Introduction to the USGS Publications Warehouse – Transcript of audio

Please stand by for realtime captions.

Good afternoon, will come to the FDL P Academy, I am a user support library and that the U.S. government publishing office DPO, I will be the MC for the room today, and GPO library assistant Ashley Dalen, will be available if you need assistance, introduction to the USGS publications warehouse, Kelly Haberstroh, digitization library and at the U.S. geological survey library. For the Q and day, add the questions in the chat and send them either to all panelists were all participants. For the question-and-answer session.

For the blue bar, the top of the screen, activate the chat, I will monitor the chat and questions will be answered at the end of the talk, and provide in the chat the links that are featured in the slides as our presenter speaks. This presentation is being recorded and will be made available shortly. I will now hand the microphone over to our speaker will take it from here.

Thank you so much Helen, I am Kelly Haberstroh, like Helen said, I work at the U.S. geological survey library at the USGS headquarters in Reston, Virginia, I have been with the USGS for 7 1/2 years now, I am the digitization library and, so I oversee the in-house scanning operation of the library, also the product owner for the USGS publications warehouse, which is what we are going to talk about today, going to give some background on what the publications warehouse is, what content it contains, some of the features of the website and some search strategies, and how the site feeds information feeds other sources, I will turn it off during the presentation but I will come back for questions at the end, so let's get started.

I wanted to make sure I have a little background in the U.S. geological survey and the USGS library, formed in 1879, and the science arm of the Department of the Interior, we perform it in a wide array of subject areas, not just geology, you can CA nonexhaustive list of areas we cover, including natural hazards and resources, ecosystems, and for mental health, effects of climate and land-use change, the USGS library, authorized by Congress in 1879, one of the largest Earth science libraries in the world, a collection of over 3 million items, one of the resources that we manage at the library is the publications warehouse, so let's get into that.

What is the publications warehouse? The authoritative catalog and publicly accessible location for accessing USGS peer-reviewed scientific publications, a public website, everybody can access, you can see the URL for it on the slide here, on the slide, and in the publications warehouse, users can access metadata about and links to more than 170,000 publications authored by USGS scientists, over the century plus history of USGS, to be clear, the publications warehouse is not the library collection of 3 million items, it is just a catalog of a subset of items that have been authored and published by the USGS. Users can obtain free online access to the free full publications published by the USGS, and articles and published by external publishers on websites as well when those are available. I just wanted to point out the publications warehouse is just one resource that is available to the public for USGS information paradox -- products. We are part of an ecosystem of resources across USGS, USGS data is catalogued in the science data catalog, and science base serves as a trusted digital repository for USGS data, models are catalogued in the model catalog, and there are a number of different resources to access maps from USGS, including the national map, the national geologic map database, and more, I will go to in those in a little bit more detail in a little while, printing the publications and the maps from

the USGS store, and of course the library collections of 3 million items are discoverable through the online catalog, and the main USGS website, and many of the resources that you see here are also indexed@USGS.gov, and use it on the website to find many of the publications data and maps as well.

All of these different resources, what can you find any publications warehouse. A little breakdown of the content that you will find there. Everything contained in the catalog is authored, and USGS publishes the own report and map series, we call typically numbered series reports of maps and also historical USGS reports published outside of the numbered series as well. Published by the USGS, they are available online through the publications warehouse, free to download by anyone, all of the new USGS published reports are disseminated through the publications warehouse directly, you can also access thousands of older reports that were issued in print and have been digitized by the USGS library and are available online.

In the middle here, you see that we do also catalog all you USGS authored and funded publications published by external entities, for these, we are talking about journal articles, conference proceedings, books, chapters and cooperative publications, and extended abstracts. Published by publishers other than USGS, we don't host those versions of the records and full provide full text access, through the warehouse but we do include links to resources, usually by object identifier when they are available, users can access these resources through their own institutional subscriptions, public access avenues, and publish, open access immediately then anyone can access them.

On the far right, pointing out the type of products that are not catalogued in the publications warehouse. Those products are things like abstracts, posters, presentations and data releases and software releases. Abstracts and posters and presentations are generally more ephemeral rather than permanent information resources and they may not receive the same level of. Review of products including the catalog, they are not in the publications warehouse, and they also have their own designated repositories in the U.S. GS, it is focused on public it -- publications.

