Please stand by for realtime captions.Good afternoon and welcome to the webinar using science.gov to access U.S. government information. My name is Jamie and with me here today is Ashley on tech-support. If you have any issues with audio or any technical issues with the webinar, please contact Ashley. She will be monitoring the chat for any issues we might have. We're going to start off with a brief introduction of our presenters. First we have Joanna Martin. She joined us in 2007 as an IT specialist helping to support and manage projects in Germantown Maryland working closely with HQ IT operations. She managed and help to develop the searchable S WP project which replaced $1 million a year program management system at half the cost. Following that she transitioned to program integration supporting various projects within the team, analyzing program information and since then she has worked closely on public access to implement DOE wide requirements for grantees to submit their final peer-reviewed manuscripts. She also served on the STI it managers interagency group and is a cochair. And her position she supports the director of various activities. She has a bachelor's in sociology from American University and a Masters in policy, planning and administration for Catholic a minute -- University of America. She has one son and resides in the Washington DC suburbs. Our second presenter will be Mary Moulton. She is the digital library and at the transportation library in Washington DC. She is responsible for depository services, collection development and evaluation of new technologies that facilitate content discovery. Before we get started, I'm going to walk you through the normal housekeeping reminders. If you have any questions you would like to ask a presenter or if you have any technical issues, feel free to use the chat box located on the bottom right-hand of the screen. I will trick -- keep track of all questions the come in. We're going to do something a little different than normal, we are going to do the questions during the presentation instead of holding them until the end. But I will keep track of all of them and facilitate that. We're also recording the session and will email a leak of the recording and slides to everyone who registered for the webinar. We will also send your certificate of participation using the email you used to register for the webinar. If anyone needs additional certificates for multiple people who watch with you, please email us and include the title, today's webinar, along with the names and email addresses of those needing certificates. If you need to zoom in on the slides, you can click on the full screen button in the bottom left side of your screen. To exit, mouse over the blue bar at the top of your screen so it expands and then click on the blue return button to get back to full view. At the end of the session, we will share a webinar satisfaction survey with you and let you know when the survey is available and the URL will appear in the chat box. We appreciate your feedback, including comments on the presentation style in the value of the webinar. With that, I will hand the microphone over to Joanna.
>> Thank you very much I think the ball needs to get turned over to me.
>> We're doing that now.
>> Thank you. I want to thank everybody who has called in today to hear about science.gov. Today I want to provide an overview of the federated search functionality that science.gov offers where research in real time 60 databases from across various federal agencies, 22,000 -- 2200 websites and over 200 million pages of authoritative federal science information. I also want to share with you some new features that have to do with public access to federally funded research results back what I'm going to recommend to everybody here, as I go over science.gov and provide an overview of how the search works and what it does and what the goals are, if anybody wants to go there as we are going through the presentation. I will provide a search example and show you a search and the results. But if anybody wants to go to science.gov and do that in real time as we go along, I certainly encourage you to. Science.gov is an interagency, voluntary initiative that provide the gateway to U.S. government science information and offers free access to the research and development results in the scientific and technical information from many federal agencies. That have missions in science and technology. I have listed them here. Of the Department of Agriculture, Department of Commerce, USDA and for service, Department of Defense, education, Department of Energy, Health and Human Services, Homeland

