Century, the more digital the conduct of research and scholarship becomes. There's almost no discipline that's untouched by either digital methodologies or isn't being propelled by, and helped and assisted by, digital techniques for doing research as well as communicating research. That's fantastic, right? That's what we wanted to see with the internet. We were talking yesterday at University of Colorado at Colorado Springs that, as was mentioned earlier, when I started my career as a journal publisher, it was in the late 1980s as the internet -- '89, '90 -- when the internet was just beginning to surface out of DARPA. I worked for the Astronomical Society. My boss was an astrophysicist. A lot of the community were already using ARPANET and the precursors to the internet to… Which we got a laugh out of yesterday, cause nobody – a lot of people were not using BITNET or UUNET, which my first email address was UUNET. But my boss saw at that moment that this was a network, this was an enabling tool created by scientists to communicate science. The irony, for us 30-some odd years later, is that the internet has been optimized for so many things. It's phenomenal for shoe shopping. I could rebook my cancelled flight from the floor of the airport on my laptop on a different airline. But it's still not optimized for sharing scientific research results. We'll talk a little bit about why – the cultural dynamics and, frankly, some of the commercial and financial dynamics that have stopped us from being able to do that.

Really, our mission is to use the internet for the purpose it was built for and really using that as an enabling tool to set the default to open in research and

scholarship. Our interest is shared across the board. What we're finding is that we've been talking about open access, open data, for quite some time and we use specific language around open access. We assume everybody knows what it is that we're talking about when we're talking about open. That's not necessarily the case. As we've done more and more work across the important constituencies in the research enterprise, one of the things that we found out is that funders have been thinking about this and talking about this for a long time. They haven't been using the same language that we've been using, but they certainly share the same end goals. What's been very interesting to us is to try to take a step back out of the library community and put ourselves in the shoes of the funders. We always said that one of the things that SPARC was designed to do was to look at creating change in the academic community by following the money stream and for a long time we had that money stream going back to library budgets where it was being used to purchase scholarly journals. We realized that in order to really be effective, we have to follow it back even farther – that what we're talking about is such a fundamental change in thinking about how research and scholarship are communicated that we needed to go back to the source of the funding and really think: How are the funders thinking about this? What are their expectations when they're funding research? In the U.S., funding research is still a pretty large scale enterprise. Over the last decade, historically, we've invested about $60 billion per year collectively in publicly funded research results. We wanted to think about why funders invested in research. They do it for the same reasons that we support the research enterprise on our campuses: to generate new ideas, to facilitate new discoveries, and to enhance our collective understanding of the world and our interactions with it. It's a public good for funders of public research and, in fact, the funders of private research in general.

They realize, as do we, that this can only happen -- this goal can only be fully met and optimized -- if you can access and fully use the outputs of this research. One of my favorite stories is Keith Yamamoto, who is the dean of the UCSF library -- library school, I'm with librarians -- medical school, standing up at a

meeting of the National Academies early on and talking about the need for optimizing the sharing of research outputs. He said, "If we fund you to do research and you can't or won't tell someone what you found out, what was the point of me funding research in the first place?" Communication of research results is part and parcel of doing that research. Science research scholarship only advances when you can see the results of the work of those who came before you, read it, use it, and build on it. That's a fundamental tenet that research funders have at their heart and it's one that we've had to work to understand and to understand how we can communicate that the work that we're trying to do is fully in support of optimizing their ability to achieve those goals. If it's true that science advances better -- research advances better -- when you can openly and freely access the results, then the working theory is that if you're a research funder or an organization that's supporting research, that policies that encourage open access to all the results will accelerate and improve these outcomes. That's really been a driving force behind why we've seen such a large scale movement in the research funding community to begin to request and, in many cases now, require open access to research articles and to research data in specific.

I'm going to stop and take a little bit of a step back and give you a 3 minute – I promise I will not bore you with U.S. policy language for the entire time, but it's important to understand that policies -- I've learned this from the hard experience of living in DC for the last 34 years, working in the policy environment for SPARC for the last 12 -- policies don't just emerge out of thin air. There's no appetite, zero appetite in Congress and in the agencies for coming up with a new policy that doesn't have some kind of precedent behind it. Starting something new, nobody wants to be out on a limb by themselves. They don't want to be first, but they don't want to be left behind. All the policy language that you'll see evolving and the directions that we're going to continue to go in, by and large, are built on foundational policies that have come before them and continue to evolve. That is critically important to understand about the data policy world in general. It's a very evolutionary environment. On the

one hand, that's fantastic because we have lots of opportunities to influence how it progresses. That's something we'll talk about -- opportunities for us to do on our campuses -- at the end of this conversation. It's also intimidating because it means you don't get perfect policies out of the gate. For somebody who wants free, immediate open access with no barriers and full open licensing now, it's hard to understand that policy making is incremental.

In the U.S., we have a really great framework that our data policies have been based on, starting with things as foundational as the positive sections of our copyright code and, moving into things that may seem a little bit obscure, but guidance from the Office of Management and Budget Circulars like OMB A-130 which talked about the importance of access to government information in the late 70s and 1980s – foundational language that talked about open and unrestricted access to public information at no more than the cost of dissemination, pre internet. The value of being able to share collectively information that tax payers have funded, whatever that information may be -- this is not specific to research data; this is just government information -- is a foundational tenet in our underlying policy constructs in DC, which is great and it gave us a lot to build on.

When the Obama administration took office in 2009, on the very first day in office, President Obama issued what's become known as the Open Government Directive, which talked about using open as the default for sharing government information. Again, not focusing on research data, but laying that foundational expectation that policies were going to be working towards making open the default for government funded information. This Open Government Directive did talk to agencies about sharing data sets and actually had a requirement that each federal agency would identify -- in 2009 this was really ambitious -- two high value data sets and figure out how to make it openly available to the public. I loved – Data.gov got a bad rap, initially. People were saying it's not sharing great data. I loved it because one of the things that they made available right away was the White House visitors logs. You're like a yenta and you want

to know who's in and out of the White House. I thought that was the greatest thing since I was on it every day going, "Who's visiting? How did they get in and what are they talking about?" It piqued people's attention, though. It made agencies understand that we're going to dip our toe in the water and then this administration is committed to driving the train forward in open. And they certainly did.

