Geo Data Portal: Translating Climate Data for Geographic Analysis

Transcript, August 29, 2012 Webinar

Danielle: Introduce Shawn Carter, who's the lead scientist for the National Climate Change and Wildlife Science Center, who will introduce our speakers.

Shawn Carter: Thanks, Danielle, and thanks to all who are joining us today. I'm pleased that we're partnering with the National Conservation Training Center to offer this series of webinars that highlight the work on climate change impacts that our center sponsors. Today's webinar is going to feature the Geo Data Portal, and it's going to be presented by Nate Booth and Adam Terando. Nate Booth is the Lead Architect and Program Manager at the USGS's Center for Integrated Data Analytics in Middleton, Wisconsin.

CIDA partners with USGS programs to provide innovative computing solutions for conducting interdisciplinary science across a growing array of massive environmental data collections. His team led the development of the Geo Data Portal.

Dr. Terando is a Climate Scientist and Research Coordinator in the North Carolina Cooperative Fish and Wildlife Research Unit in NC State. His research focuses on the impacts of climate change ecosystems and agroecosystems, and the complex human-environment relationships that drive these processes.

With that, I'll turn it over to Nate and Adam.

Nate Booth: OK, I think I have control. Hello, everybody. It's a good turnout today. I appreciate everybody taking time out and listening to the webinar. I'm going to present first and then I'm going to turn it over to my colleague Adam. Then I'm going to provide some closing remarks. Then we're going to open it up for questions. Let's jump right into it.

The agenda for the talk will be to give a bit of background on this project that was supported by the National Climate Change Wildlife Science Center program back in 2009, which has led to the Geo Data Portal. Adam will lead with a series of examples and I'll follow on with a few additional examples. Then we'll provide a summary and we'll get into some discussion.

Back in 2009, we began this project. The title of the project was "The Development and Dissemination of High Resolution National Climate Change Dataset." It involved a partnership of PIs including Jaime Collazo from North Carolina State, Lauren Hay from USGS Modeling of Watershed Systems, Katharine Hayhoe from Texas Tech, myself, and Adam.

Today, we're going to focus on the second component of the project, which was developing a system for disseminating these large downscaled climate datasets.

The motivation for the project, or at least this part of the project, was we wanted to make it easier for scientists and managers to discover what data exists in terms of these downscaled climate models, make it easier to access those large data, and process those data for analysis.

I've also listed here the four members of the Geo Data Portal development team. Two of them are here with me today. The team includes Dave Blodgett, Tom Kunicki, Ivan Suftin, and Jordan Walker.

The initial design, or conceptual model, for this tool that we purposed back in 2009 is illustrated here, and the idea was that we had recognized that for supporting various kinds of models, such as hydrologic models, there's a need to access a variety of datasets, including downscaled climate, but also additional landscape data as well.

We recognized that folks were spending a lot of time accessing these data and reprocessing these data, oftentimes, in similar ways, somewhat duplicative. We wanted to devise a system to simplify this process and provide some consistency for preparing all these data for analyses. This is what we started with, and this is what we've moved towards with the Geo Data Portal.

Before the Geo Data Portal, this is how it sort of worked, and this may resonate with some of you, and it's probably the way it still works for quite a few folks in the community. Oftentimes, when you wanted access to a large downscaled climate model, you made a phone call. If you were lucky, you were able to access something like this, which was a large terabyte hard drive.

They put it in the mail truck, deliver it to your office, and you'd work on it, in a GIS and other software, to prepare it for what you really wanted to do, which was to run it into a model or a series of coupled models. Then, like I mentioned before, this frequently was revisited one project to the next.

So what we've intended to do with the Geo Data Portal is to short circuit a lot of that processing, and make it easier to access these various large datasets, using the Internet as the mechanism for connecting all these various data resources with scientists.

We'll get into these examples later in the presentation, but in addition to accessing climate data, and summarizing climate data for geographic units like watersheds, which is shown in the lower-right here, the Geo Data Portal can now also be used for real-time weather, such as NOAA's NEXRAD indicated rainfall dataset. It also can be used for landscape variables, like the National Land Cover dataset.

Those of us that work in the informatics field have been moving towards this paradigm which is nicely written about in this paper called "The Geo-Processing Web." The idea is to use again, the Internet as a mechanism for moving data around and providing sophisticated and powerful processing of data through a web interface or through the Internet. This whole idea, this whole paradigm is quite popular in many places across the community including a large initiative within the National Science Foundation called Earthcube.