security, we're providing journal articles for various other agencies from the Department of Health and Human Services, Department of Transportation, environmental protection agency, GPO, NASA and the National Science Foundation. The types of -- what I will refer to it as STI, which is information that includes journal articles, technical reports, conference papers, videos and these are the formats we can provide them in, video, audio, images, multimedia and scientific and technical data sets and collections science.gov is a voluntary inter-agency which is the flagship product. [Indiscernible] it is a high level science information manager and all of the various agencies. CENDI is been around since the mid-80s and science.gov was launched as its flagship product in about 2002 where it provided for the first time a public search to these vast stores of STI. Then in 2004, the federated search part of science.gov was implemented, which offers real-time relevance ranking of the research results and STI made available by the federal agencies that I previously referred to. I will go over federated search and how that works and how it is different in a couple of slides forward. Again, it is governed -- not only CENDI provides the primary financial support for science.gov, but science.gov is governed by a group of experts recommended by CENDI in the form of the science.gov alliance. That is a voluntary group of these departments that I listed. The science.gov is cochaired right now by Mary Moulton, from D.O.T. and Department of Energy. And OSTI, the office of scientific and technical information, host the website and provides the federated search technology as well. If you go to the homepage you will see we have a relatively new look and feel to the interface. We streamlined it about three years ago. It used to be a lot busier looking, now provides a simple search box and by entering a search query or search keyword, it will do a fulltext search and you can also search through in advance search. If you click on advanced search you can see that you can narrow the searches by the various collections of information, which in the advanced search, we categorize by different science categories. You can also search by the title and the author and do some other things with the advanced search. But really, it is very simple to enter your search term and just do a fulltext search. We provide clustering, which I will show in the search details in a minute. We have an alert service, and the results the come back, which again are the federated search real-time search technology where they come back in relevance rank order and can be filtered by various categories. Where the content is available. So right now you have categories that include text, which are mostly technical reports, conference papers, and other textural information. It is a catchall category for text. There's also a category for multimedia, videos, if there are videos or audio files, images, that kind of thing, there is a multimedia category tab. There's a data tab. And there is now a new tab called public access. That is really what I'm going to try to highlight mostly through the presentation. That is one of the newest focal areas in terms of providing access to peer-reviewed scholarly publication, journal articles, that result from the science agencies. The search technology we use is federated search. That is a little bit different than the types of search engines you are probably most used to using. Even as opposed to what Mary reminded me to point out, a new search on search.gov which searches various government webpages. It instead goes out and queries the agency selected collections of what we call sources, which are typically databases for the most part, but it will go out and search them simultaneously. It is a deep web search. It takes a little bit longer, and it returns back a set of top-level results. So it connects to these collections, and the results come back aggregated. They are returned again in relevance rank order, providing top-level results. So it will essentially go out using a special query that is built customized depending on what we're trying to look for in that database. The query will go out using the search term. It will bring back those results and provide just the top level results, which are -- which really -- so when you go into your search, it is not going to bring back every single result from each of the repositories, it will bring back the top of to 100 search results from each of those sources in relevance rank order. That way instead of having to sift through thousands and millions of research results to find the top level result you want, this search provides a great solution to not having to go and do that. That is kind of how the federated search works. Versus the other types of technologies, which again, which will do a static crawling. The next slide is a search

example. This is where, if you want to go to science.gov and you have two screens are looking at right now, I wanted to do a basic search query on CRISPR. You can see from this box, it is pretty easy. You put in your term, and then you will get search results screen looking similar to this. I'm going to to go to the next slide which gives you more detail on what the search results will provide. When I did my search for CRISPR, there were about 1300 top-level results that came back for my search. And they came back in two categories. To be honest, I am not sure, there is too much in those categories right now. That you can see that we have got the 1300 top-level results that came back in these categories. And that will include -- that gives you a combination of tech reports, conference papers, accepted manuscripts and journal articles. So you can kind of flip-flop between the text and public access tab. The public access tab, of course, is going to give you these journal articles and manuscripts from all of these federal funders. And the relevance of each result is denoted by the stars below. I think everybody is probably used to that way of determining what is more relevant than what is not. And again, up to 100 of the most relevant top-level results from each of the sources are returned. So again, so that you don't have to go through and sift through thousands, research shows that you typically don't go past page 2 or three, so this provides a really good useful tool to finding those top-level results. There is also clustering, which is offered by the application that does the federated search, which is called explore it everywhere. And that can help you cluster. You can see if you go to the website, you will see that you can refine that by certain topics. That is searching the content and putting it into clustering categories. So if you want to kind of narrow it down that way, you can do that as well. So when you get into your search, and you finish your search, you're going to see the search results page. I am in the public access tab here. When I go do my searches, it is doing the searching in real time. So you will see a status bar which is over under the red bar to the right where the arrows are that will say -- it will be like 10 of 72 sources complete. You will start to increase as the search is continuing in real-time. When it is complete, you can open up that -- you can really open it in any time, but you can open up the 72 sources complete box, and it will be a pretty big box. Will say search status, and then there will tell you how many top-level results it has returned from all of these sources from how many results are really at those sources. So it will provide you -- I will pray -- appeared to be a smaller so, but again, it is the top level result. It is not intended to find all the results. If you look here, and you're not going to see some of the agencies that have more results because I cut off the box, but for example, this has returned 85 results. That is because it he duped the result so it is not the 100. If you wanted to dig deeper to find things that were less relevant for whatever reason, you can go to the source by clicking on the source name in the search status box. That was something that the alliance was helpful to users. We try not to focus on all the results, just the top level results. So that is a really important thing that is a distinguishing feature between science.gov and just day -- the standalone repository that is centralized. That