The Obama administration followed up that directive at the beginning of their second term with an Executive Order that moved the ball forward. It's really important for our community to understand that this executive directive actually began to define. What do we mean by open? When we're talking about sharing data, is it ok to just put up an Excel spreadsheet and say download this and have at it? Is that meaningful open data? Having the President of the United States say open and machine readable is the default for government data and we're going to start to talk about formats that define effective open data was quite enormous. As I'll mention later on, there's a move afoot in Congress to codify this language into legal statute, coming from a couple different directions which is fascinating. When you hear members of Congress… I went to a coalition meeting on the Financial Transparency Act which is talking about financial data generated by the government. Two members of Congress -- I'm not making this up -- who support the bill talked about not only standardized and structured data, they also used the terms we need articles marked up in structured language like XML.

Professor Krystyna Matusiak   They're talking about metadata?

Heather Joseph   They're beginning to talk about metadata. They may not understand it in a scientific research data context, but when it's regulations or financial information that they're interested in taking down barriers for U.S. companies to get data on emerging companies to invest in, then they begin to care very deeply. They're beginning to take the nomenclature on board. They probably thought I was nuts. I'm doing a little dance going there's Congress people talking

about XML. This is huge. This is gigantic. They are understanding, coming from different perspectives, the importance of open data. This laid the groundwork for that to happen.

I want to emphasize, it's not just the U.S. who's moving in this direction. As we got farther into this decade and as we started to hone in on research data and data policies began to emerge, the European Commission has been moving steadily in the same direction as the U.S. and has, in fact, outpaced us a little bit. In 2014, they issued what they called their Horizon 2020 Policy framework. They do seven years of funding for European-wide research funding. Horizon 2020 covers 2014 through 2020. The goal in the European Commission's framework for this is to make all articles reporting on European Commission funded research and all data resulting from European Commission funded research fully open by 2020. They have a very aggressive target and timeline and have honed in on moving the ball forward.

The White House, right after issuing the open and machine readable directive, issued a second directive. It wasn't an Executive Order. It's a memo from the Science Advisor talking about the importance of these kinds of terms and conditions applying to research data. I'm sure you're familiar with the Holdren Memo which looks like Horizon 2020, but with no set end date and no very specific requirements other than agencies that fund research in the U.S. had to come up with plans for policies and policies for requiring open sharing of research articles and data. That has progressed. Both of these things -- what I want to emphasize -- were new, large scale requirements for researchers to begin to make that mindset. The handwriting's on the wall that we're saying, as the single largest funders in Europe and the single largest funders in the United States, that they're agreeing that open needs to be the default in these large scale policy efforts.

It's interesting to look at how both groups define open. We can talk about open and you might think it means immediately, freely available to read, but when

we're talking about data that's not enough. It's actually not enough for articles, either. The working definition that we use for open access and open data is freely, immediately available coupled with the rights to use the output fully in the digital environment. You can see that the European Commission is really getting explicit. You need to access this information to read it, but also to mine it, exploit it, reproduce it, disseminate it. They're beginning to say the utility of this material is just as important as the accessibility of this material. The White House directive is kind of lighter language but is certainly signaling the importance of utility, as well. The intent of these things is clear. Where we are now, especially in the research data policy environment, is that as a community working with the policy makers on whatever front we can, we need to get specific about the terms and conditions to truly get to open.

Again as I just mentioned, for us -- and I readily admit this is aggressive -- for articles, from SPARC's perspective, the ideal terms and conditions would be no embargo period, immediate availability with a CCBY license – attribution only. You can do anything you want as long as you properly attribute the work to the original author. For data, we would love to see no embargo with a CC0 license which essentially has those same terms and conditions. We're not naïve. There's a little asterisk on these things. We're not naïve. We know that that's not humanly possible right now or maybe even ever. It's important, we've found, to have an ideal; to be able to articulate this would be in a perfect world where we would get to. We recognize that not all disciplines are going to get there at the same time. In fact, maybe not all disciplines should get to that ending. But, to be able to articulate it to drive towards that goal has been critically important in this incremental world to make progress towards having things move in the right direction in any kind of large scale.

I love how the European Commission talks about defining open for data. They're pragmatic. It's just perfect. They say we would like it to be open on day one carrying CC0, but the reality is we're going to work to make it as open as possible and as closed as necessary. When we think about medical data,

anything that has personal identifiers, anything you look at in longitudinal studies that can be tracked back to people, you think of course. The asterisk is there for a reason. We do recognize that there are many, many instances where it would be counterproductive and even dangerous to say this is the default for everybody and you have to do it, no questions asked. So, I really like this pragmatic approach. I will say when we talk about open access to articles, I'm not quite as sanguine. I wouldn't say this about articles. I definitely draw a harder line – SPARC draws a harder line cause this is information that that has no personal disclosures, it's not in the pre-competitive space, and we're not talking about the potential to patent this material. I know you have to leave to go to an IP meeting soon, so I won't talk any more about intellectual property so you're not OD'd on it today.

With articles, that really doesn't matter, but in the data environment we believe that it does, with one big caveat. The subset of data that underlies a paper -- in specific, the data set you would need to be able to effectively reproduce the result -- we think is very, very important to have immediately available with the paper. We do think that making a policy that governs just this subset of data is both possible and necessary to help both improve and continue to support the quality of research and science that we do. Also, I don't think it's ever been more important from a research community's perspective to be able to be data driven and to be able to point to good results in higher quality science to guarantee the veracity of scientific results. There are scientific facts and when we can point to the data that supports them and supports their reproducibility, it's an important time for us to understand how critical that is.