This slide which I borrowed from Cliff Jacobs also shows several prominent journals that are talking about big data, the challenges of big data, and the approaches for dealing with these large

datasets in support of geosciences. Again, this is another example of what we're trying to help satisfy in this community with the Geo Data Portal.

The approach that we took was first to understand these common patterns of accessing large data and processing the data for analysis and modeling. We spent quite a bit of time understanding the various communities both the folks that were generating these sorts of datasets and the folks that would be using these datasets. The Geo Data Portal provides this translation between these very different communities -- the atmospheric community and the geosciences community -- hydrologic modelers, ecological modelers, et cetera. The Geo Data Portal serves as a broker between these communities.

In order for us to provide some sustainability to this design what we did was we tried to respect the way that these various communities like to work with their data. From a geoscience standpoint, they like GIS, we like GIS. From an atmospheric sciences standpoint, they are more familiar and comfortable with other types of tools and software. The Geo Data Portal sits between those two communities.

Another guiding principle is that we wanted to provide these capabilities which we'll describe through examples, not just through one tool but through a series of tools that we'll also describe.

Then finally we've spent a lot of time dealing with how to optimize this solution for dealing with large datasets over the Internet.

The tool is based on common standards. I'm not going to get into a lot of details on this slide, it's mostly for reference, but we've put great attention towards using common standards again to provide some sustainability to the approach as well as to have reusability between the various datasets -- not having the tool focusing on one particular dataset, but rather a whole suite of datasets that can provide interfaces according to some standard in some standard way.

A lot of that work, a lot of the standards development that has been evolving to support this sort of approach has happened within the Open Geospatial Consortium and that's where we spend a lot of our time.

Now I'm going to move into just an animated demo of how one would interact with the Geo Data Portal to access climate information. Once you come to the website which we'll provide at the end of the webinar, you first describe the area of interest that you have. It could be anything from one to a series of watersheds. It could be a national wildlife refuge, it could be a political boundary -- a state or a county -- any arbitrary area across the landscape that you want to summarize geographic or climate information to.

What we've done is provided a GIS file that describes the upper Flint watershed in the southeast part of the U.S., south of Atlanta. Second step is to come into the catalog that is behind the Geo Data Portal to identify the various datasets that are accessible. We've spent time with a variety of atmospheric scientists in making linkages to their data holdings within the Geo Data Portal and some of those are shown here.

Once you select a dataset of interests you can see a full metadata record that can describe the methods that they used as well as the variables that are accessible as well as, importantly, proper accreditation to the authors of the dataset.

Once you choose your dataset, you provide some configuration to how you want to process that data as well as a time frame or a time range. Then oftentimes, not surprisingly, the processing will take a few minutes to even an hour or so, so you can specify your email, submit it, and then once the processing is done, you get an email response that echoes back the processing steps that you put in your recipe as well as a link to the file to download.

That's quickly how one would operate the Geo Data Portal and now we're going to jump into a series of examples which will hopefully make this much more interesting and real for how this can provide value to your science or your other analyses. Can you switch it over to Adam, please?

Danielle: Yes.

Adam Terando: Thanks, Nate. We're going to pick it up with a couple of examples that scientists have been working on when they have been using the Data Portal. Like Nate says, from a scientist perspective, when you're doing global change research or climate change research, there's lots of data and lots of models and so the question is how can you spend less time dealing with individual datasets or individual modeling products and more time getting to your analysis and doing some science.

This is pretty evident when you can see that there's the CMIP3 archive which has all the GCMs from the last IPCC assessment. You can spend lots of time searching through here to get the proper model run and the proper variable. There's meteorological observations you can get from a different website. There's the climate wizard which has visualization tools; it has another downscale website.

There are also dynamically downscaled projects such as NARCCAP, where you can also retrieve data. Then there are local or regional projects such as with the La Florida project sponsored by USGS at Florida State. Even USGS in-house we have our own dynamically downscaled projections.

There's lots of data, again, lots of models. It can be a bit confusing and a bit overwhelming to be able to spend your time getting these data and figuring out how to get it not only for the types of models and variables you need but for the specific place your analysis is for.

Can we use the GDP to facilitate the research process and spend less time on these sorts of activities? We have a couple of examples here. The first one is from Dr. Mike Osland from the National Wetlands Research Center. He was using the Data Portal to look at climate impacts on potential changes in the range of mangroves in the Gulf coast and along the Atlantic seaboard.

The hypothesis is that winter climate is an important driver for the salt marsh mangrove forests in the southeastern US. We know that mangroves are very sensitive to minimum temperatures, and then you have this division between mangrove and more salt water marsh type systems along the coast.