I wanted to pause here, to better illustrate the type of content you will find in the publications warehouse, showing some of the most popular USGS publications that are available through their. When I was looking through the metrics I was pleased to see the most popular publications actually demonstrate the wide breadth of science topics covered by the USGS, publications on minerals, water use, mental migration, statistical methods, remote-sensing map projections in all different types of science. You can see here that the most popular publications are mostly more recent USGS publications, available to download the full text for free of course, we do have have scans here as well, a map projections manual from 1987 here, and a map from 2000 that have been digitized from our print collection in the library and made available online.

We also have a journal article here, authored by the USGS, and external scholarly journals, and that is also popular as well.

Hopefully they'll provide a taste of kind of what you will find in the publications warehouse, after this presentation, hopefully you will be inspired to explore some more.

Before we get into the website some more, I wanted to provide some additional background information on the publications warehouse, established in 2004, it has been around for about 19 years at this point, and under the USGS library, located organizationally today, it is the product owner but we

do collaborate with other groups within USGS, science centers and publishing group and policy group and more, it is a collaborative effort to bring all of this together.

Who makes of the staff of the publications warehouse, I will give more background, it is a small team and quite spread out but generally we have three functional components, the first one is cataloging, we have a small team of part-time contractors hosted at the USGS wildlife healthcare Center in Madison, Wisconsin and the librarian who oversees them there. Everything is catalogued in the publications warehouse is verified and corrected in curated by the cataloging team. The metadata scheme in the publications warehouse is custom, and the cataloging team catalogs based on a set of established rules that we have internally, but the metadata can be cross walked to the Dublin core.

All newly published USGS products, in addition to cataloging legacy, ones that have been previously published but not yet captured in the publications warehouse. Update and make corrections to enhance existing records as well.

Development support and maintenance of the publications warehouse application, including the services, database and user interface components, and all of that is supported by contractor Illinois, and also part of the time of the operations team within our larger program within USGS, who provide operational and infrastructural support the warehouse and other USGS applications as well.

To clarify why winning the support, the publications warehouse, the backend, not off-the-shelf or out-of-the-box tool or solution, it is custom built to fit the needs of USGS, so all of that development and operational infrastructure support that would be provided by a vendor if it was off-the-shelf or something is instead here in the house and USGS.

The core staff cut digitization, digitizing print USGS publications and maps in house in Reston, betting the fulltext digital content to the publications warehouse. Over 90% of all of the publications in the warehouse are available and ready to download, it continues to go up as we digitize more reports in the library. The digitization library, the other part of the job, I lead the effort there, to contractors doing scanning work in Reston, in addition to me we have a couple of library staff helping with the coordination and processing and uploading of those scans to the publications warehouse. As you can imagine, the scanning was limited during the Covid-19 pandemic, we had a little dip there, now much more time available with those scanning backup. A couple more details about the digitization operation.

We are scanning the items in house, in the library, the priority is scanning and making available the last 10% of USGS numbered series reports and maps that are not yet available online. We do prioritize requests that are made from patrons, both within USGS and the public, if there is a particular report or map you would like to see and there is no digital version available in the publications warehouse yet, let us know and we will get that up in the scanning queue, we want to support current research as much as possible.

The location is unique, we have materials and many different formats including bound volumes with large foot up plates of large maps and microfilm and microfiche, CD-ROMs and micro this and more, many remaining materials that still need to be digitized require special handling, and season images, and in large brown that was, in the book, and envelope, overstuffed with the publication that was available at the microfiche only, a variety of different scanners, to support the in-house digitization effort, and a multitude of different formats, adapting to the needs of the material, and new scans added to the publications warehouse added all the time.

Okay, we have a general understanding of what is in the warehouse and the groups involved in the care and feeding of this, and get into the website itself, some of the features, as I mentioned before, in the publications warehouse are edited and approved by the cataloging team and was there ready to be published, made to the public in the warehouse catalog. On the website.