We're doing fairly well in terms of moving the ball forward. We're nowhere near 100% coverage or getting to the end game. In Europe, open access policies have been implemented for articles across the EU, but pilot projects for data are currently in place that are covering roughly 20% of research data. That percentage will go up each year as we get closer to 2020. In the United States, the agencies moved at different paces. That Holdren Memo came out in 2013.

It's 2017 and at the beginning of this year, we finally saw all of the agencies who were covered by that directive finally submit either plans for policies or have their plans implemented. That's 22 science agencies accounting for 99% of our total scientific R&D spend now have policies. Those do include some iteration of research data policies and that hits us where we live on our campuses.

What are they asking us to do as researchers? What are they asking us to support on our campuses? It's a first iteration policy. In the White House directive, U.S. agencies are told to maximize access to research data while still -- we're not as shorthanded as the European Commission -- protecting confidentiality and personal privacy; recognizing proprietary interests, business information, and IP; balancing the value of long-term preservation and how can we pay for this with administrative burdens. While I may read that and go "oh," have a tone of voice, this is our campus responsibility. We're the grantees. The chickens come home to roost to make sure that we're doing these things the right way on our campuses. So, we take it at SPARC very seriously. Our ideal is to try to work with the agencies and our campus constituency to try to reduce friction in compliance - have these policies move in the direction of as much harmonization of requirements as possible, so that compliance is as low, low, low overhead as we can humanly make it on our campuses.

Policies have some decent overlap in terms of what they require. All of them require that you submit a Data Management Plan as a grantee up front. You don't submit a Data Management Plan and say "here's how I'm going to make my data open" after you get your grant. You have to tell the agency "here's what I plan to do" when you're applying for funding. The NSF really broke ground on this, making this a requirement several years ago. The idea that this would be a first step. The signal was you're going to need to share your data. The second message was we don't know, as the funder, what the right thing to do or how to do it correctly for you is. So, we're going to ask the community for guidance in helping to create these Data Management Plans. Again, we have that iterative process in the data world where it's great because you're not

being told this is exactly the box you have to fit your data into to comply, but it's difficult because you're not really given a whole heck of a lot of specificity about how to do this, as we'll talk about in another minute. That is a commonality across almost all plans. Some guidance is given in the policies and plans for where do you put your data. If you're going to share your data, what's an acceptable place to put it? Obviously, it looks different for terabytes of astronomical data -- the location looks different – than for something coming out of NSF social science program. Again, a lot of opportunity for community feedback and also for library and campus involvement in public-private partnerships to think about mechanisms for preserving and making this datable -- this datable -- data accessible.

Datable. Accessible data. It's a new t-shirt. Have you seen the t-shirts with "Data is the new…" and it's always something trendy? The one that I saw in a plane leaving, I guess it was Austin, was "Data is the new bacon." I was like that's the best t-shirt I have seen in forever. That's awesome.

Commonalities also include attribution of data. We're talking about metadata. We're talking about different standards. What is important? Data has to be useful. It has to be clean. It has to be attributable. We need to know the provenance. The policies also require… Kind of interesting. Agencies don't know what they have in terms of data. So, part of these policies has also been for agencies to try to get a handle on the results of the research that they're funding. Requiring inventory, indices, ways to locate data; the role for libraries and librarians in this enterprise I think is enormous. Some of our agencies have pretty terrific libraries. NIH has the National Library of Medicine. Department of Ag. has the National Agricultural Library. But they have so many different programs and so much data that's being generated that assistance from our community is valuable. They sometimes don't know to ask, so making ourselves visible as expert resources is a really important strategy in helping to drive this forward. So lots of… Well, I would say a reasonable number of commonalities among what these 22 agencies are asking us to do in terms of the data

environment. The rub is there's not a common set of standards for any of these policy components. They're all up to community input and development. Again, that double-edged sword. That's good and bad.

What's interesting is that leading up to the change of administration, this looseness was interpreted by some policy makers and particularly in Congress that maybe we need more granular direction, maybe more prescriptive legislation. I don't that there's very many people in – certainly in the SPARC community who feel like that would be a fantastic thing to do. We would like to reduce friction, but we do think that the way to approach that is to be proactive in helping to iterate these policies and to drive compliance that meets with our ability to effectively handle them on campus. We certainly do not want to be told there's only one way to do this and unless you do this, you can't get federal dollars for research. As I mentioned, there's other pieces of legislation that are moving in Congress right now, not prescripting research data. One that you'll see us advocating for and that you'll see maybe on your campus people supporting is the OPEN Government Data Act. This bill codifies the open and machine readable as the default language of the Obama administration. Congress is still a few steps behind legislating where the Executive Branch set the direction. We can talk about that later on if you would like to.

This new administration has created a fluid political climate. It should say "is also posing." Sorry, I was… See, the fluid political climate when you live in Washington is very challenging and sometimes mucks with my ability to type properly. It's posing some very interesting challenges in the data world. I know you are all aware of the concerns that have arisen recently over what some are calling at risk government data sets, scientific data sets, particularly in the environmental and climate change areas. Some are calling it endangered research data sets. There's been lots of media coverage about activism around recognizing that scientific data, research data, is critically important in many venues. It's been the library community that has, by and large, begun to self-organize and organically step up to the plate and say there are things that we

can do. We have expertise and we have the capacity to begin to -- not talk about let's rend our clothes and talk about this is awful, this stuff could go away -- what can we do to step in and take action. You may have seen that over the course of the last several months, since January, there have been spontaneous Datarescue Events -- again, by and large, organized by our academic and research libraries on different campuses – where the library is going out, identifying data sets that are potentially endangered or at risk in federal agencies, and organizing ways to take that data, take it onboard, and steward and protect and preserve it locally on campus.

They started as individual efforts among our libraries. This is a real, amazing, organic, activist – it's the activist gene that I've seen in the library community coming to the surface. You laugh. I have a slide, it's not in this deck, but when I first started at SPARC I was so surprised by how willing the library community was to be politically active when the PATRIOT Act was being debated. One of the challenges was to a public library in Mystic, Connecticut. They wanted to look at user history on a computer in a library in Mystic. The librarians organized around "no, privacy is crucial for our patrons" and they had t-shirts that said "Libraries: the thin blue line between you and the FBI." I thought, "I love these people. I want to work for these people."