How might winter climate change impact where mangroves could be located in the future? The first thing we need to do is find some data to be able to address this question and so we can go to the Data Portal.

First thing you can do is upload a shapefile for the area that you have. It doesn't have to be a shapefile or an area that is specific to what the Data Portal already has. You can upload something that you're working with and the Data Portal has no idea, has never worked with it before, but it'll still work here.

For instance, here we have these coastal polygon IDs which are attributed to National Wetlands Inventory data. Once you've uploaded your study area then, again, as Nate was saying, you pick different attributes that you need to configure the Data Portal with. Then you can pick your climate variable. In this case, we're looking at a downscaled dataset that has projections of minimum daily temperature for the next century, for the 21st century.

Then you submit it for processing and you wait to get your email back. Then you get your email saying everything is all ready. And then you have your data and you're ready to go, and you get a nice table output.

It could be netCDF file format, if you're familiar with that, which is pretty common in the climate community, or it could be a CSV file or a comma separated file as well.

Here, we see that this is that relationship that Mike has found. He could help build this model with the Data Portal, as well, since there is observed climatological data in there. You see here, this is the predicted percent of tidal saline wetlands that become dominated by mangroves as the minimum temperature increases. This is what we have currently in 1970 to 2000 and then you see how that corresponds to the map of the mangrove presence today, pretty much limited to southeast coast of Louisiana as well as around Florida.

Then you can look at the projections, based on those downscaled climate projections that we obtained from the Geo Data Portal.

This is one scenario here. Let's see. I'm going to switch over to my little wand. This is with a B1 emissions scenario. You can look at this for a lower emission scenario, so more of a climate stabilization type scenario. The colors here correspond to the predicted percentage of each polygon that is dominated by mangroves.

Versus a higher emission scenario. This is the A2 emission scenario. You see the change and the difference in the projected areas of mangrove dominance.

Another example, and again, this is from Mike Osland. We want to look at, say, how does precipitation affect the coastal wetland occurrence along the gradient on the western coast of the Gulf of Mexico.

Starting out we have some idea of how the system might behave. In this case, that precipitation has a big impact on the type of wetland that you would see along the coast.

You get a change in the ratio between unvegetated and vegetated wetlands in Texas.

The question is how could this change? The objective here is to use precipitation and the coastal wetland occurrence data to develop some kind of model of abundance of these types of wetlands along the coast.

We already have the coastal wetland data and so where can you get the climate data for your observations and your projections?

Again, use the Data Portal and you can get your past climate data. In this case, we're looking at precipitation and you see the timestep here; this is daily data. You have it for these individual grid IDs.

Then you can develop your model. You have your percentage of tidal saline wetlands that are dominated by unvegetated and you look at the relationship between that and precipitation.

Hopefully, it's really improving your ability because you're taking out all that time devoted to getting sometimes terabytes worth of data and then manually slicing and dicing that dataset up to get it into both the format and in the area that you need.

Then the final example I'm going to show you is another aspect of the Data Portal that we've been working on and right now we call it the Derivative Climate Portal. This has been with a user group with myself and some other folks in the Fish and Wildlife Service or in the LCCs and in the USGS and in NOAA as well.

The idea behind looking at this Derivative Portal is we wanted to look at climate indices because we realized a lot of times if you get a climate projection of, say, mean temperature change over 70 to 100 years, it's really hard to link that to the biology or the ecology. Animals or plants or organisms aren't necessarily going to respond if the mean temperature changes from 60 degrees Fahrenheit to 65 degrees Fahrenheit. What does that mean exactly?

You might be able to find in the literature that there is different thresholds or indices for a particular species that are very important and so can we speed up this process through the Data Portal and be able to calculate some of these indices on the fly and then access those?

Here's the website. We'll go over this again. Again, this is a little bit separate from the Data Portal but the same project so it's the Derived Downscaled Climate Projection Portal.

This is what it looks like. I'm just going to step through it here and show you some of the capabilities. We're not really focusing particularly on the specific downscaled dataset for this because, hopefully, eventually we'd like to expand this and be able to use lots of different climate datasets, but just to see how it works so far.

The first step here is you have this interface. To get rid of all the junk like all the junk in the website above and below the picture, you click this little button here and that gets rid of that.

The first step is to pick the index you want. Right now we have a set of pre-calculated ones. We would hope to be able to expand these as time goes on.

You pick what you want to look at. I'm going to pick observed and projected days where the maximum temperature was above some threshold. Then you pick your threshold. For the days

above max temperature. Right now, there are three thresholds that have been pre-calculated; but, again, in the future, we'd like to be able to do this on the fly where the scientist or manager can put in their threshold, whether that's 88 degrees or 102 or 70 degrees, whatever makes sense for that species that they're looking at.