Here's just a little bit about the website design, the publications warehouse website is responsive, you can easily access it on your phone or tablet or desktop computer, everything resizes nicely so you're not scrolling horizontally across your phone. The website is built on semantic HTML 5, so the content is more accessible and meaningful to all users, including users -- humans and machines, it uses a U.S. web design standard, the standards for federal government website design, it is standards-based, and the application of the website to the cloud in 2020, there is some increased stability and support from the operations team that I mentioned before.

I'm not going to do a live demo today but show you everything on the site with some screenshots in the next few slides.

So this is the main homepage of the publications warehouse, the first thing that you're going to see when you go to pubs .ER .us.gov, I wanted to highlight some of the features here, right in the middle of the page, the basic search box, talking a little bit more about basic search of the next slide, right below that you will see the advanced search button. You can click that to open all of the options that are available for advanced search and I will give you more details about that in a moment. Few ways to discover publications from the front page here without searching. At the bottom of the page, we have this continuously updating our cystlike feed of new publications that have been added to the publications warehouse. It always shows the most recent few publications here. At the right side, the publications warehouse Twitter feed, we do have a Twitter account that we treat out new and interesting USGS publications usually twice a day.

At the very top navigation bar here, you will see the explore button allowing you to browse the content by report, type and series if you don't want to search. Also the top of the page you will see a link to our documentation, providing frequently asked questions as well as documentation for the API web service that we provide. And lastly at the top of the page you will see the contact button, which is available when you are looking at any page on the website, this contact form will send our team an email, and respond, within one business day. And a little bit about a basic search year, most users use the basic search, the option, we are all familiar with on most websites, and in the basic search, and title series name and author, abstract, year and larger work title. Fields within each record to return results. And the Boolean search operators like and, or and not, are not supported in the basic search, and doing ebullient like search using the advanced search options, which I will show you next.

Here is some of our advanced search options that I mentioned earlier, I've highlighted some of the most useful options in fields to search year, many more options beyond these I'm showing you here, you can be very granular about the search if you wish with all the options provided, free text, and some of them are brought -- drop-down fields, drop-down menus, and those dropped on many fields are controlled and have a controlled vocabulary to choose from. As I mentioned earlier, you can use the advanced search options to do Boolean like searching, so if you use multiple different fields within the advanced search, they will perform a search that contains all of the search terms, a.k.a. an and search, on the right, they search for a series name and open file report, and year published 1973, when we perform the

search the publications warehouse look for publications that meet all of those conditions. So it is performing and search for them all.

If use the same field multiple times, the publications warehouse will perform and or search for that field. For example on the right, we have a search for series name, open file report, circular publication title water, year published, and that is 73 and 1975, and the same field multiple times, and using the series name twice and the year published twice. So for those repeated fields the publications warehouse will perform and or search, we will get results for publications in the open file report or circular series and results for publications published in 1973 or 1975.

So the advanced search is pretty powerful, and you are able to refine your results in quite a detailed manner here.

I mentioned earlier that there are a number of other more comprehensive resources within USGS to find maps, the publications warehouse does contain a selection of maps as well, mostly those that were published in one of our numbered series. I've given an example of how you could build meds at -- advanced series query to build the downloadable maps in the publications warehouse and you can axis those for free, and we are limiting the search to the USGS numbered series results and results that contain one of two types. Plates or sheets, the or where you will find the downloadable maps.

A screenshot to the right, a little bit blurry to me, the search brings back over 21,000 results, many maps to download if you wish. And you can narrow it down and start getting more specific, for example you can add in Colorado, the basic search bar in addition to the advanced search fields we have set, and the downloadable maps that contain the word Colorado in the title or abstract metadata. And you'll see over 1700 results from that search.

Once you build the search query and hit the search button, this is what your search results will look like. You can see in this case, this is a basic search for the Colorado River. You can see your search query at the top. And easily edited, and search again, and brought it, directly from this page, the search results are shown here at the bottom, and it will show the title, contributors, series, report number, year published and a little snippet of the abstract to try to let you know if it is going to be relevant to you before you go ahead and click into the main page. The title is a clickable link you can use to access the full citation page that we will take a look at any minute.

Right above those search results, you will see the download options if you want to analyze this set of results or export it in another way, you can download the result to an RAS file that you can upload to your favorite citation manager, or several delimited formats available as well, TSD and Excel, downloading the results to a spreadsheet is another great way to do some additional filtering and sorting of the results, a lot of metadata, useful to drill down even further.