So, these Datarescue events that have been started as one-offs are now beginning to find more of a coordinated home. There's a meeting in Washington, May 8th and 9th, of national library groups, Internet Archive, a set of funders – private funders at this point, I don't know if there's any public funders that are coming, a whole host of interested organizations – Mozilla Science, people who are interested in saying those one-off events of Datarescue were really important. That was a great first step. How do we look at this and begin to build it into something that we can sustain as a community effort and that we can encourage large scale investment in infrastructure in supporting? Infrastructure not only with our federal agencies, but locally on our campuses. The libraries have raised our hands. We said we're willing to do this. This is

incredibly important information. We're positioned with the skillset and expertise to help make sure that this stuff goes forward. Now let's come up with a plan to do this under a more organized umbrella. This fits in so nicely with this notion of, ok, so we have these data sharing requirements from the federal agencies that say you've gotta put your data somewhere; where's it going to go? We have a tremendous opportunity, collectively as campuses, to say: what is the role that we're equipped to play and that we want to play; what's the level of funding that we would need to be able to really step up to the plate.

Something else that's been happening in the background is that public funders have been in a bit of turmoil right now. I mean, there's a very uncertain funding climate for them. Private funders of scientific research have been self-organizing, as well. Led by the Gates Foundation, the Arnold Foundation, Soros Foundation, Sloan, Doris Duke, Robert Wood Johnson, they self-organized a group called the Open Research Funders Group. They're meeting every six to eight weeks to talk about their own policies for supporting open access to the data and the articles that they're producing. But, also, they're beginning to talk about what role can they play by working together collectively. This group represents together, now, it's nine funders with a collective $6 billion a year investment in scientific research, with other funders interested in joining. What can they invest in that will help support making open the default in this environment? This is an ideal target for them to think about: investing in infrastructure on our campuses across libraries. To be able to do this, there's an immediate need. So, we'll see what happens after the conversation on May 8th and 9th. Coming back to a campus near you soon, hopefully.

Finally, I just want to end with a few thoughts on: what does this mean for our campuses; what are other things we can think about doing on the local level; what roles can we play to help keep things moving in a positive direction? First and foremost, we can help generate and support tools to help our communities make sense of these emerging requirements. Sorry, water break there. The emerging requirements are coming from funders, particularly the public

funders. We've seen a couple of these emerging. If you go to the SPARC website, one of the first things that we did when the policies started for data sharing and article sharing was to create with one of our member libraries, Johns Hopkins University Library, an online resource that you can go to and you can look and compare and contrast. What does NSF require me to do in data sharing versus Department of Energy? The plans are all in there. The policies are chunked out into requirements and you can take a look and see. If I've got a researcher on my campus… Are you leaving me? Nice to meet you. If I've got a researcher on my campus who has funding from several agencies, how do I support them in complying with both of those policies as easily as possible? If you've got a researcher who's submitting a grant and they're turning to you and saying, "I don't know how to create a Data Management Plan. What is a DMP? Where do I start?" there are openly available tools like the DMPTool – Data Management Planning Tool – which helps walk you through elements of creating a Data Management Plan. These are all things that are coming out of our campus environments, out of our library environments. There are super helpful and they're meeting immediate needs for questions that we are hearing, that people on our campus are asking. They don't know where else to turn. They're going to the Office of Sponsored Research and they're trying to help, but it's a really tremendous, effective, important role that we on our campuses, especially in the libraries, can step up to the plate and begin to provide.

A second opportunity is obviously advocacy with the iterative nature of these policies. As I mentioned earlier, sometimes the agencies don't know that they need our help or our input. Sometimes they put out calls for comments that are organized. But in general, they tend to welcome positive suggestions. Or this isn't working and what would make it better? What would make it easier? What would make your policy more effective is if you would do x, y, or z. If you're not comfortable as an individual or an individual campus speaking out, you can always come to one of the library organizations. Come to SPARC. If you're an ARL member, if you're an ALA member, ACRL. We all have policy offices… Offices? We're six people at SPARC. It seems funny to have a policy office. It's

three physical little, square spaces. Just email us. We can help bring those messages out there. We don't know what you know on the ground, so we want to advocate for the right thing. We do have the privilege of having built very, very solid, mutually respectful communications channels into the majority of the federal agencies, into what was OSCP and what hopefully will continue to be OSCP. Lots of mechanisms for us to be able to help channel what you're seeing on the ground and what would be useful to you directly to folks who are doing these policy things. It's clear to us that we need to evolve these policies at paces and in directions that are not only acceptable to our campuses and folks who are doing research, but sustainable by the research enterprise. So that means if there's a need for additional investment, we might think we know where those investments need to be made, but we really don't know unless you tell us, "I would like to be part of Datarefuge. I'd like to take on data sets at this level. But in order for me to do it, this is what I would require." We're sometimes shy about saying we need more money because libraries, we're not traditionally the target for people tossing money at us randomly and we're not going to become that. But, we do have incredibly critical expertise at a moment in a time where there are people who are going to be asking you, "can you do this and what do you need to do it?" Being prepared to answer that question really directly and really honestly with solid numbers is going to be very, very important.

I skipped around a little bit here, but in terms of making our voices heard, a goal for us… We'll use the word harmonizing. This came out of the Canadian policy making environment where they have three agencies and they all work together. They didn't talk about "we're going to coordinate our policies," they talked about harmonizing the elements of their policies. I just love the word cause it makes policy work seem like you're just going to go and sing a duet together. It just makes it feel more human and achievable, honestly. Sometimes you think we argue over legislation, whether it should be hereafter or heretofore, that level of ridiculous granularity which does make a difference. When you're thinking about harmonizing it and talking about we just want to be working with you in a way that works for both of us and that the result sings – if

you'll pardon the pun – is really what we want to do. So, suggestions, what you're seeing, observations are critically important to bring forward cause you guys are the boots on the ground on your campuses all over. We really do need to hear from you about what you need.