Then you're going to choose your emissions scenario. These are all the emissions scenarios that were used in the last IPCC assessment -- carbon dioxide, other greenhouse gas emissions. They go basically from highest scenario of emissions at A1FI down to the lowest, which is B1. Then you can pick what do you want to look at for the map? Do you want to look at the mean of all the models or do you want to look at an individual climate model?

I'm going to pick the ensemble here and then you're going to pick your time period that you want to look at as well. Since we're looking at a map, we really want to look at the average over a time period so we pick 30-year averages. You can pick which period you want to focus on.

Once you've made the decision, here I'm looking at just the A1FI emissions scenario for the ensemble mean of all the models and I'm looking at the past or the hindcast projection from the climate models.

Then you can look at the future, as well. Now I've switched to the end of the 21st century projection and you can see the change in the colors here. The scale here is showing the projected number of days that are above 90 degrees Fahrenheit.

Then you can easily really quickly switch to a different emissions scenario so you can see here the big difference between A1FI and B1. You can also then change how you want to depict the data. Instead of just looking at the raw projection, you can just click this button up here and it'll show the change from the 1961 to 1990 baseline period. That may be a little clearer information as to what's going on.

Here you can see the A1FI projection and you see large projected increases in warm days above what was seen in the past. Here it's just summarizing that. We have the A1FI projection, the B1 projection, the low emissions, and then we see the change from the B1 to the A1FI. Here, it really becomes very stark the difference in the two emissions scenarios.

This is the emissions scenario that we're really on track for right now. We're actually a little bit below the emissions in this one and this is the very optimistic climate stabilized scenario.

Then I'm going to show one more thing you can do with the Derivative Portal. I'm going to switch it up so now I'm looking at frost days or the number of days the minimum temperature was below 32 degrees Fahrenheit. This might be a little bit more of an ecologically meaningful variable for a lot of species.

What I'm going to do now is before I didn't do anything with this pull down menu but now I'm going to pick one of these areas of interest and you see there are these different options.

I'm going to pick LCC boundaries. That's a pretty useful one for the work I do. Now, what I'm going to do is I'm just going to click on an LCC that I want to look at. I'm going to pick this one.

I'm showing you all this stuff that I cut out before, but once you do that it'll calculate on the fly the time series. The average number of days where the minimum temperature was below freezing averaged over that LCC, wherever their data. You can see this is the decline in the number of frost days, not surprising, as the temperature gets warmer.

Just to zoom in here on what this is actually showing, there is a lot of different things that you can manipulate with this time series here. You can change the emissions scenario; you can show the range of the projections; and you can also download the data.

Here's just an example of what this would look like. The shading here is the individual model results and then the dark line is the mean of all the models.

Here, I've turned on two emissions scenarios, the high and the low emissions scenarios, so you can see the difference. Here, I've taken off the range, so uncheck the min and the max, and it just shows the mean projection for all emission scenarios and that makes it a little easier to compare across the emissions scenarios.

Finally, I'm going to download the data and I'm going to choose this as a CSV file. Then you're ready to do some analysis. Here, you have the years and then you have the output for each climate model, for each emissions scenario.

This is just looking at it in Excel, but I'm more of an R guy. Just a couple of commands in R and you're ready to go with some analysis and hopefully make your research life a bit easier.

Finally, just to show really quick, we're moving on into some additional global change datasets that we're going to be serving off of the Data Portal. A lot of these are from the Southeast Regional Assessment Project, which I've been involved in, which is funded by USGS and the Southeast Climate Science Center, which are these websites down here.

We're going to be serving the urbanization projections, sea level rise projections, as well as landscape change, which is based on our vegetation dynamics work. The point is that the Data Portal is a really flexible format and we're hoping that we can get a lot of these datasets, even aside from climate data. Some of these we've already put onto the Data Portal that will facilitate the research process.

With that, I think I'm going to hand it back over to Nate.

Nate: OK. Thanks, Adam. I see I'm being made the presenter. I'm just going to run through a few additional examples and then we're going to open it up for questions. Another thing that we wanted to be able to support with the Geo Data Portal is to not only make it easier to access an individual model but, like what Adam showed in the Derivative Portal, access to various models and various emissions scenarios and then see how those various scenarios are impacting other sorts of models.

In this case, we're looking at how they affect a hydrologic model for the southeast. This is just a quick two slides from Lauren Hay and Andy Bock from the USGS National Research Program and the Colorado Water Science Center, both in Colorado. Let me see. I think that's a little big. OK. Then advance this. There we go.

