Using Science.gov to Access U.S. Government Science Information – Transcript of audio

Please stand by for realtime captions.

Good afternoon, everyone. Welcome to the FDLP Academy. I am with my GPO colleague, Ashley Dahlen. We have a trick webinar for you today . Using Science.gov to Access U.S. Government Science Information . With us today our presenters, Mary Moulton and Joanna Martin. I want to reach you a little bit about both of them heard Mary Moulton joined the National Transportation Library in 2011 following her career in academic and private sectors. The National Transportation Library is a virtual library. Mary leads and develops a digital services, information organization, digital duration and discovery and uses FDLP resources. Mary is the co-chair with an interagency group of senior sciences with several federal agencies. She has a BS in plant science and MLS and an MS in entomology. Joanna Martin joined the department of energy in 20 07. She's worked with a variety of products over the years with a special interest in the research results. Joanna became involved early on and she is a role of the co-chair along with Mary. Joanna has a BA from sociology in the American university. With that, I will turn the patrol microphone over to Mary and Joanna.

Thank you, Joe. Can you hear me?

Yes, we can.

Wonderful. I am going to start this presentation off and we will turn it over to Mary Moulton midway. I want to thank you for those introductions. I want to thank all of those for joining us today. First of all, I think that most everyone here know something about science.gov. But is curious to understand more about it and I want to describe the tool I will focus on some of its history. I think it is the most interesting aspect of the site. The latest and upcoming features for science.gov. I just realized that I skipped over this overview. I apologize to everyone. I'm going to move over to science.gov. When I say retreat results, what I mean is what we are covering here, journal articles, technical reports, conference pages, videos -- which way the videos of the conferences with different multimedia. Those are kind of just a handful of things that we are talking about. Science.gov is overseen by the science alliance. Mary and I are the co-chairs and it is the majority of agencies that fund technology research. That is across the government. Pretty much, they all participate with overlap there. You can see they have good coverage of representation with the difference results that we make available using our necessary items. Science.gov was launched in the early 2000's, 2002 to be exact. They came together to talk about ways to collaborate on information that started in the mid-'80s. Thing started to get more digitized and science.gov came about. It became a one-stop shop, if you will, to search the government stores of information. How can we make this information available in an easy way that will provide access to this information to the public? We offer real-time of the research results in a relevant way. We are providing online research results, at the top level of research results. That is really what makes science.gov special. To this day, Department of energy through the office of scientific and technical information, OSTI -- and its platform on behalf of science.gov. In terms of the search technology that is used and I think, again, this is really what makes it special. We do a federated search, but the federated search that we provide is physically -- federated search is a technology that searches across multiple disparate content sources. What we do is -- we bring in technology that will search the entire surface of the web. We will get into the deep web. That is really what many of the sources at science.gov searches. It goes out there, the search technology that we have -- that we use. It is a software called [Indiscernible] and there is other