>> --

>> We have one comment. James says he searched for the same search he did on the screen a got 44 text results.

>> Okay. I just did a surgeon got something different. So, okay. I am happy to talked him after. I can't see what he is seeing. Sometimes -- again, this is in real-time, so sometimes you may want to just try to search again. Right now I am doing the search. I'm not challenging what he is saying, I am saying maybe he will want to go back and try that again. Sometimes perhaps where you are with the connection, it is a little slower. I can't speak to that detail right now but I am happy to talk to him afterwards. I am going on the main search assets. Likes searches for research results. I going to turn it over to Mary. She is going to talk more about public access. But if we need to go back to looking at the search more, I am going to do some searches myself to see about the last comment to see if I can duplicate that, but Mary, I'm going to send this over to you.

>> Thanks. Hopefully everybody can -- hear me. Thank you. I am going to talk a little bit about some other features of science.gov and a little bit about public access as well. In addition to having the

federated search engine, and discovery within the search results, we also provide access to Tran 19 opportunities. -- STEM Opportunities. At the top of the slide you will see the option to select STEM opportunities. Basically those are opportunities for students where they can work for a government agency, and they will be able to search for STEM opportunities by program type, discipline, location, agencies often have opportunities outside of DC, as well as the agencies sponsor. So there are descriptions of the opportunities and links for applying. Other features, of course, involve information about public access. Public access is a very important, broad federal wide initiative. The government agencies that sponsor scientific and technical research either within their agency or by funding research outside of their agencies are now required to make the results of that research publicly available. Just a little word of caution here, many of us have been providing public access at various levels, but the real implementation of public access to the results of federally funded research went into effect within the last year. So for example, at D.O.T., a little over a year ago when we put out our grant announcements, we required for the first time that our grant recipients would agree to make the results publicly available. So what that means is, in another year or two or even three, when they are finished, completed with the research cycle, and they have publications that were the results of that funding, they need to send the national transportation library, where I work, they need to submit that content to us so that we can ingest it and index it and make it available through our institutional repository, the national transportation library. And that is happening all across government. So when you are really going to start seeing a lot of the results will probably be -- I am expecting the first wave will come in about a year from now. The national Institute of health preceded us in implementing this initiative, and they have been making their results available now for several years. So the other agencies are now playing catch-up. One of the reasons why having a great tool like science.gov available to all of us is that the agencies do have this mandate to make the results of the research they are funding publicly available. And there are several interagency working groups, CENDI is certainly one of them. I am involved in of the group, the interagency working group on open science . All of us see science.gov as one important method for providing public access. And here is why a federated search is important, not just for the public access aspect of making results available to all of you, but also because users or searchers of information, of government information may not know which agency was the sponsor of a specific research study. In my area, and transportation, we are not just roads and bridges and planes, trains and automobiles, we get involved in safety, research, behavioral type studies, urban planning, sociological type studies, and we know that we are not often a destination for people who are searching around for information. So that really is how I got involved in science.gov. When I started looking at the search logs, I realized that we were getting a tremendous number of referrals to our repository from the science.gov search. So that is a great thing. We do a tremendous amount of outreach, science.gov alliance does, but also my library, the national transportation library does. But it is often not enough because we are very focused on the transportation research sector, but we want everyone to be able to find the content they need for whatever it is they are working on. So back to the public access, science.gov does provide the ability to limit search results, and we have a category now called public access. So basically this is a surge of peer-reviewed scholarly publications that are the result of federally funded research, and all of this is being made available through our government repositories. And currently, this is from 15 agencies that are participating in the alliance. You can narrow your search to public access by selecting that tab that you see the Aero is pointing to at the top of the page. And so, that way you will just be focused on public access results. Another feature involving public access is that, from the home page, you can learn more about public access. We have a link on the science.gov homepage that takes you to a list of all of the agencies public access plans. And that is important because there is a lot of diversity among the science agencies, the government agencies that are sponsoring science. And so, there are going to be different nuances involved in how are you are making this available. In this is great. You can have access now to the plans. So you will better understand the