What they've been doing is developing this approach that's going to look at a variety of downscaled models and, again, a variety of scenarios within each and then access those data for individual watersheds within their hydrologic model, run it through their hydrologic model, and then compare the output for the hydrologic model for each of those various runs to one another.

Lauren and Andy have lots of preliminary graphics to show how this is coming out and it's all work in progress. The main point we wanted to make in this presentation is that this sort of thing really was not possible in any practical sense before the Geo Data Portal.

To access any one of these datasets and to process that data into hydrologic units and to run them through a model would have taken months individually. What we're able to do is this sort of important comparison now in a matter of weeks. That's a great efficiency improvement.

Second example, I'm going to show a couple of slides here that are using some of the new urbanization model outputs that Adam just mentioned. The Geo Data Portal for another hydrologic model project, “Watershed characterization and climate forcing”, at least that part of the hydrologic modeling project that, again, focusing on the Southeast Regional Assessment.

You've seen these sorts of screenshots before. This is the watershed of interest. What we're going to do is we're going to access the National Land Cover Dataset. The dataset that is specified here is from the USGS National Map for land cover. We're going to access the PRISM, climatology dataset, for this watershed and we're going to access the urbanization model output that Adam just mentioned. Then, from each of those runs we can quickly look at how that watershed is characterized and then look at the forcing data for that watershed.

This is the breakdown of the National Land Cover Dataset categories for that watershed.

Here's the time series of percent urbanization, for that watershed, from the urbanization model.

Next, here is the prism data, the min/max temperature and precipitation for that time period for that watershed.

This is another neat example. This is more of just a demonstration but it's a neat example of some of the unintended uses that the Geo Data Portal is now facilitating.

Rich Signell is a scientist from USGS, Woods Hole. With the drought this summer he was curious just how much was precipitation down during the dustbowl of the '30s. He used the Geo Data Portal for this analysis.

The other thing we wanted to show with this example was how you can interact with the Geo Data Portal, not just through the website that we've been showing but through a programmatic interface that's available through Python, which is becoming increasingly popular amongst tool boxes of scientists these days. There's a nice Python library called IPython that helps you sort of tell this story.

Here's an image, from one of Ken Burns' documentaries, showing the dustbowl.

We're next going to open up a connection to the Geo Data Portal and we're going to specify a particular county from northwestern Oklahoma that we're going to summarize a historical

climatology for. Don't worry about the code here, it's just mostly being shown to illustrate how you can access the Geo Data Portal using Python.

We're going to specify a time period and then, of course, one of the nice advantages that you have with the scripting language with Python is access to all sorts of visualization libraries. So after running that command against the Geo Data Portal with the Python script, Rich was then able to develop this time series graphic which indeed shows this major decrease in precipitation in the late 1930s.

Then, this is the last example. This is an example from a scientist in the USGS Wisconsin Water Science Center, Steve Corsi, who's using the Geo Data Portal for accessing real time NEXRAD indicated rainfall to support near shore water quality models. In the Great Lakes, the beach water quality is impacted largely by tributaries and we want to be able to characterize the flux from tributaries through models that take inputs including rainfall.

This tool, called EnDDaT, which is built on top of the Geo Data Portal, let's you specify your watershed for an area that's near your beach and then can access real time rainfall and summarize it and then bring it into your model.

That's another example that was unanticipated at the beginning of the project but just another example of how with building this system on open standards we're able to adapt it for other sorts of data besides just the long time series downscaled climate model output data.

OK. So, we're just about 40 minutes in. We're going to do our summary and then we're going to open it up for questions.

Hopefully, these examples have shown how the Geo Data Portal can provide a convenient way to access these large datasets. We're continuing to work to provide access to additional data and new datasets as they become available. We've put the URL for the Geo Data Portal here on the bottom of the slide. We encourage you to try it out and to provide feedback to us through the form on the website and we'll try to follow up with you and incorporate any of those comments into future work as resources allow.

With that, I'll just thank everybody for their attention and I think we're going to turn it over and open it up for questions.

Danielle: Great. Thank you for a really great presentation, both Adam and Nate. The way we usually do questions for our webinars is by having people press the "raise hand" icon on the main screen, it's in-between the participant list and the chat box. It's a little white hand. If you press on that, you can raise your hand and I'll call on you in the order that you raised your hand. We did have one question typed into the chat box from Lucas Fortini. Lucas asked, "Any future plans to include Pacific and Caribbean Island data into the data and/or the climate index portal?"