types of a federated searches. It uses customized connections that are developed in coordination with the federal agencies -- the connectors which are really specific with code. And they connect to the science sources. It is specifically what we want to bring back. We are bringing results back that may go outside of a scope. It is really what we want to search those connectors for. We don't have a database to do this. We are going out and searching -- like I said, all of these content sources. It will go out in real time. It will search these databases or categories. It will bring back the top level results. That is really for performance. It would take a lot of time, if it searched all of the results. Probably, we would not get anything on that. It brings back the top level search results. It is all the results from our repository. It utilizes content sources relevant ranked results that they use for their repositories. With utilizing that, it will utilize [Indiscernible] for relevance rankings. It will bring back the top 100 search results to bring back what is most relevant within that topic. It will be what will be theoretically considered for that search query that will prohibit the user from having to set thousands and thousands of results. That is the intent of science.gov search. Science.gov utilizes the search technology, of course. We can go out and it will search full text. It has advanced searching. If you want to use search in our advanced search, we have categories. It is for a particular content sources. It is typically not needed, but we keep that around. I encourage folks to play around with it, we also have alerts. It is relevant ranked top level search results which can be filtered -- it will really bring back to you what you want out of your search query. The latest feature that we have implemented in public access. Mary will talk a little bit more about public access. That is a federal initiative that came out around 2013. Public access was kind of bring across the federal government. A memo was issued by by the office of technology and policy. Federal agencies had to make all of the research results available and they used various repositories of their own. This was a way for more agencies to make more literature in general articles available to the public. What we want to do to was to create connectors with the federal agency repository so that you can go search for something and search across all of those repositories. That can bring back all of these results. Everything is public access. This is mostly going to be focused on scientific research that would be more on the literature. That is something that we have worked hard to make those connectors and make that information available. It will bring back results -- you will see that the results are [Indiscernible], there is a multimedia tab. We are reconsidering the need it to keep [Indiscernible] -- I think something that will be talking about. It would then be the repository or tools and something that will be trying to add to. Another thing that will be looking at will be talking to the group about trying to identify sources for software. [Indiscernible] when it comes to these results and we stay up to date in terms of adding those. In this screen, I wanted to show you the public access connector. If you go to the advanced search, you will see on the advanced search that there is little arrows and expandable boxes. At the very bottom there is a [Indiscernible] this is the connectors that we built for the Federal Reserve. If you wanted to go and you could go to the advanced searches or you can go to this public access. I just wanted to show you the range and how many repositories are out there that we have worked to build to provide results. This slide, I would like to spend a little bit of time on. When you go into a search, I will try to use the pointer here -- I do not use WebEx all that much, please, bear with me. What I want to show you is -- I want to show you -- here -- I think you can see it. I am not sure. Hang on. I'm going to try this one more time. Monkeypox -- I think you can see this. I wanted to do a search. We have all heard about COVID. I wanted to do something that was a little bit more contemporary. We are living through the monkeypox virus. I want to see what I got back. If I use quotes, then it would give me a [Indiscernible] instead, I had 71 sources that are out there. I'm going to do another pointer. Wait a minute. I don't know if I can do multiple pointers. I can. You can see where there is 71 of 71 sources complete. When I did the search, I had to sit there for a second. It searches in real time. It does go out at that very time and search all of these individual sources. It does take a little bit longer if you went into another database ended your search that way. There is sort of a status bar where you will see -- it might say 65 of 71 sources complete. The goal is 71 of 71, then you know all your sources have been

completed. You can open that up. When you open that up, you will see which sources brought back results. I will tell everyone that is on the phone that because it is real time and we do rely on the end points, where agencies may be doing maintenance, they may be having another problem or something else going on. A source may time-out. In case, I recommend waiting and trying it at a different time. Because it is real time, you will get different results. I just wanted to explain that to everyone. I just wanted to see what the status of those searches are. That is something we try to monitor. We do try to monitor the connectors. If you are seeing this, you can always get in touch with us through science.gov . There is also another -- when I did the search, it was 71 of 71 sources. I had it set up through categories. I think maybe four -- I am trying to think. I think it is the text category, multimedia, data, and public access I highlighted public access. Because I wanted to see the journal literature. I wanted the peer-reviewed content. That is why I focused on public access. That does not seem like very much. That is because it is the top level results. It does bring back the top 100 relevant results that have that search query. That way you do not have to go through thousands of results. If it did not go through top level, you get thousands of results. You can see how many results are returned where I have the pointer. If you have any other questions after this about how that works [Indiscernible]. I emphasize that because there will be this assumption that all results are brought back, but it is just the top level results that are brought back. If you want to see the collections that have returned the results. I will try the pointer one more time. Let me see here. There is this. It says all collections. If you go to that drop-down, you will see all the collections that have returned. You can choose which results you actually want to bring back if you're more interested in that background. It could be from central, or one of the other agencies. A lot of these things might be in the search as well. If you go to the search results page, you can kind of navigate and add some of this a filtering. That is the way you search. If you're going to go and search for something -- that is the search example. If we go to the next slide, we also have a tool that is related. It is worldwidescience.gov. Worldwide science, instead of focusing on the federal and technical information, it is looking across -- not just national, but international databases in the world. It does search science.gov, but you can find international science and technical information. If you are interested patrons or working with a focus that you work with, the results here can be [Indiscernible] using over 10 languages. With that, and going over what science.gov does and how it searches. I encourage you to ask me questions when we do our chat. Also, we will provide our contact information. I like to turn it over to Mary Moulton from the Department of Transportation . She is the national transportation library [Indiscernible]. And the content that it makes available to the public. Talk a little bit about public access. With that, I'm going to turn -- I will drop the ball over to Mary. Thank you all for listening.