I'm just going to close by saying publicly funded research data, making it openly available, has been a trend that we have seen catch fire not only in the United States, but globally. We've seen the added impetus, interest, and drive of private funders saying we recognize that this is crucially important. We have an interesting political climate here in the United States where we don't really know. I would be lying if I said we could read tea leaves and we understand what the future of scientific and scholarly research data looks like from the public sector here. We don't know. It's a rocky road. It's uncertain. But that has opened up an opportunity for us collectively, as a community, to step up and get more actively engaged and to help drive it in the direction that we feel is the most beneficial for our researchers, for our campuses, for our society. So, I hope that you will all heed that opportunity and jump in with ideas, with thoughts, with suggestions. I will stop here and talk about any aspect of this that you like. Thank you very much.

Questions? Concerns? Yes? And can you tell me who you are when you…?

Doug Rippey	Sure. My name is Doug Rippey. I work in the libraries as a Metadata Technician. Therefore, I'm gonna ask you a metadata question. Do you have metadata that we could put into our discovery system on the multiple agency policy document online tool and the Data Management Plan Tool, so that we could make those live, visible, and discoverable the same way as Chicago Manual of Style, dictionaries, and products we purchase?
Heather Joseph	We do not and that is an excellent suggestion. I think that would be something that we would want to do in the next iteration of this tool. We have grand designs to make it much more interactive. If I could ask you to do me a huge

favor, can you just put that request into an email, send it to me so I can give it to the team, and say this is something that would be fantastic to do early in the next iteration? Because I suspect you are not the only person who will think now that we've got the plans in there, we're talking about: what should we do next, how do we make it most useful? If that is one thing that we can do, we can look to put together the resources to get that done. Would you mind doing that?

Doug Rippey

I would like to.

Heather Joseph

That would be excellent. There I am, up there, email-wise. Nick, if you're listening, it's coming to you. Yes?

Krystyna Matusiak

Krystyna Matusiak, faculty in the Library and Information Science program. I want to say that librarians and LIS Faculty should be champions for all that, for open publications and data. So, I think the most important thing is that faculty on campus really participate. My question is: How do you build community now with faculty? There's not enough awareness and it's just an extra thing that faculty has to do. So, if you have suggestions…

Heather Joseph

So, another great question and there's not one straight, silver bullet, easy answer. There's a couple things, though, that I think are important. One of the things that we've recognized is that we need to be able to make making the choice to publish in an open access journal or to deposit in an open access repository or share your data easy and integrated into the workflow of faculty members. How do we do that? There's a couple, I mean, there's lots of different ways that we can go about enabling it. Some are technical. Some are saying, well, when we're onboarding faculty, we should be talking about "this university values open access." We need to have more, not during tenure and promotion. That shouldn't be the first time that faculty hears that open is important on campus. So, we're starting to think about ways to raise the awareness of open as something that the campus supports.

What does that mean? Well, frankly, it means working with administration to be able to express support for open. If it's a resolution that says open should be the default mode on our campus. We've seen some campuses take that route. We've seen some campuses take the route of we're going to have a campus policy that basically says if you're a faculty member, this is the expectation. Still on most campuses, we haven't seen that sort of holistic integration, this notion that… We actually had, it was great, we were doing a forum like this and we had a young assistant professor starting her first tenure track job at the University of Vermont and she said, "You missed the opportunity to get us, which is at new faculty orientation, when we're literally too scared to leave the room because we think we're going to miss the thing that we need to be successful on campus." She said, "If you told me that I have to give priority to making my data and my articles open, then, I would have written it down and I would have done it diligently." There's that sort of technical intervention, but that has to be coupled with the acceptance that open is, in fact, a priority that needs to incentivized and rewarded on our campuses equally or in place of publication in traditional subscription access journals that carry high impact factors, which tend to still be subscription journals.

How do we have that conversation? We have some ideas about how to do that. We've been thinking about this for quite some time. We have a research project that some of our early career faculty, including the young woman who raised that question, applied for grant funding and they're collecting tenure and promotion guidelines from across campuses in the U.S. They're looking to see what is actually rewarded and can we find any instances where open is actually rewarded. Just simply collecting that information gives us a baseline to start from. Really, the question that needs to be asked by our university leadership and our research funder leadership: we may be requiring that you do certain things, but are we rewarding them adequately? It's kind of interesting because as much as I feel like we, as the library community, can do almost anything we set our minds to, I don't think we're the right community to be leading reform

of tenure and promotion guidelines. We can raise the issue and say, hey, we're being asked "I would like to do this, but are we going to be rewarded?" We can't really change those guidelines. But, I do think that research funders, in particular this open research group of funders who are getting together, they're starting to say we're trying. They believe that open is the way to go, as the default, as an enabling strategy to achieve their core missions. The Gates Foundation has stood up and said -- it was in the press release for this group -- they said, "We believe that we should use open access to our data and articles because it helps us achieve our core mission of saving lives in the global health initiative." Right to that core mission. When they're asking themselves questions -- Ok, we're requiring it, are we rewarding this behavior adequately? Are we rewarding behaviors that advance our missions? -- that's the question that we need our campuses to ask. Are we actually rewarding publishing communications behaviors of our research outputs that are fueling our core missions? We'd like to see that conversation happen in a neutral space. Maybe have the National Academies convene, research funders, administrators, and faculty members who these things are going to apply to. Simply ask them to ask that question. Are we currently rewarding behaviors? We know the answer's going to be no, not adequately, or open access would be – people would find it much more appealing. Once you answer the question "no," we would like those groups to begin to make recommendations about what do incentives look like that reward behaviors that do advance our core missions. When you look at university or college or community college mission statements, they have some iteration of the same language: production of knowledge, dissemination of knowledge, public engagement, engage citizenry. All of those things, that when you look at opening up access to research results, educational resources, open is an enabling strategy that's right into advancing the core mission of, I would wager, every academic institution that we have. Getting that recognition at the level that it needs to be heard at is going to be key for us to do that. It doesn't mean we stop our grass roots efforts, but getting that recognition and having the value of open, the outcomes that open can help achieve, I think are crucially important.