Nate: Adam, do you want to take that one?

Adam: Sure. Yeah. Ideally, we definitely would like to. There are some more dynamically downscaled datasets available in the Caribbean. I think, actually I might punt it back to you, Nate. But, I mean in general, I would love to get those in there. We have one project now where we're

working on some statistically downscaled projections for Puerto Rico. But yeah, I think it would be great to add those in.

Another thing that we should probably reiterate is that using the standards you can, if you have any sort of OPeNDAP dataset then you can use the Data Portal to access that. It doesn't have to be one of the 13 that are specifically listed there on the library of datasets.

Nate: Yeah. So just to follow up on that, during this presentation, we've purposefully tried to keep it non-technical. But for making additional data available, provided those datasets, those models, are available, we've gone through this several times now and we have sort of a step by step set of instructions to make your data available in the sorts of standards that are needed for those kinds of data to be accessible within the Geo Data Portal. So we can follow up on that.

Danielle: OK. Esther Stroh, you can ask your question now. You'll have to press star six to un-mute your phone from the global mute.

Esther Stroh: OK. Well, first, I just want to say wow. But I do have a question and I missed just the first couple of minutes. Now, I know that the operability won't be the same everywhere, but does the portal have some data available for other places besides North America, like Europe, specifically Northern Europe?

Nate: Adam, I'll probably defer to you on the various models that are accessible. We do have a global runoff model that's accessible through the portal. Adam, do you want to help me out on any of the other climate models that might be accessible?

Adam: At this point, I think mostly it's been U.S. datasets, and Canada. But that's a good point. I would hope that we can get those up there. I don't know if Dave is still sitting next to you, Nate, and if he knows of anything else. But again, you can still use, for the raw climate projections or if you have a European dataset, if it's in that OPeNDAP framework, you can still access that through the Data Portal. You can always use it to subset data that are out there that follow these open standards.

Esther: OK. Thanks.

Danielle: OK. To follow-up on the first question, Theresa Valentine asks about data for Alaska, if that will be incorporated.

Adam: Right now there is one set of downscaled projections for Alaska. I think they're provisional, on there. Those were done by Katharine Hayhoe. But those are up there. I know we were talking about getting other downscaled datasets on there as well. I'm not sure if they've done Alaska yet. There is one dataset up there now, but it is still provisional data.

Danielle: OK. Jeff Stoner, you can ask your question now. You'll need to press star six to un-mute your phone.

Jeff Stoner: Yeah guys, that was a great presentation. Are you planning on follow-up presentations for other groups? This is the first I've heard about this and I'm sure there's a lot of people that would be very interested in this capability and tool.

Nate: Yeah. Hey, Jeff. Certainly. Yeah. Just let me know what forums you think would be useful and we'd be happy to put those on.

Jeff: OK. Great. Thanks. Nice job guys.

Nate: Thanks.

Danielle: OK. David, and I apologize if I mispronounce your last name, Szczebak. You can ask your question now. Press star six to un-mute your phone.

David Szczebak: Hi guys. Thanks a lot. That was a great presentation. This might be kind of an ignorant question because I really don't work with downscaled data very much. Do you have any guidance in terms of working with that data that we get from the portal and elevation data? Is, generally, elevation included in the downscaled data that we'll get?

Adam: No. That's a separate dataset. I think there is an elevation dataset that's already uploaded in the Data Portal, if I'm not mistaken.

Nate: I think the question might be more around correcting for elevation in the downscaling techniques, Adam.

Adam: OK. Yeah. For the most part, they are already taken into account. That's one of the first things that is usually done when you downscale because that's such an important local effect on climate that the raw climate models usually won't be able to capture because it's just too coarse of a resolution. So that information is already embedded within the downscaled projections. It's just more of a matter of what's the resolution of the downscaled climate projections. You don't want to go too fine of a resolution because you just don't have the information to actually...You're not improving your accuracy any more if you go too fine of a resolution.

There's, typically, some kind of compromise you have to make there in terms of getting it at a grain size that will be able to capture the local effects of climate change but also still being at a scale where you have some confidence in your ability to model processes at that resolution.

That's almost like a topic for another day. But I know there's lots of guidance coming out about those things from the Department of the Interior. I think Forest Service also just came out with a guide on downscaling.

David: OK. Great. Thanks.

Danielle: OK. I have a typed question from Joy Marburger. Joy asks, "Can you analyze multiple variables simultaneously for ecological predictions related to climate change?"

Nate: Yeah. Whatever variables are available within the dataset you can access and summarize those simultaneously through the portal.