Thanks, Joanna. I need a moment. Great.

[Captioner cannot get audio]

It really does play a very important role in a civic and social life. We were able to make an accurate in, we are the government. We made an acronym that fit the way ROSA P. A little bit about public access. I'm going to take a bit of a deep dive here. My experience is that outside of the government, a lot of people are not aware of the details of public access and how this has really, just in a few years since we have lamented this, it has had a really profound impact on our agencies, as well as the research community. Public assess means that the public has access to publications and data that were funded by the government agency, scientific research programs. Basically, that means that all the work that has been funded with public funds -- agency funds, will be made available to the public when the research projects end. There are a couple of exceptions to this. We do not put out private information about people. We do respect confidentiality and other securities concerns. We will not make publicly available

items that might influence or erode national security. For the Department of Transportation this is a very small subset of what we have . Everything else is publicly available and discoverable. The public access plan applies to all D.O.T. employees. If you are working inside of USDOT . Your research results need to be submitted to ROSA P. They need to be made publicly available. Was this a problem before public access? Not really. You can tell -- when you come to our repository and look for the contents, that we have items going back to the late '90s. I think sometimes people hung onto some of their research. We have been very aggressive about promoting public access across USDOT as well as external to the folks that we provided research grants to. We have got good participation now across D.O.T.. That is really reflected in the collections. Since we implement the public access plan we have really been able to grow our collections and the items all have significant integrity and importance to researchers. We have already experienced a positive impact from the public access. The other issue is that, of course, we fund external research. I mentioned we have a lot of partners at the state Department of Transportation and universities. Externally funded research is mandated under our public access plan. Folks when they are finished with their research, their research results have to be submitted to the national transportation library for publication in ROSA P . That is mentioned specifically in our plan. If you are interested in looking at the plan, at the top of the page, I put a link -- the DOI, I put a link to the plan. You can take a look at that. There is a lot of confusing terminology around access. I just want to take a few moments to demystify this for you. If you're working in an academic library -- really, any organization right now, you are aware of how published articles are behind a firewall, usually. Unless you have some type of institutional license agreement, the folks who use your library who are at your institution are not able to access the content. They have to pay for it. Either buying individual articles are going through interlibrary loan. Of course, if you are using a lot of information in a specific domain, that can get unwieldy. This has been, for decades now, an issue -- a concern for librarians across all different sectors. What open access does -- this is not public access. I am talking specifically about open access. There are open access journals now that provide unrestricted access and unrestricted reuse of publications. They are usually cooperated under a creative Commons license. You will see this on the document. All of our items are in the public domain, which is another type of a right. It is a public right. All of our documents are also open access. Our public access plan covers final peer reviews manuscripts accepted for publication. We do not put publish articles in ROSA P. As Joanna was explaining, this is a big deal for science.gov because we all coordinate with each other in how we index our content. It is very easy for the science.gov search engine to pick out those things that are labeled peer-reviewed manuscripts and journal articles. I believe that some other agencies are allowing a click through to the articles in some cases and partnering with organizations like CORUS. We are not one of those agencies. For us it is all of the obligations that are produced right D.O.T. and the submitted research reports from our researchers and peer-reviewed manuscripts. That is what's is covered in our plan. An open access repository -- just to make sure that we are all on the same page here. It is a digital platform for research results. Your organization may have an open access repository. Just to repeat, access is free, immediate and permanent. Anyone can use it. We follow OAI-\PMH repositories. This is what makes science.gov feasible. Our content is indexed by Google, Google scholar, being -- this technology is really enabling for search and also for people who use our repositories. That orange symbol on the left is the open access symbol. When you see this next to a repository or content in a repository, it means that it is open access. A little bit about our collections. We collect a diverse -- a diversity of content. Most of our objects -- most of the content in ROSA P is submitted by content creators. Again, those are people who were inside our agency, but also externally funded researchers . We also digitize content at the request of submitters or colleagues. For example, if you look in the blue box on the left, you will see that those are print copies of the national transportation statistics. This was a publication -- this is a publication that we continue to put out. Although, only in digital format now. Back in the '70s, it was a print publication. You might think that national transportation statistics is important, and you would be right,