Once you select a resource from the search results and click on the title, you will be taken to the citation page for the publication, there's a lot going on here, I will walk us through. Starting at the top, you will see the title for the publication in large font, the report series and report number, journal articles, the journal title here, moving down you will see the contributors for the publication, regular authors just show up with the by indicator here, editors and compilers would show edited by or compiled by. Below that you will see the digital identifier, or deal wife for the publication, moving down into the links section, you will see all of the downloadable content available, this that I'm showing on the screen, is the series publication published by the USGS, so the full text is freely available for you to download.

You will also see any associated data or software links that we have for this publication in the links section as well, and they will take you to another repository where the data or software is held.

Down here you will see you can download the citation, in a couple of different formats if you wish, if you would like to imported into a citation manager or double encore.

Moving on, you see the full abstract for the publication, and the citation page here, search by keyword and return results that have the keyword in the abstract. And over here on the right, you will see the contact information for the publication. This is included for all newer USGS reports, if you have questions about the content, you can reach out to the contact provided.

If you continue to scroll down the citation page, you will see more features available. At the top here you will see the suggested citation for the publication, which is available for all the newest USGS reports, the citation is in USGS citation format because of course we have our own citation format. They are created and maintained by the publishing group here in USGS.

Below that you will see the study area, an interactive map you can zoom in and out of and apply different layers if you would like. These polygons are created and added by the publications warehouse cataloging team. Not like actual shape files, just meant to give a quick visual reference for the study area for convenience to users. Finally if you continue to screwed on the page you will see the rest of the metadata for the publication, including a lot of different things but includes the publisher, the USGS science center, a description of the product and some additional location metadata as well and at the very bottom you'll see a link to a metrics page, which will give you some statistics on how many times as been visited. And Holly times they have been downloaded. And I wanted to go little bit more in depth, and available in the data model for the publications warehouse here. As you might know, a digital object identifier or DOI for short is a unique persistent identifier that is meant to provide permanent access to a resource even if the URL posting location changes in the future, all series publications, even the ones published in the late 1800s, have cross referenced you Isaac assigned to them, and make them more citable and discoverable. Published by external entities catalogued in the publications warehouse, we provide the external resource when it's available, and that will take you to the external publishers site, where it is located.

In addition to cross referenced DUI for publications, the data site DOI is linked to the data associated with the publication, in 2016 it became a requirement for USGS authors to release it simultaneously or before the associated publication is published, it became a goal for the publications warehouse to make the connections on the citations page, and the link is provided to the USGS data release associated with the publication on the citation page, and that is a data site DOI.

And and the IDs, open researchers and contributor ID, and persistent identifiers for people rather than resources, they help to disambiguate authors with the same name and help track a researchers work through name changes in organization changes. And enter them, into the review and approval tracking system, so the publications were for warehouse stores the IDs for the USGS publishers, I mean authors. On the citation page, and the publications warehouse. The little green circles next to the author page, it will take you to the profile, where you can see the entire list that has been captured there, and the Orchid IDE, you can unambiguously find the publications in the warehouse by searching for it.

The citation pages, a colorful little circle with a number in the middle, and the doughnut, and the metric score, for me publication, and online attention score for individual publications, assigning the score based on how me times the publication has been mentioned in online sources, social media or policy documents and more. Click on the donut to see additional information, for that publication, see what sources have given the publication attention. And has the metric score, I will have the badge on the citation page. One of the tools that we used to support the USGS, the publications, and external venue, the journal article,: pay wall -- a service that identifies open access versions of articles either through the publisher themselves and other repositories like a institutional repository, and the publications warehouse uses the API to automatically bring in links to open access versions of the articles when available, adding them to the warehouse citation page for the publication. And as I mentioned, looking at the citation page before, they all have a metrics page available which shows you the page views and download statistics for the publication, and view the data in the chart or the table format. And the metrics viewed, only goes back 12 months, and can actually retrieve metrics, and further back, requests when needed. And then lastly the future just released a couple of years ago, and the USGS publishing group has begun creating XML for each USGS numbered one that they publish, and required in the plan, the publications warehouse, the automated process, and transformed into the webpage on the fly, basically when you click on the HTML link for report on the citation page, you will really be able to view the whole publication on the webpage. You can see the full content right there on the browser.