Danielle: OK. Nancy Green, you can ask your question now. You can press star six to un-mute your phone.

Nancy Green: Hi. Thanks for a terrific presentation and thanks for just developing this. It looks like a great tool that will have a lot of applicability. I have a question about spatial scale. I notice in some of the slides you had predetermined spatial scales, like LCC boundaries and other things like that. How much flexibility is there if people want to send you a GIS file or specify just some irregular boundary, basically?

Nate: Sure. This is Nate. Two part answer. For the Geo Data Portal which gives you just a time series output from the model, from the downscaled climate model or other datasets, you currently can provide any arbitrary boundaries of any scale that you're interested in by uploading a common GIS file format called a shapefile format. For the Derivative Portal, which Adam showed, which gives you access to indices, which is one step further in terms of processing, currently, yes, that portal is limited to a set of areas that are predefined. We'd be open to considering other sets of areas. We're hoping that if resources allow, we can provide folks with the capability of summarizing those sorts of derivatives for any area of interest. But that's not available currently.

Nancy: OK. Thanks. Really helpful.

Danielle: OK. We have a typed question from Cindy Swanson. Cindy asks, "Does the dropdown menu have the ability to look at land designation? In particular, run models for wilderness areas?"

Nate: I'm guessing that the dropdown menu she's talking about are the predefined areas of interest. I guess it's a similar answer to the last question. If you have a GIS file that describes wilderness areas, you can upload it and you can summarize these data for those particular areas of interest.

Adam: Yeah. But again, for the Derivative Portal, it's only those predefined ones right now. I would be interested in building more capability into that. I guess Nate can confirm or deny this but it's a tradeoff on the processing time to do that.

Nate: Yeah. It's just something that we've heard from other folks. It's good to hear others have that particular interest and it's just a matter of prioritizing that, again, as resources allow.

Danielle: OK. Phil Morefield, you can ask your question now. You can press star six to un-mute your phone.

Phil Morefield: Thanks a lot. I just wanted to ask a couple of questions. The first relates to additional datasets. In this presentation, you mentioned NARCCAP's specifically so I'm wondering if there are any plans to incorporate that particular dataset in the future. I'm also thinking about the CMIP5 files that are largely available now and if you have any plans for those datasets you can talk about that for a minute. And then, also, I just wanted to ask when the Derivative Data Portal is expected to be out of the beta phase.

Adam: Nate, I'll let you handle those.

Nate: OK. One at a time. The NARCCAP, we don't have access to it yet. We've heard that request. We'd like to develop access to it. For a lot of these questions, we're at a bit of a disadvantage because the fellow that's really responsible for a lot of the architecture and has been

working out access to new datasets, his name is Dave Blodgett and he's on his honeymoon this week so he's not here to answer these sorts of questions. We're pinch hitting for him.

NARCCAP is not available now. It's something that we've heard and we will work towards it. As I recall, there were some data access restrictions that we were needing to deal with for that particular one.

CMIP5, we are working on with Lawrence Livermore. You'd have to follow-up in terms of ETA on that. I think it's going to be another few months, at least.

The last question was...what was the last question?

Phil: When does the Derivative Portal move out of beta?

Nate: Oh, right. The Derivative Portal is in beta because the underlying downscaled climate model that it's based on is also provisional. As Adam mentioned, we intend to provide those Derivative Portal capabilities with other climate models. Once the dataset that's used currently is no longer provisional or when we make the Derivative Portal available for models that are published, then the Derivative Portal will be out of beta. That seems a bit confusing, I realize that, but that's the reason it's in that state currently.

Phil: Thanks.

Danielle: OK. Ed Laurent, you can ask your question now. Press star six to un-mute your phone.

Ed Laurent: Hi, great presentation. I was wondering if you were using a grid standard or the integration of your different projections and for the export of the data, the military grid reference system?

Nate: If I understand your question, again I'm at a bit of a disadvantage on this one. I'd ask you to follow up with Dave Blodgett who we can provide contact information for. The Geo Data Portal does have the ability to do projections, if you upload a shapefile of any projection the Geo Data Portal will convert it appropriately.

Also we've spent a whole lot of time carefully being able to intersect things that are in a geographic reference system like our GIS files with atmospheric data which is a whole different situation with the way they deal with spherical earth and their grids. That's about all I can say on that and again, I just ask you to follow up.

Danielle: OK, we have a typed question from Ryan Boyles. He asks can Adam perhaps comment on measures of error or skill that might be available.