I'll follow that with just one little plug. Open Access Week this year, if you guys follow Open Access Week, we do global Open Access Week in October. We organize around a theme every year. Last year, we organized around an action-oriented theme. How can you act on open? What can you do? What's one thing, big, little, medium, whatever, that you can do to advance open on your campus? This year, we're going to do it a little bit differently. I think it speaks to the heart of your question. We've asked ourselves at SPARC… We're not shy about saying -- we've been talking about this for a long time -- we've made some progress, we're not all the way there yet. What are we doing wrong? What aren't we getting across effectively enough? Why aren't we bringing people on board? We can blame tenure and promotion. Fine. What's our strategy for dealing with that? Part of it is that we brought in a communications expert, this wonderful woman, Deborah Rosen, who's done communications for national campaigns on all kinds of different topics. She knew nothing about open access, nothing about us. She immediately came to us after doing a landscape analysis and interviews and said you guys are wonderful, love you, fantastic thing, here's what you do. You talk about open access as if it's the end game, as if it's open access for the sake of open access. Like, open's better than closed, right? Yeah, you're gonna get about 20% of the population who thinks open is better than closed and who is willing to go out on that limb with you. What you really need to do is say open access in order to do what. Open in order to… Open data in order to share data quickly to prevent a Zika pandemic. Open access to journal articles in order to accelerate progress towards a cure for cancer. Open access to educational resources in order to make college affordable to every student. And it was that. Who is your audience? Who are you talking to? Why should you care about open? What is your job? What is your goal? What is your personal stake in this enterprise? Speak directly to can this help you, personally, professionally, and then societally. So, we're going to organize around the "open in order to" theme and ask people to ask themselves those questions and see how that shifts the conversations we have on our

campus, as well. So, we'll see what effect we get from that. We're very grateful. Deborah Rosen, if you ever need a good person to talk to. Yes?

[Male Voice] Is there a way that [unintelligible] nature and operation friction [unintelligible] our initiatives [unintelligible]. But, as far as [unintelligible] have you heard anything about [unintelligible] their positions [unintelligible] commercial interests.

Heather Joseph That's the most interesting segment we can ask the "open in order to." Why is Elsevier promoting open access? There are some in the company who will say, "But I genuinely believe that it's to promote research, dissemination of research." But, commercial companies are tasked with making a profit to return to shareholders. That is their job. Is it legitimate to say "open in order to maximize profits to deliver to my shareholders?" Is that something that we want to support? I'm not gonna answer that question. I know how I would answer it. I gave a keynote at OASPA - Open Access Scholarly Publishers Association, the trade association for publishers, and presented that idea and the commercial publishers did not like it at all. I can understand why. But, I do think, for us, we have to be clear-eyed and we have to ask those questions and if they surprise us and can say open in order to profit maximize, it's good for the research enterprise these ways, are you taking those profits and reinvesting them in our libraries, in our research enterprise? What's happening? Or is it just that we all suddenly, as a community, say we're fine with 40% profit margins going into shareholders' pockets instead of an alternative? We'll see. But not, I would say… That strategy hasn't been effective in moving. They're not suddenly going, "Oh, that's what we've been doing? Gosh! Didn't realize it." So, we'll see. We'll see what happens. It's an interesting time. Other questions? Yes?

[Female Voice] I have a couple questions.

Heather Joseph Shoot. Fire away.

[Female Voice]	[unintelligible]

Heather Joseph	So, Smithsonian is covered by the blanket directive, the executive order, and it was also one of the agencies that's covered under the Holdren Memo, the second… the research data and articles. So, they are. They do have a public access plan that's published. You can find it on the site. So, yes, it does. They have some interesting mechanisms that they're working on for publications, but data as well. So, yes?

[Female Voice]	Also, then, do you feel like you're finding deep resistance because of the [unintelligible]?

Heather Joseph	In articles, it is the commercial publishing industry and lobby. It's not a secret. They have fought back. They have lobbied hard in Congress, in the European Commission, against articles. In the data world, we have not run up against commercial interests. The most challenging constituency are the individual researchers and individual research communities. I understand that. Part of the reason that I try to be so careful to put out the "as open as possible, as closed as necessary" language is because, immediately, researchers go and look at the worst case scenario or the scenario of "don't make me do this on day one because…" or "don't make me do this because…" My husband is actually a social scientist and he said to me, "I'm with you on articles. I'm good. But, don't touch my data. It took me five years to collect this and it's gonna take me time to get through this. Don't tell me, please, to make it open." And he's like, "It's not that I don't want to advance discipline. I want to wrap my mind around my work, first." I can't, not just because he's my husband, but I can't argue with that. Also, of course with privacy and medical things. But researchers, as a constituency, are the ones that are the toughest constituency to move en masse on this. Yes?

[Male Voice]	Can you talk about the area where [unintelligible] kind of speed that process along [unintelligible] part of the issue is that, but it's also that it's not

documented well. [Unintelligible] no effort at all [unintelligible] provided by somebody out there?

Heather Joseph   It's certainly a service that could be provided by somebody out there. It's kind of outside of the expertise of SPARC. I mean, again, we're like six full time people. What we're trying to do is find those pressure points, right? This is more the building tools. It fits between advocacy and building specific tools. There's a whole conversation. We're talking this week about open scholarship, but there's also this other title movement called open science. We're all the same. Open data. Open access. Open ed. All part of open science, which I think feeds up into open scholarship. We're all talking about the same things. The open science movement, though, there are some players that we're working hard to try to – we want to all be working together. Again, I use this political climate not as an excuse, but as an eye opener for how important it is for us not to differentiate into little silos in the open community. We have to be working openly, linked arms, strength in numbers, standing together and supporting each other. The open science community, I think, encompasses groups like Mozilla and encompasses groups that have different skill sets, different pieces of expertise that they may be in a better position to move forward with what you're talking about.