results that you are getting from the individual agencies. You will also find information about agencies submission systems. So if you are a researcher or investigator and have received federal funding, you can get information about submitting your publications for inclusion in these open access repositories. Here is where it gets tricky, but you don't have to worry about this. The other thing that kind of complements -- complicates government-sponsored research is that some is funded by more than when the -- agency. So as a submitter of your research results, this can be helpful to know more about the submission systems. And as a searcher of information, this can help you so that you don't have to know all of the agencies that may have been funding a particular study. The other really wonderful thing about the science.gov alliance is that we are part of a larger worldwide science alliance, and so all of the contents -- content indexed and searched by the science.gov search engine is also made available through the same technology to the worldwide science.org group. So you can access all the science.gov alliance information as well as participating countries, governments across the globe. You also have the ability to do your searches and translate search results in 10 different languages. And there are several of them listed here, 10 languages. So that really opens up to you literally the world of scientific research. Just a little bit about what is next. In support of public access, we are launching this new feature so that you can search across the entire collection of the 15 participating agencies. You can search across all of our collections with one search, and just to point out once again, on the home page, the simple search, you can check the search box underneath next to public access, which will just search the public access content. So that means that your search will automatically limit the results to those Karlovy peer-reviewed publication -- scholarly peer reviewed publications and manuscripts. I would also like to put a plug in for my colleagues maybe on the phone who work for agencies, federal agencies, that sponsor scientific and technical research. If you are interested in having your agency participate, we would love to have you and have your content. So if you're interested, you can contact either Joanna or myself. And we would be thrilled to give you more information about how you could join the science.gov alliance. So with that, I'm going to turn this over to questions and discussion.
>> It looks like we just got one in. Other special characters or arguments for search? For example, the phrase multiword searches, etc.
>> -- We have one question so far. Two. The first one are there special characters or arguments for searching? For example, phrase, multiword searches, etc.
>> Can you hear me?
>> Can you hear us?
>> -- I un-mute. I so -- I am so sorry. -- I mutated. -- I on mute. I am so sorry. I am going to refer this to Lynn Davis.
>> This is Marilyn. Yes, you can use a bouillon searching from the advanced search page. It is our experience, based on our Google analytics, that the majority of users don't. However, speaking to this group, I know you're really interested in that. Over time, we have had really nice detailed searching information. We have not kept that up today. Again, more because there has not been as much overall use, but we would be happy to -- if you guys want to vote for that, we would be happy to reinstate that information for you.
>> Okay. Another question. The public access tab says it includes peer-reviewed scholarly publications resulting from federally funded scientific research. Does this mean that the results are all peer-reviewed?
>> I can answer that one. So when we approached this public access project for science.gov, our intent was to develop queries that would go to the public access repositories that each of the agencies were using to make available the research results. In this context, I am talking about generally the final period accepted manuscripts. So we went to each of the agencies, whatever repository they had decided to use, whatever they documented in their public access plan -- some had not even built their systems yet, but at this point, we have been able to identify most of the agencies repositories because they have