Interested in same example, and the report 2021, it is a good example, Helen has just put that in the chat, thank you. Okay, and the features of the website, talking about the resources that use the publications warehouse API, the content elsewhere, a basic flowchart of how information products move through the systems in USGS and were they ended with the resources, we will be briefly looking at each of these other resources, and the index, using the information from the publications warehouse. And it has a public API application programming interface available, providing a web service that can be queried through a technique and return data in a structured manner, the API is open to anyone to use, available on the website, and some nice documentation, as a link in the resources at the end of the slides here. When you query the API you will receive that structured one like this, this is an example of part of the JSO when that you would receive, JSON, the data behind one of the citation pages that we looked at earlier.

One of the resources that uses the publications warehouse API, is the library catalog, and online public catalog of physical resources, as well as the late Tronic resource subscriptions at the USGS library, in addition to those, integrated publications warehouse records into the library catalog using the publications warehouse API, you can search for the content for the three through the library catalog itself, you can limit the search scope to just publications warehouse content in a library catalog and and the filters, offers offers additional search options beyond what you can find directly on the publications website. What I find really useful is limiting it to everything, and both the physical library collection as well as the publications warehouse. And the screenshot here, I search for the title, and the search returned results for the publication in the warehouse, the top result& Check the online source, as well as the physical copy, and that bottom results there, and if you like to see, you might want to have the tangible physical item in your hand. Strike a screenshot of what a publications warehouse record looks like in the library catalog, the warehouse has a custom metadata scheme, I can be crosswalks, to create the records in the catalog. Is updated with new records from the publications warehouse, lightly, every 24 hours we see new and it dated records appeared. The main USGS website, and the publications warehouse API, as a couple hundred science centers all over the country, and produce products and publications in each science center has its own webpage on the main USGS website. It is important to our organization that each publication is attributed to the science center office that produced it. Each

citation page and the publications warehouse has a contributing office tag on it that identifies which office produced it.

So using the publications warehouse API and contributing office tags, the main website automatically pulls in and displays publications and the specifics of a specific science center has produced other individual webpage.

Is of the upper Midwest science center webpage on the main USGS website, over 1900 publications that have been automatically associated with the webpage, because of the API, you can click on any of these publications, from the page here, you will see the metadata data from the publications warehouse and follow a link back to the actual publications warehouse record for. And similarly, staff also have profiles, on the main website as well, using the same logic but with people, pulling and publications from the API, and using the orchid IDs it automatically pulls and is single staff members publications and displays them on the staff profile webpage here, really appreciate the feature, automatically builds the publications list for them.

Science-based, a trusted digital repository, is home to much of the data that USGS releases to the public, on our science-based data released landing pages, and the publications are referenced in the related external resource section, and link to the publications warehouse where the publication is. The USGS model catalog is another catalog for models and similarly science-based associated publications are referenced on the model catalog pages and links back to the publications warehouse.

Probably one of the most important users of the warehouse API is Google and Google scholar, it is fully indexed by both of these search engines. Search for the publication and Google scholar, the warehouse link is often one of the first things to turn up in the results, a lot of web traffic comes from Google, we did see an increased spike of usage from the site we got indexed Google in 2013.

Another great resource, from the publications warehouse is science.gov, federated search portal, access to U.S. government science information from a number of federal agencies and we have worked with science.gov to improve the connectors they have built, and make sure that the publications are effectively searchable and effective through the tool. And another user of the publications warehouse API, and a database of geoscience literature, the American geosciences Institute, and publications in their database with the help of the publications database API as well. And then finally a resource that I'm sure many of you are familiar with is the catalog of catalog publications or CGP, in order to comply with the statutory mandate in 484, USGS submits a list of all the new publications monthly, added to the CGP, and made discoverable to users throughout the outlet. Okay, before Christopher Topeka, would to further expand on the resources from the USGS, and the publications warehouse contains a subset of USGS published maps, there are ones available, a lot of interest to folks, here are some of the resources that I mentioned before, the national map contains a suite of products that help support all of your mapping needs, including the one that allows you to browse double maps and data from the USGS, and and more information about, and allows you to browse access and download USGS topographic maps, from all different time periods, and it is the spot where you can purchase printed maps as well. I just wanted to give a little bit more detail about the database, MG MDB in particular, and you will see some overlap between the content, and the publications warehouse. And an archive of geoscience maps including those from USGS and state geological surveys. The areas where you will see overlap with the publications warehouse is with USGS published maps that are part of the USGS numbered series, they're both in the publications warehouse, and we work with the MG MDB, and we reduce efforts and share files in the catalogs and link to the associated NGMDB record from the associated warehouse and vice