We had a visitor at SPARC last week, Johannes Vogel, who is the European Commission's new point person for Open Science Policy Platform. He's the curator of the Berlin Natural History Museum. So, his approach to open science… He was just wonderful to talk to, first of all. He's one of the most interesting, energetic, positive people I've seen in this space. His primary interest in open science and his definition of open science was having everyone feel that they had a stake in science and that meant the public and citizen science. His end game was public engagement. Getting public engagement requires that the materials you're putting out there are in a form that the public can readily engage with. So, on the one hand that means accessible language and things and accessibility for any kind of reading disabled, any type of

disability, making things accessible. But then, with data, that the data's clean, that it's in context, that it's annotated properly. All of those pieces are essential components. I was thinking he's coming at it from… We're coming at it from the campus perspective. He's coming at it from the public perspective. It's really interesting to walk through what are you doing to get there and how can we connect? It seems to me that what you're talking about is one of those important, connected pieces. We should and will look for places… Again, if you send us suggestions and we can't do it, we'll try to reach out in the wider, open world to see if there are people that might be able to take your idea and work with you and run with it.

We're doing things, too, just apropos to your question about commercial interests, cause we don't want to just dismiss commercial as being bad. There's nothing wrong with making money. We think there's maybe better places to make those kinds of margins, although no other company is doing it. I actually have a slide that shows relative profit margins of Apple, Coca-Cola, some oil companies, and commercial publishers. The commercial publishers outstrip them in percentage of pure profit. But, there's a value in having the commercial enterprises understand and support our agenda. We really need businesses, tech companies, entrepreneurs to stand up and say, especially now, open as an enabling strategy will help us create new businesses, create jobs, fuel the economy. We believe that that is the case. We've had economists do work in that area, but unless we have the commercial sector standing up and saying it's a good idea, we're not – we're always gonna be at a disadvantage in terms of policy advocacy.

So, one of the things we did as an organization this year was SPARC joined a coalition called the Data Coalition that was founded to advocate for things like the Open Government Data Act, the Financial Transparency Act – which is where we heard Congress people talking about XML and structured markup of articles. PDF isn't good enough, you need something you can text and data mine. I was like, my heart is singing with gladness. But, one of the reasons that

we joined, frankly, was to make sure that the library community was in coalitions and at the table with members like the Data Coalition has, which are Intel, Amazon, larger players who are interested in our space but who we can talk, then, in coalition meetings. We're supporting issues that are core to them. We're explaining our interests and how the library community's interests overlap and then asking them to support us as we go forward. We don't dismiss the importance of commercial enterprise by any stretch of the imagination. We want to be as strategic as possible in figuring out ways that we can engage and work with them effectively as possible. So, we kind of do that as well. Very, very interesting. You feel a little bit like a fish out of water, sometimes, when they're talking about these large scale ICC disclosures and I'm like, "I'm working on journals." They're like, "That's cool, that's cool." Kinda get a pat on the head. But, they're understanding it in a different way.

The irony of the two Congress people who were – there were three who were talking… But for those of you who have been engaged in open access policy advocacy, you'll know that after the NIH came first out of the box putting their policy forward and having it passed into law, one of the things the publishing industry did was lobby to overturn the NIH policy by advocating for two pieces of legislation sort of back to back. One was called the Research Works Act that Representatives Maloney and Issa co-sponsored. They're the two Representatives who are supporting the Financial Transparency Act and co-sponsoring it who used those words about how important these elements were. I mean, it was great because they're on the referring committee to our legislation public access, the fast road legislation. We're like, whatever the education process is to understand and to kinda come around, it made me more optimistic that when they see it, if it's important, if it's good enough for financial information, why would it be any different in the scientific environment. We're looking forward to addition conversations in that committee at those offices. Who knows? Maybe we'll see positive direction.

Any additional questions? Yes?

[Female Voice]   Um, so um, I was just looking at the correlations [unintelligible] congress [unintelligible] was in the context of the state community stuff [unintelligible] and what other librarians are seeing that there's things that negatively impact sharing data or sharing research, that there's this real strange fear that their data is gonna be misinterpreted, misused [unintelligible] talking about [unintelligible] but it's really not that [unintelligible]. Like even if I have this [unintelligible] data [unintelligible], but then anybody can do anything they want with it. It's almost like there are undertones like, unless you're starting off and you have a PhD in astronomy you won't understand my data and I don't want to share it cause I don't know what you're gonna do with it. I just wondered if you've seen that in your work and how do you respond [unintelligible]?

Heather Joseph   We definitely see it. I think it's an understandable, human, backlash reaction of fear in an environment where you don't know anymore. It feels like up is down, down is up. Can my data be taken and misinterpreted? Yes. Is it a real possibility? Yes. Is it likely? Probably not. Should it stop you from trying to openly communicate with as many people as possible? My personal – I'm not speaking on SPARC's behalf cause we haven't talked about this, but... Absolutely not. I really feel like sunlight is the best disinfectant. Right? You make things openly available. You can refute. You can defend. You can back up. If you can't, then you might not be so confident about sharing your data, but if you're confident in your data and you understand that it may not be perfect. This is science, right? This is how it advances. I think we should be sharing it.

The flip side of that, in the era of alternative facts and things, one of the things the library community is doing - and you may be aware of it - is looking at ways to provide high quality information and sources that people go to that may not be in the library. So, Wikipedia. This idea… There's a campaign called – it's a hashtag – 1Lib1Ref, one librarian, one reference. Wikipedia's asking librarians to just take ten minutes, go to a Wikipedia entry that you have expertise in, and

supplement it with one high quality reference that you can source back that reference. So, to really up the quality of public information, we're trying to look at other kinds of actions like that that we can do with people like Wikipedia, where it's low overhead for our community but where you can help raise the quality bar of information that's available. This is the kind of stuff, the stuff that we do, we're not just doing this for our own help, right? This is important. We have to be confident in what's out there and we put out there and we can stand behind it. I think that's one of the ways that we can help to combat this fear level and help ratchet it down. We have each other's backs. We have researchers' backs, they have ours. We have the communities' backs. We can point to here's how do you know something is good. Here's how we can help you. The idea of data dictionaries and things like that. They're exactly the kinds of things that we can do to help combat fear. We're not gonna change that human reaction, but we can help mitigate its affect by utilizing our skill sets to their fullest degree. Does that help?