some result in their. So the scope of that was to target the peer-reviewed accepted manuscripts. Generally, they are not going to be copyrighted versions of those manuscripts. They're just going to be the final peer-reviewed manuscripts. I think [Indiscernible] may be rolling in certain things, I can't say you will not find some copyrighted one. That is not what we are targeting, but some of the agencies have put things in that are copyrighted because maybe one federal author on it -- but to answer your question in short, yes, everything should be peer-reviewed. We did our very best to identify whatever would be peer-reviewed. I know also some of the agencies such as the Department of Education does not publish a lot in journal titles. They just don't. So they do their own reports. They do their own publications that are not necessarily technical reports, but there are publications that look a lot like journal articles but they are peer-reviewed. So you can feel quite confident that everything is peer-reviewed. There is always a caveat, but I want to say yes just with those caveats to be clear.
>> This is Mary. I would like to add to that very good explanation of the peer-reviewed content. I have an example for my agency. What our public access plan is requiring, the researchers who accept funding, that they submit their final peer-reviewed manuscript. We are not going to deal with anything that is copyrighted or asking people to retain the copyright. What we are going to do is embargo the content for a year. This is where it gets really interesting and why as a user of science.gov or someone who wants to look at this public access content, each agency has a slightly different public access plan. Some other agencies may have a longer embargo period, which means they are not going to make any of the information available for a year, for 18 months, for two years. Some may be pointing you to another platform, but what we are doing is just requiring the final peer-reviewed manuscript that was submitted for publication. There is nothing illegal about this. What it allows the researcher to do, the seeker of the information, is to read and use that information, and the expectation is that the author will get credit for the work, and actually studies have shown that usually the journal article itself is what is essentially cited. So what we are doing is perfectly legal, and if there is any confusion about this or what you are seeing when you pull the content back from a specific agency, it might be helpful to go and read that agency public access plan, which you can easily get to from the front page of science.gov.
>> This is Joanna. Thank you Mary. I think the devil is in the details. But the think about science.gov is it is not a database. It is a portal that uses federated search technology to access various databases. So each of the agencies, at least in reviewing all of their public access plans, which I did at one point in time, the goal for each of those agencies was to make their peer-reviewed content publicly available. That is what they had to do in terms of their public access planning. So they all make it available, a little bit differently. Most of the agencies, I am pretty sure most of the agencies are all going by the 12 month embargo. So I think that is pretty standardized with the agencies. Some of the agencies are waiting to make their manuscript available, but some, like DOE, what we do is create a citation record to say, here is the name of the article, because we will get it and it will come in, and we have to wait before we can make even the manuscript publicly decibel it needs to be one year from the date of publication. So still create a citation record and then put a date on it. if there was a question about some result available at a future date. That is because we want to let the public know we have this manuscript, it is going to be available. You can also create alerts with some of these systems. So if you want to be alerted when it becomes available, you can do that. I think it is fair to say that, in fact, I am going to say we worked very hard to ensure that everything in this public access tab, at the very least, it is peer-reviewed. Some reports maybe thrown in there that I -- are not necessarily journal articles, but you can be quite confident that is all peer-reviewed content. If it is an open access journal for example, we will still put in there. That will just be the open access article. Those are various ways we are making that peer-reviewed content available. But I appreciate what one of the attendees said about their students can only use peer-reviewed content. I think you can feel quite confident about that.
>> One of the things that we are exploring, something that of come up that I wanted to clarify, because I think the way the search works in terms of making the top level results available, which again, we only