versa, they link to the publications house. But the NGMDB offers some additional download options that the publications warehouse does not, so TIF files that are of higher quality than the usual files for maps and some are geo-referenced in the NGMDB but on the publications warehouse. Just so you know, you have several options to discover maps in the USGS.

I did just want to share some statistics from the past fiscal year hear from the publications warehouse, fiscal year 2022, just to give you an idea of how many users were seen at the publications warehouse site, we do get several million visitors and page views per year and to give you an idea of the scale of cataloging that we do in a year, we added over 4600 records, to the publications warehouse in fiscal year 22 and edited or updated almost 14,000 more.

Wanted to point out the publications warehouse is consistently a top website in the interior in terms of usage, usually in the top 10 or top 15, hopefully providing a service to the public. And just before I wrap up the presentation today, I wanted to point out more additional resources related to it, a nice little library guy that provides an overview and searching tips for the publications warehouse that I shared with you today, and put out guidance documentations, and requirements for indexing, processes and workflows involved, and individuals for the warehouse, on the website here, general FAQs on the website, and more specifically focuses on fundamental science practices, USGS, scientific integrity principles that guide our work. At USGS, FSP and FAQ related to the publications warehouse on the website as well, before we have web service documentation available on the website for the API as well as some swagger documentation. And so that is all that I had for the slides today. And happy to take any questions, and will do my best to answer them.

You always answer correctly Kelly. [Laughter]

So far we have two questions, if you have additional questions, added to the chat. The first question is, does the search function support wildcards?

Great question, I did not address that, it does not currently support wildcards, unfortunately you can't use the star, and part of the word and bring back results, and the basic search is pretty flexible, and if you know exactly what you want, it is very powerful.

What is the highest alt metric score.

That's a very good question, the last time I looked in alt metric for USGS authored publications, I think the highest one was around 2200, 2000 or so. I would have to confirm if that is still accurate.

That is incredible if that's the number, wow. Phyllis one of your favorite publications?

Not prepared for this, probably one of my favorite publications that we have featured on the Twitter account, for hollowing before is a water publication a circle publication, to that it has no scientific basis, I would have to find a link to that and put in the chapter you guys.

Tells little bit more about the minimum standards workflow for high-quality scans that go into the publications warehouse.

We do have standards for all of the scanning that we complete, we scan everything at a minimum with 400 dpi -- PPI, pixels per inch, and black and white pages with techs on them, scanning grayscale, I think we are like a three-star on the FA DGI standards, if you're aware of the federal digitization standards.

We have a lot of tools that help us create high-quality scans, the hardware we have in house at USGS supports that in the software that we use to complete those scans helps us create high-quality products.

I do have some other presentations on digitization, specifically at USGS, that I would be happy to point you to. There are some public recordings that are available from the library that maybe we can send out afterwards.

That is a great idea, we will keep those attending posted about the additional links.

One more tiny question, is there work right now to provide USGS style citations for the older publications?

Great question. Like I said, those USGS style citations are being curated by our publishing group within USGS for all of the new series, I would think they would go back to probably at least 10 years at this point. But there is not currently an effort to create USGS style citations for the older series reports, I do think it is something that is desired by the community, and something that we could suing at this time, and we are not producing them for older citations.

Rep of the talk at this time. Your presentation was fantastic.

We are sharing a survey link in the chat, take a moment to let us know how you felt about the webinar, it has been recorded and you will be notified shortly thereafter when it is available to you. If you enjoyed today's webinar, check out some of the upcoming FDO P Academy webinars, 2 links in the chat, the first is about the Academy, FDLP, and the calendar of events, thank you for a great presentation and have a marvelous day everyone.

Thank you. [Event Concluded]