Adam: Yeah, so right now, it's just the raw output. I've talked with Dave Blodgett, again on his honeymoon, before and we've discussed the possibility of adding those types of features on here. One thing would be operationalizing some different methods to include measures of uncertainty in these projections -- such as Bayesian model averaging or other methods that would get you probabilistic projections and would incorporate that uncertainty into these projections. Another thing that would be good on the Derivative Portal I think would be to add in the observed datasets

on to there as well. That would be pretty easy to then calculate measures of bias off of that. It's a good suggestion and it's something that I think we're moving towards.

Danielle: Our next question's from Tom Kurkowski. He asks, "Is all your metadata standardized on the ISO standard? I believe the FGDC is encouraging agencies to migrate to this international standard." And then he has another question that asks, "Are you utilizing GeoNetwork as your metadata manager or server?"

Adam: It sounds like one for you, Nate.

Nate: We are using GeoNetwork and yes, we are basing all the metadata on ISO 19115, which is embraced by FGDC. That's one of several standards that the Geo Data Portal is based on. We have a whole series of other presentations that deal with the architecture in more detail and these sorts of matters that we'd be happy to provide or share, as time allows, with folks like yourself. Again, feel free to follow up.

Danielle: I have another typed question. If there are any more questions just either type them in the chat box or press the little white "Raise Hand" icon. The next question's from Nicole Athearn. She says, "We're excited to see a Python API. Is there documentation on how to address data packages and models?"

Nate: The Python stuff is pretty new. I feel like I'm ducking out of a lot of these questions but I'd encourage you to follow up and we can put you in contact with the folks that are developing that. It is our intention to absolutely turn those into Python packages with full documentation but I don't have a concrete ETA on that.

Adam: Along those lines too, I've been really interested in being able to do something similar and say MATLAB or porting it over to R which I think has a pretty good Python interface. Personally I'm really excited about that as well.

Danielle: OK, we have another typed question from Forrest Melton. "Is there a process in place through which research teams can submit data for inclusion in the Geo Data Portal? Can the Geo Data Portal host new datasets or does the Geo Data Portal require that the partner host the data on their own OPeNDAP server?"

Nate: Yeah that's a very good question. We've devised a system so that all of those options are possibilities. Obviously, for the Geo Data Portal to do the hosting and serving of the data, there are resource considerations, but if a team can make the data available over OPeNDAP and we would work with that team to configure it in an appropriate way so that it works, it performs well. Then that's an approach that would work right away. We need to go through a bit of a process before we can reference that dataset officially through the website but certainly that's an option.

Another option within USGS that we've recently been working on is with an effort, a project called ScienceBase and what ScienceBase is all about is hosting project level data. The effort that we're on the way with ScienceBase on is where a team can upload their project data into ScienceBase and then there will be a linkage to the Geo Data Portal where others can access and subset those project data using the Geo Data Portal. Those are three approaches that we're supporting at the moment.

Danielle: OK, Forrest followed up with that and asked who would be the best point of contact for proposing new datasets?

Nate: You can follow up with myself. I assume, this is Nate, you'll get my contact information after the webinar if you don't have it already.

Danielle: Then another kind of communication question, Christopher Barnes asks if there's an email group that can be used to announce updates or developments to each of the two portals, do you guys have an email list that you use for anything?

Nate: That's another very good suggestion. We don't do that, we participate quite frequently with an effort called the National Climate Prediction Platform which does have such a capability and if you're tuned into that you're likely will have seen that we've presented updates to that forum quite often. That's a very good idea and we'll take it under consideration for some sort of LISTSERV or email outlet.

Danielle: Great, are there any more questions? [pause]

Danielle: OK, I think that was a really great presentation so thank you both Nate and Adam. Thank you everybody for all your great questions and for really actively participating. Shawn, did you want to say anything before we sign out?

Shawn: Just thank you everyone, I thought it was tremendous and I really appreciate the turn out and the questions, everything was great.

Danielle: All right, this webinar will be recorded -- or has been recorded and it'll be edited and posted within probably the next week, by the end of next week. Everybody who registered for this webinar will get the link. And then, I think that the Climate Science Center is also working on posting it somewhere so that people can access it easily. So, that will be forthcoming in case you were not able to view the whole webinar or you know someone who wasn't.

So, thank you, everyone, and we'll see you again soon. Shawn, do you want to tell them when the next webinar is going to be?

Shawn: I don't have the date handy, but it'll be approximately the...

Danielle: I do, I have it. So, the next one for September is going to be September 27th, which is a Thursday, and it's at three p.m. All right. Thank you all. See you next time.

Adam: Thank you.

Nate: Bye bye.

Transcription by CastingWords