provide up to so many results because we never intended to provide every single result. There are some things that are starting to be considered, such as, we know the NIH funds a lot of health and medical related science. They have vast stores. We may be changing some of the criteria for how many top level results we bring back depending on sizes of repositories. So you may start seeing changes with that over time, but again, the ultimate goal is to bring back the very top level results, kind of like trying to find the diamond in the rough. Trying to not have to go through a needle in a haystack. And that is really what you will find if you get more familiar with using science.gov. But anything you notice as a librarian, I think it is very helpful to get any feedback on ways we can change. I just wanted to clarify those points.
>> I see there is another question. I know you're fielding these questions.
>> Does the search return data sets as well if they search for them directly?
>> Good question. I want to say that we do have some data sets in here, but I would not say that science.gov at this point -- and Mary you can comment further -- I don't think that science.gov right now is really -- we really have not tackled making the data sets available as much as the textual journal article type information. The textual stuff. That is not -- but I think that is probably going to be on the horizon.
>> Yes. I think that is a really good assessment of the current situation. And again, this is going to vary a little bit by agency. And I can talk a little bit about what we're doing at the national transportation library right now. We migrated a year ago to a new platform where we could catalog and link data sets to publications, and we started to do that for our internal D.O.T. publications. So when you search science.gov and get search results from the national transportation library, it is possible that you would get linked data in publications. But we have spent a huge effort over the last two years to implement that, and I think that where it may be somewhat unusual for a government agency to be doing this -- so a lot -- the type of data is going to very a lot. And again, speaking from my own agency, data is not just like a table of numerical values. It can be audio, it can be video, it can be images. We have all of that survey data. We have all of that at supplementary datafiles in our repository. So not every agency is able to deal with that. And we have actually been talking about this for many years now, and there is the data.gov registry of data sets, which is one place where you can go to look for open data. But I do anticipate, this is not the end of it, there is a lot of discussion that is happening about this right now, and I think moving forward you are going to see more and more either catalog records that include a link to a data set which may be located someplace else, or actual data sets that are being crawled. But we are at the very beginning stages of implementing the data aspect of public access. So stay tuned. This is an exciting thing we are embarking on right now. And I think it is only going to get better.
>> Another question. Since it's not a database, can we call this a repository from different Fader -- federal government Paris Staples -- parastatals?
>> I think you could. We pretty much call it -- I think we call it a portal. It is funded jointly by the federal agencies, that's for sure. So if you look at the definition of that, I think that is sort of what that means, that it is kind of controlled wholly or partly by the government. So I suppose you could call it that.
>> Another way to think about it though is more like an index to the different repositories. It is in fixing -- indexing the full text as well as the metadata on any of the items in the repositories. This happens in real time. You're always getting up to date search results. That is why I would hesitate to use the repository term to describe it. It is more like [Indiscernible] index created on the fly every time you do a search.
>> I think that is important to highlight, the real-time nature. That is why people can wanted to what you search right now depending on the responsiveness of the other agency repositories. You may get something a little bit different. It is hopefully not vastly different, but it can differ a little bit because of the way the to elegy works, but the benefits to that is that if something is put into a repository today and let's say this afternoon you're going to find it probably. It is not like you have to wait for things to get updated, but I think it is probably not calling it a repository, but maybe a gateway or index. That is a good point. I also just went and did a search with the data, it looks like some of the way the data is being

pulled in, it is being pulled in is that we actually -- I believe we have like data sources. So if somebody at an agency has a particular -- like here we have one that is the United States for service data catalog. That is actually a database that contains data information. And so, that might be a way our next steps will be, if the agencies have places where they are making their data available, even though I think most of the agencies are trying to do what Mary described as interconnecting data to the article, and then maybe even related software, kind of interlinking all these things. But I think of agencies have particular data centers, that we can build connectors to those. So they will get pulled back. But I don't think that we have a vast amount of that right now.
>> Okay. --
>> Also, to be honest, if anybody out there as a librarian, if there is a data source that is a federal data source, something that you think your patrons want to access, it would be great to hear. Because those are things we can deal with and can build connectors to. And generally speaking, we have mostly representatives for most of the science and technical agencies. We have a couple that we don't right now, but we have most of them.
>> We still have a couple more minutes for questions if anyone has any more questions. Please feel free to chat them in the bottom right-hand side of the chat box. Ashley has sent out a link to the webinar satisfaction survey. We would appreciate if you would take a couple minutes and fill that out. It helps us with programming here at GPO. And while I am waiting for any more questions to come in, I'm going to let you know we have a comfortable more Academy offerings coming up. The new. -- Repository library Institute is coming up on October 10 and October 11. It is a partial day course. You can sign up for that. And we have train maybe for the Olympics using government resources to learn about appropriate exercise on the 16th at 2 PM Eastern. If you're interested in either one of those offerings, sign up. I am not saying any more questions. Just want to take the time to thank our presenters for this really awesome presentation, and also thank you to all of the audience for joining us today.
>> This is Joanna. Thank you very much. I appreciate all the participation, and I am sure I speak for Mary, but if anybody has any questions, specific or want to talk through anything related or have suggestions, please don't hesitate to reach out to us directly.
>> [Event concluded]