Wikipedia's pretty open and they love librarians. They want to work with us on things. Again, if you have ideas for working with them, they're phenomenal and phenomenally open as an organization. They recognize where their strengths are and where their weaknesses are. I think it was wonderful to see them reach out and to really be deliberate about saying the library community is where we go for good, high quality, verifiable information. It's a crucial role that libraries play and I love that kind of validation for the importance of our work and the effectiveness of what we do, not just on campus but extending it into the creation of resources that the world uses.

[Male Voice]   Question.

Heather Joseph   Ok, please.

[Male Voice]	Same conference, but [unintelligible] institutional level policies and mandates have put [unintelligible] ourselves better [unintelligible] data rescue [unintelligible] all this stuff out. [Unintelligible.]
Heather Joseph	I do think… Institutional mandates have been designed to serve different purposes. I think some universities or some campuses have hoped that it will populate their repositories. Certainly, that was Harvard's – one of Harvard's – main priorities and DASH is reasonably well populated, right? That was a priority they had, a strategy across all of their schools to explicitly do that. Some campuses, the policy is more of a statement of intent rather than we're gonna populate our repository. Some, frankly, I'd be the first to admit have wanted to use it as a lever for getting more deposits into their repositories and it hasn't worked. I think one of the things we need to do is look at what's effective in a campus policy, what isn't necessarily effective in a campus mandate.

One of the things that's important or useful – I promise I will address the data refuge question in a minute. But in terms of the utility of campus policies, besides signaling the intent, one other element that's really important and that has been crucially helpful is when your campus policy is a rights retention policy. Again, Harvard, MIT, University of California, they essentially say the university reserves an irrevocable, worldwide, non-exclusive distribution license so that researchers feel more comfortable complying with the policy and the university's actually positioned itself to say, legally, we have the ability and the rights to the output of research on our campus. It feels like a small piece, but it's something that has really made the universities that have put rights retention based policies in place feel much more confident about asserting, "we're going to require, request, defend." The researchers like it because it means if there's a problem or the publisher has a question with their individual article, they don't come to the researcher. They come to the sponsored research or the copyright officer. It's the university muscle who the publishers have to take it up with and that's been really effective in making some publishers change policies. So, I think those are good elements in terms of populating repositories. There's a lot

we can unpack, there.

I think data refuge is a fantastic thing for people to think about and to think about the potential utility of our existing repositories. But I think there was also another session at CNI Roundtables on the future of repositories. Certainly the coalition, the Confederation of Open Access Repositories, or COAR, just issued a white paper on next generation repositories. Really, really important because one of the things that we're seeing is that while we were very well intended across our campuses in building our initial suite of repositories, many of the repositories are built – and I'm gonna quote Herbert Van de Sompel, here, who is one of the technical architects of the architecture of our initial set of repositories – we built them as digital librarians early on in the evolution of the internet and of the web. We didn't focus on building them as web native enabling technologies. Herbert gave this fantastic talk, actually at a COAR meeting, where he said, "You know, when we were first designing the architecture for repositories, we were doing it as digital librarians and people were saying things like, 'Uch, HTTP protocol' and we didn't really think about embedding that as central. Here's why. Last year, we used Gopher and Gopher went away. We were thinking, meh, this is just another communications protocol that'll probably be ephemeral. We just didn't know." So the focus of our repository architecture has not been to build web native enabling technologies. I think that's one of the things we really have to think about.

I think data refuge gives us an opportunity to rethink. Are our repositories structured properly? Do we actually all need an individual one? What could - with some infusion of capital to help us reinvest, knowing what we know now about what makes things discoverable, useful, and really valuable in this web native environment – what might we rethink about the nature of our repositories? I really would encourage you, if you're interested in that, take a look at that white paper. Herbert's one of the authors of it. And maybe we can… I think it's a valuable conversation for us to have. It doesn't mean our repositories failed. It makes me crazy when people say that. Are they as

successful as they could be? No. Do we understand some of the reasons why? Yes. Can we fix them and make them more useful and a crucial part of the infrastructure that supports open? Abso-freaking-lutely. Do we need more help and money to do it? Again, yes. I won't say abso-freaking-lutely twice, but I just did. I mean, I do think that that's an important thing. We have to be realistic. Yeah?

[Female Voice]   [Unintelligible]

Heather Joseph   Entirely possible. I don't – technically I don't know. I don't know if LOCKSS, if Vicki or David, are gonna be at this meeting, but I would bet that they are. It seems like… Um. Mike Furlough runs Hathi. So, Hathi will be there. Things like that, the shared infrastructure. People who have this experience in doing this on that scale need to be at the table and we need to draw from the lessons of these fantastic community collaborative projects. What can we take that's good? How do we use these experiences and drive things forward? I don't want to reinvent the wheel. This is a great opportunity. One of the things I talked with the librarians at Penn, who are the folks who got this started with data refuge, and I admire this community for many reasons, but they embody one of the main reasons. They said we got this started. We don't need to own it. We want the people who know more or who are better positioned, now, to pick it up and run with it and we'll support any way that we can. It may be being in DC, but there's so much, they're the reverse of territorial and the reverse of saying our success means we keep it, we close it. They're like yeah it's a good idea. We feel validated that the community wants it. We want it to be as successful as possible. Take it. Run, larger community. I just, I love that. I think that's how open beats closed. I think that's how we, as the academy, win over commercial. So this, I really feel like it's a big opportunity.

Other questions? Alright. Thank you guys so very much for coming and talking. Thank you guys online. I hope you found it useful and thank you.

[End of Recorded Material]