if you thought that. We embarked on a big project where we assembled a complete run of these and digitize them. Those are in the collection. That is an example of something that we would digitize and make available. The two items at the bottom on the left, the two reports at the bottom are born digital items. They were never in print. We currently have about 60,000 items in ROSA P. They reflect all modes of transportation and related disciplines. We not only collect materials that are created by USDOT, as I mentioned, stated D.O.T. , local and tribal road agencies, universities, and other transportation organizations. Science.gov benefits users and the national transportation library. Through science.gov -- science.gov is a significant driver of users to our site. It is a collaborative effort. In the total landscape of government repositories and government libraries, we are a small library. We are very focused, as I demonstrated it. Being involved in science.gov allows us to leverage the resources of other agencies. It is an efficiency that we have kind of hobbled together across the government that allows us to do this. It benefits us because, as you are probably aware, it is sometimes difficult to determine which agency may have sponsored or published research results. Those can be two different things, by the way. That national transportation location that I have on the screen a few minutes ago -- there were several iterations of that that were published with the department of energy. This was in the '70s, when we had our first energy crisis. For several years in a row, there was a publication that we did with the Department of Energy. It was published by U.S. Department of transportation, but the corporate creator was the Department of Energy. You get kind of tricky there. With science.gov, you do not have to know who published it, who did the research. Science.gov will lead you to the correct agency and the cracks repository. Transportation research is interdisciplinary. It covers all of the engineering disciplines, but also medicine, behavioral science -- that has become very important for safety is concerned -- for example, getting people to use seatbelts, planning, geography, geology. One of the things that I did when I first started working at D.O.T. 12 years ago, I asked for the search logs for our then repository. I was really astonished to find so much of the traffic that was coming to ROSA P was from science.gov . I could tell from the search logs that people did not know that they were going to find content at the Department of Transportation . It definitely is the case. We know that we are reaching a wider audience. Our content has to compete in a global information system. If you think that what the World Wide Web is -- we are very competitive. Science.gov is a tool that helps to make us competitive. That is what I have this afternoon. We are going to take a few questions, I believe. Before we do that, I do want to acknowledge and thank Joanna for providing the leadership and resources for science.gov . It is really a great research tool for everybody. It really helps us to get our content out there for you. Thank you.

Thank you very much, Mary and Joanna. Fantastic webinar. Do we have any questions? Daniel made a comment. I don't understand that. I don't understand the question.

I think his question was for my presentation, Joe, when I was asking if people could see my pointer.

Okay. I see that. Let us see. Any other questions? Any questions for Joanna and Mary? I had a couple myself. The connection -- would going to science.gov be exactly the same as going to one of the agency databases? In connection be a 121 [Indiscernible] there or is it different in any way?

I can take that. There could be some differences, insofar as, to the fact -- that agencies on their databases, may have filters on the front end. Maybe you are going and I'm going to search -- a lot of times, I would say, most of the software these days you will do a full text search and then you will narrow your results after you get them. I still find that some agencies will use repositories on the front -- may be right at the beginning. Maybe you want to see this, but you only want to see this, this, and this. The query will be on the front now. We cannot automatically know that what is going to be. If someone went and did a search on somebody's repository and they did a search string, they would get a set of

results. If they do not do any narrowing and get search results, it will one through a relevance rank. You will see some similarities. Member, that after we take the relevance ranking that the repository is using on that side, we are going to run one on our side. There are going to be some differences, but it is only to enhance the relevance rank order of the entire [Indiscernible] of what is being searched on the science.gov site. If you do a one 21 search, for the most part, if there is top level results -- you should see those in the results. It is a hard question to answer. I hope I did an okay job.

I thought there might be major differences. That is good.

We get these cases from time to time. Because it is real time too -- I always say -- before I could say yes or no, it depends. Really, it is searching. It is still going to search. Let us say it is looking at ROSA P and it is looking at three other agencies. If I go to ROSA P, I sometimes will do comparisons. Do I get this in my search result here and my search result here? If you do see something that is completely different, I would want to look into that a little bit further. I would expect to see some consistency there.

That is great. That kind of goes with what Jim Gillespie asked. Hold on one second. Let me grab this question. He says, if I'm looking for a known item, would be best to start with ROSA P or science.gov ? You seem to be saying that science.gov would cover it, pretty much.

I would say that. As a general rule, I would first -- the point of science.gov is initially to prevent you from having to go to the repository itself. The idea is if you have a term -- agencies that fund certain things or research certain areas -- a lot of people make assumptions. We do a lot of work in the basic sciences. People do not realize that. If you definitely are looking for results from a particular agency, yes, go to the agency source. If you're looking for something on a topic and you do not care what agency it comes from, you are just looking for federally funded authoritative information than I would say go to science.gov. You may find -- oh my gosh, I cannot believe that, again, that the Department of Defense has some of these research results that you would not expect you keep your mind opened if you're looking for a topic. If you do, if you are specific -- I really want to find things from Department of Transportation, or I really want to find things from this agency, or this -- yes, go to the source. That is one of the things when science.gov was put out it was because a lot of people did not realize what other agencies were funding.

Thank you. [Indiscernible - muffled] Christine says when looking at the 71 sources covered by science.gov and some sources are timed out, what is the best way to get those results?

That is a good question. I think a lot of this is going to rely on, in that case, somebody's judgment. If something is timing out, again, I would suggest going back. If it keeps timing out, then letting us know through science.gov. If it is something that -- if it is biomedical and you are like, I am searching on something and I know it is biomedical -- I would probably go to Central directly. If you notice something that has to do with CO2 emissions, maybe you want to go to a D.O.E. source. In some cases, I would try that. There could be something because of the connector and how is working with the agency at that time. That repository -- it is not me though that repository is necessarily broken. It is probably getting a lot of traffic. We try to work through a lot of these issues with the system admin's at the agencies, sometimes things happen.

Great. Thank you. [Indiscernible]

On the science.gov website and there is a help or a FAQ. I'm going to look. If you want to go to another -- we have -- just looking for the FAQ. We have a description on the search itself. If you want to look here -- if you go into the footer of science.gov , you have got to expand on the homepage. Go down and hover. You will see that there is a whole overview of the best way to find what. It will go over what you can for your query and you can use operators. I need to go review that just to be sure. It'll give you some -- you can use operators and all of that. Hopefully, that answers. Send a message, Christine, if you are looking for something. We can help you with that. I would just go and review. You don't really need to put it -- I would still put it in quotes. It does not say you have to. Send us a message and we will work with you, Christine.

Great. Thank you. We have got a few more minutes to get in any last questions. Please, get them in. We are up against the 3:00 deadline. This has been a fantastic webinar. If I could slide in a quick question. The thing about preprints came up during COVID, preprint servers -- they wanted to get things to scientist quickly before it was published. I don't know if science.gov was thinking about that or has addressed that issue at all. Are you familiar with that?

I'm going to turn it over to Mary.

I don't want to take away from anyone else.

That is a very interesting topic. Anybody who has worked with preprints knows that they dominate certain fields. Physics is one that's come to mind. Interestingly, we had some preprints submitted recently to ROSA P. I have not previously seen these. These preprints came off of a server. We cannot put them in ROSA P. What I am unfamiliar with is the preprint servers that are open to everyone in a specific discipline. I am thinking about physics because it used to be a physics and chemistry library and. To my knowledge, we don't have any equivalent in government. Although, some of the repositories may have some preprints in them. We are collecting them deliberately and comprehensively. Unbelievably, there are some copyright issues now.

Thank you. I just read about it during COVID. They wanted to get information traded as fast as possible around the world during that period just something I read about. Joanna is saying if you have questions please email them there. If you can put the satisfaction survey in it, I don't know if we have time. We are running up against the time. There is the survey. You can see this presentation tomorrow. This link will be in our webinar repository. You will see that tomorrow or the next day. You can see the last two or three years will be up there, including this great webinar. I would like to thank Mary and Joanna for a fantastic webinar. Sorry, we have two eggs that pretty quickly. Please, think about presenting again. You have presented before and this is great. Thank you, audience. At thank you, Ashley. Check out our FDLP calendar. We have got a webinar tomorrow -- a terrific webinar read please, think about checking that one out. We have got a couple more scheduled for August. Office of minority health resource Center -- National Service of minority health and resources information. That should be a great webinar tomorrow. Anyway, thank you, all, one more time. Come on back to the FDLP Academy . Check us out tomorrow for another great webinar. Thank you very much, everyone. Have a great rest of the day. Bye-bye.

[Event Concluded]